

Workgroup: Internet Engineering Task Force  
Internet-Draft:  
draft-buethe-opus-speech-coding-enhancement-00  
Updates: [6716](#) (if approved)  
Published: 23 October 2023  
Intended Status: Standards Track  
Expires: 25 April 2024  
Authors: J. Bueth, Ed.    J.-M. Valin  
          Amazon            Amazon

## **Integration of Speech Codec Enhancement Methods into the Opus Codec**

### **Abstract**

This document proposes a method for integrating a speech codec enhancement method into the Opus codec [[RFC6716](#)]

### **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 <https://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 25 April 2024.

### **Copyright Notice**

Copyright (c) 2023 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 (<https://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 Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

## Table of Contents

- [1. Introduction](#)
  - [1.1. Requirements Language](#)
- [2. An Illustrative Example](#)
- [3. Requirements for the Enhancement Method](#)
- [4. Requirements for Opus Decoder Integration](#)
- [5. Interoperability](#)
- [6. IANA Considerations](#)
- [7. Security Considerations](#)
- [8. References](#)
  - [8.1. Normative References](#)
  - [8.2. Informative References](#)
- [Authors' Addresses](#)

### 1. Introduction

Since the specification of the original Opus codec [[RFC6716](#)] new data-driven speech codec enhancement methods emerged which outperform classical enhancement methods by a large margin. This document proposes a method to integrate such enhancement methods into the Opus decoder including a set of requirements that ensure

- (1) consistent performance of the enhancement method itself,
- (2) preservation of decoder performance (e.g. seamless mode switching), and
- (3) preservation of basic interoperability when tuning the Opus encoder for use with the enhanced decoder.

The document furthermore contains a description of the linear-adaptive coding enhancer (LACE) and its integration into the Opus decoder as an illustrative example.

#### 1.1. Requirements Language

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 BCP 14 [[RFC2119](#)] [[RFC8174](#)] when, and only when, they appear in all capitals, as shown here.

### 2. An Illustrative Example

We use the linear-adaptive coding enhancer (LACE) [[lace-paper](#)] as an illustrative example to highlight the specific challenges of integrating a speech codec enhancement method into the Opus decoder. LACE is trained to enhance the output signal of the SILK decoder, the speech coding mode of Opus, and [Figure 1](#) depicts a high-level

overview of the Opus decoder with LACE added as an enhancement model.

The first requirement for a speech coding enhancement method concerns the performance of the method itself. In this example it relates to the question how the SILK decoder output compares to the LACE output. In [[lace-paper](#)] this has been evaluated on clean speech samples using a P.808 listening test as well as the objective method PESQ, which showed consistent improvement for all tested bitrates. For a general enhancement method it will be necessary to specify testing material and performance criteria to prevent unintended quality degradation of the Opus codec.

The second requirement concerns performance of the Opus decoder as a whole. Depending on the bitstream the decoder may have to perform mode switching, e.g. between SILK and CELT, or it may combine the SILK and CELT outputs when the codec operates in hybrid mode. Changes to the SILK output signal by an enhancement method, such as added delay, phase shifts, or level alterations can therefore negatively impact the performance of the Opus decoder even if the first requirement is met. LACE solves this problem by adding no delay and by being approximately phase and level preserving. However, since many enhancement methods are non causal and non phase preserving, these requirements may be too strict for a general enhancement method.

The third requirement concerns interoperability. The Opus specification provides significant freedom for tuning the encoder and the presence of an enhancement method in the decoder may change the optimal encoding choices significantly. In the present example encoding e.g. wideband content at 6 kb/s still leads to fair-to-good quality when using then LACE-enhanced decoder while the quality of a legacy decoder is significantly worse. To make full use of these new enhancement methods such encoder tunings should be allowed but basic interoperability with legacy decoders or other enhanced decoders needs to be ensured.

