Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: May 27, 2013

JM. Valin
Mozilla
C. Bran
Plantronics
November 23, 2012

WebRTC Audio Codec and Processing Requirements
draft-ietf-rtcweb-audio-01

Abstract

This document outlines the audio codec and processing requirements for WebRTC client application and endpoint devices.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on May 27, 2013.

Copyright Notice

Copyright (c) 2012 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction 3
2. Terminology 3
3. Codec Requirements 3
4. Audio Level 3
5. Acoustic Echo Cancellation (AEC) 4
6. Legacy VoIP Interoperability 5
7. IANA Considerations 5
8. Security Considerations 5
9. Acknowledgements 6
10. Normative References 6
Authors' Addresses 6

1. Introduction

An integral part of the success and adoption of the 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 client implementations.

2. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD 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 clients, a minimum set of required codecs are specified below. While this section specifies the codecs that will be mandated for all WebRTC client implementations, it leaves the question of supporting additional codecs to the will of the implementer.

WebRTC clients are REQUIRED to implement the following audio codecs.

- Opus [RFC6716], with any ptime value up to 120 ms
- G.711 PCMA and PCMU with one channel, a rate of 8000 Hz and a ptime of 20 - see section 4.5.14 of [RFC3551]
- Telephone Event - [RFC4733]

For all cases where the client 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. Clients MAY use the offer/answer mechanism to signal a preference for a particular mode or ptime.

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 kHz to 48 kHz and up. 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.

AUTHORS' NOTE: The idea of using the same level as what the ITU-T recommends is that it should improve inter-operability while at the same time maintaining sufficient dynamic range and reducing the risk of clipping. The main drawbacks are that the resulting level is about 12 dB lower than typical "commercial music" levels and it leaves room for ill-behaved clients to be much louder than a normal client. While using music-type levels is not really an option (it would require using the same compressor-limitors that studios use), it would be possible to have a level slightly higher (e.g. 3 dB) than what is recommended above without causing interoperability problems.

Assuming 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 client 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 active speech have a level of 2600 (-19 dBm0) for an average speaker. If the client does not have control over the entire audio capture, as is typically the case for a software client, then the client 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 client 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, clients MAY increase the gain before playback.

5. Acoustic Echo Cancellation (AEC)

It is plausible that the dominant near to mid-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 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 clients, echo cancellation will improve the user experience and should be implemented by the endpoint device.

WebRTC clients SHOULD include an AEC and if that is not possible, the clients SHOULD ensure that the speaker-to-microphone gain is below unity at all frequencies to avoid instability when none of the client has echo cancellation. For clients that do not control the audio capture and playback devices directly, it is RECOMMENDED to support echo cancellation between devices running at slight different sampling rates, such as when a webcam is used for microphone.

The client SHOULD allow either the entire AEC or the non-linear processing (NLP) to be turned off for applications, such as music, that do not behave well with the spectral attenuation methods typically used in NLPs. It SHOULD have the ability to detect the presence of a headset and disable echo cancellation.

For some applications where the remote client may not have an echo canceller, the local client MAY include a far-end echo canceller, but if that is the case, it SHOULD be disabled by default.

6. Legacy VoIP Interoperability

The codec requirements above will ensure, at a minimum, voice interoperability capabilities between WebRTC client applications and legacy phone systems.

7. IANA Considerations

This document makes no request of IANA.

Note to RFC Editor: this section may be removed on publication as an RFC.

8. Security Considerations

The codec requirements have no additional security considerations other than those captured in [I-D.ekr-security-considerations-for-rtc-web].

9. Acknowledgements

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

10. Normative References

[I-D.ekr-security-considerations-for-rtc-web]
Rescorla, E., "Security Considerations for RTC-Web", May 2011.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC3551] Schulzrinne, H. and S. Casner, "RTP Profile for Audio and Video Conferences with Minimal Control", STD 65, RFC 3551, July 2003.

[RFC4733] Schulzrinne, H. and T. Taylor, "RTP Payload for DTMF Digits, Telephony Tones, and Telephony Signals", RFC 4733, December 2006.

Authors' Addresses

Jean-Marc Valin
Mozilla
650 Castro Street
Mountain View, CA 94041
USA

Email: jmvalin@jmvalin.ca

Cary Bran
Plantronics
345 Encinial Street
Santa Cruz, CA 95060
USA

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