Webex AI Codec

AI Transparency Technical Note

August 2024

Contents

At Cisco, we believe that Artificial Intelligence (AI) can be leveraged to power an inclusive future for all. We also recognize that by applying this technology, we have a responsibility to mitigate potential harm. That is why Cisco adheres to our Responsible AI Framework (the "Framework"), which is based on six principles of Transparency, Fairness, Accountability, Privacy, Security and Reliability. Cisco translates these principles into product development requirements, which ultimately form part of the product development lifecycle alongside our Security by Design, Privacy by Design, and Human Rights by Design processes.

Accordingly, Webex by Cisco (Webex) features that leverage AI are built with transparency, fairness, accountability, privacy, security, and reliability at their core. Each feature powered by AI undergoes an AI Impact (AII) Assessment – a best-in-class review of how the technical underpinnings of the functionality measure against the Framework Principles.

Webex AI Codec was built with the Framework Principles at the center of how we deliver the AI-powered technology. This Technical Note describes more information about Webex AI Codec and how Cisco responsibly leverages AI deliver the functionality.

webex by cisco

Feature Overview

Webex AI Codec is an AI-based speech enhancement and audio codec technology aimed at delivering high-quality noise-free speech under normal and challenging network conditions, for example packet loss, which may occur in real-time communication systems.

The Webex AI Codec consists of three main modules: an encoder, a quantizer, and a decoder. For any given audio input captured from a microphone (i.e. noisy, reverberated speech), the Webex AI Codec encoder removes noise and reduces reverberation, while simultaneously transforming the audio into a high dimensional latent representation referred to as the embedding space.

The quantizer compresses the denoised speech signal representation with a low bitrate, for example 6/3/2/1 kbps. The compressed speech representation is subsequently packetized and sent over the network. At the decoder, the Webex AI Codec reconstructs the speech signal and prepares it for sound reproduction, via a loudspeaker as an example.

Webex AI Codec is currently used by:

Cisco Webex Platform	Endpoints
Webex Meetings	Webex App
Webex Calling	Webex App to Webex App calls

Model Overview

Introduction

The Webex AI Codec model enhances speech (noise removal and de-reverberation) and performs coding at low bitrates. The model is used in conjunction with the Automatic Gain Control, Speech Level Estimator, and Deep REDundancy (DRED) algorithms to create a real-time communication system that is resilient to network impairments (i.e. packet loss). Automatic Gain Control ensures that the input signal is within the dynamic range in which the model performs optimally. The Speech Level Estimator calculates the signal power of the speech that the Webex AI Codec encoder codes. The DRED module packetizes the current audio frame with copies of the previous frames so that if a packet is lost, the Webex AI Codec can reconstruct the previous audio frames once the next packet is received, recovering audio that would have been lost otherwise.

In addition to the modules mentioned above, Webex AI Codec leverages other components normally used in real-time communication systems, such as jitter buffer and packet loss concealment techniques. Additional information on how Webex AI Codec can be used for the implementation of resilient communication systems can be referenced in the Webex AI Codec technical paper.

Model Architecture

The Webex AI Codec model architecture is inspired by modern neural speech codec techniques as described in the Webex AI Codec technical paper, including additional information on the training of the Webex AI Codec. The architecture, including a fully convolutional encoder, a residual vector quantizer, and a decoder is depicted below:
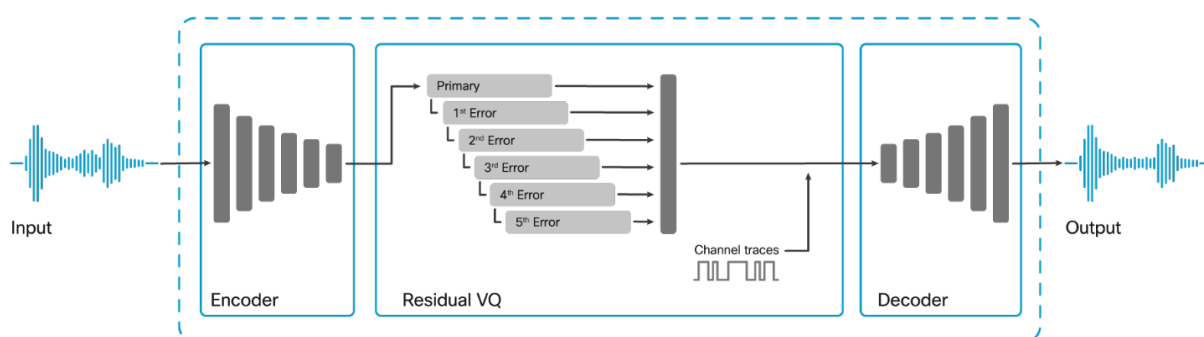