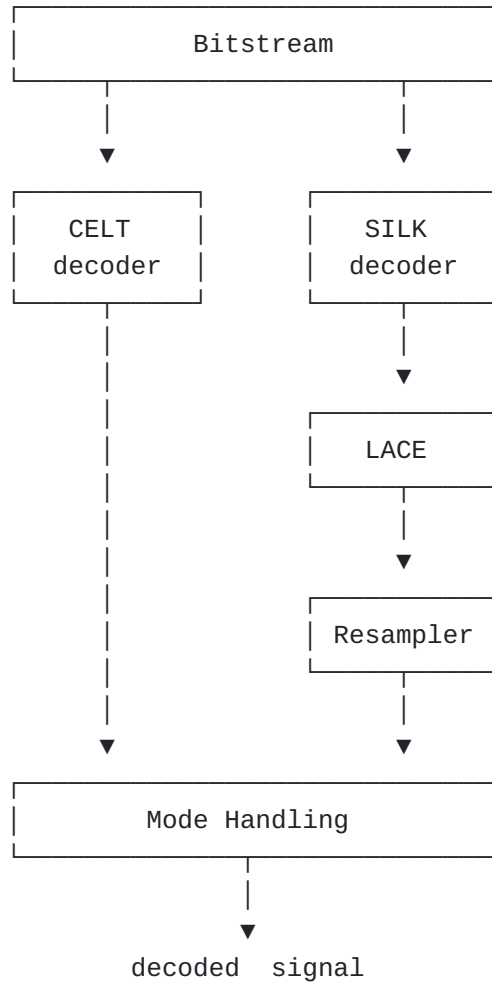


Figure 1: A simplified Opus decoder diagram including LACE as enhancement module

LACE has meanwhile been superseded by the Non-Linear adaptive coding enhancer (NoLACE) [[nolace-paper](#)] which shares all basic properties of LACE outlined above but provides higher quality. This stresses the advantage of specifying requirements for an enhancement method over specifying the method itself.

### 3. Requirements for the Enhancement Method

TBD

### 4. Requirements for Opus Decoder Integration

TBD

### 5. Interoperability

TBD

## 6. IANA Considerations

The decoder should be able to signal the presence of an enhancement method to the encoder over SDP. The exact mechanism is TBD and the following options are open for discussion.

- (1) update audio/opus media type registration [[RFC7587](#)] to include a parameter `speech_enhancement` with possible values 0 and 1
- (2) assign an extension ID, e.g. 33, from the registry defined in [[opus-extension](#)] to implement speech coding enhancement. This has the advantage of a double use, meaning the extension ID can both be used to signal the decoder capability to the encoder and for transmitting side information to guide a speech enhancement method from the encoder to the decoder. However, it needs to be proven that side information is useful.
- (3) update [[opus-extension](#)] to include extension IDs beyond 127 for data-less extensions

## 7. Security Considerations

TBD

## 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, <<https://www.rfc-editor.org/info/rfc2119>>.
- [[RFC8174](#)] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <<https://www.rfc-editor.org/info/rfc8174>>.
- [[RFC6716](#)] Valin, JM., Vos, K., and T. Terriberry, "Definition of the Opus Audio Codec", RFC 6716, DOI 10.17487/RFC6716, September 2012, <<https://www.rfc-editor.org/info/rfc6716>>.
- [[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, <<https://www.rfc-editor.org/info/rfc7587>>.
- [[opus-extension](#)] Valin, J.-M., "Extension Formatting for the Opus Codec (draft-valin-opus-extension)", April 2023.

## 8.2. Informative References

[lace-paper] Buethe, J., Valin, J.-M., and A. Mustafa, "LACE: A light-weight, causal Model for enhancing coded Speech through Adaptive Convolutions", 2023.

[nolace-paper] Buethe, J., Mustafa, A., Valin, J.-M., Helwani, K., and M. Goodwin, "NoLACE: Improving Low-Complexity Speech Codec Enhancement Through Adaptive Temporal Shaping", 2023.

### Authors' Addresses

Jan (editor)  
Amazon  
Germany

Email: [jbuethe@amazon.com](mailto:jbuethe@amazon.com)

Jean-Marc  
Amazon  
Canada

Email: [jmvalin@amazon.com](mailto:jmvalin@amazon.com)