Figure 1 – Webex AI Codec architecture

Usage Guidelines

The Webex AI Codec, and derived features, cannot be used outside of Webex applications and Cisco devices. For the developer SDKs, application developers will need a Webex license, which includes the requisite permissions to develop custom applications.

Model Inputs and Outputs

The Webex AI Codec model handles audio data as input (i.e. noisy, reverberant speech) on the transmitter side and returns audio data as output (i.e. clean, de-reverberated speech) on the receiver side.

To perform speech enhancement and coding, the encoder converts the audio input into embeddings, for example, a vector representation of the audio data. A residual Vector Quantizer subsequently quantizes those embeddings and transmits the resulting codeword indexes over a communication channel to the decoder on the receiver end. The decoder uses the received information to synthesize high-fidelity output speech at its output.

webex by CISCO

Data Sources for Training and Evaluation

Webex AI Codec is trained with a mix of real and synthetic data curated to meet model acceptance criteria and maximize diversity of speakers (demographic & language) and acoustic environments. The data was either purchased, licensed, downloaded from publicly available sources, or responsibly crowdsourced. The final test sets were manually reviewed to verify correctness of the annotated metadata.

No Personal Identifiable Information (PII) was collected as part of the model training and evaluation, nor is it collected during the inference stage when the model is used to process audio data in real-time. The Webex AI Codec model is not trained or improved using data collected from the Webex end users.

Model Evaluation and Performance

The model performance is evaluated using both subjective and objective tests to assess audio quality and robustness of the system against challenging acoustic and network conditions. Subjective tests include A/B preference tests and Multiple Stimuli with Hidden Reference and Anchor (MUSHRA), ITU-R recommendation BS.1534 Method for the subjective assessment of intermediate quality level of audio systems. Objective tests include Perceptual Objective Listening Quality Analysis (POLQA) and Packet Loss Concealment Mean Opinion Scores (PLCMOS).

Performance evaluations verified that the Webex AI Codec enhances the user experience by effectively reconstructing audio that would otherwise be lost due to packet loss. This enhancement leads to fewer disruptions during conversations, thereby helping to boost productivity. However, these assessments also revealed the following corner cases that are sporadically triggered by challenging acoustic conditions:

- In high noise level environments, brief speech segments can undergo alteration, with temporary changes or distortions in the speech characteristics.
- In highly reverberant environments, the audio output from the codec might sound "metallic" or "robotic."
- When processing audio from simultaneous speakers in highly reverberant spaces, the codec might generate unintended audio artifacts and potentially truncate short speech segments.
- Users who use medical devices to generate speech, for example electrolarynx devices, should select "Music Mode" in the Webex App to reduce the occurrence of potential audio artifacts.

Cisco is committed to enhancing audio quality by minimizing these infrequent anomalies resulting from acoustically challenging scenarios that pose difficulties for any speech codec.

Safety and Ethical Considerations

Cisco assessed the Webex AI Codec and its use of AI against the Framework Principles through an AI Impact (AII) Assessment. Through that process, Cisco analyzed the use of the Webex AI Codec against safety and ethical considerations, including whether and to what extent the model mitigates bias, integrates fairness, and whether and to what extent the model may compromise human rights.

The underlying model was trained on diverse, balanced datasets that are representative of the relevant user base and subsequent feature development. For the Webex AI Codec, the relevant user base consists of end users of the Webex Meeting and Webex Calling platforms. The datasets used for the training and evaluation reflect the characteristics of our user base, including languages, accents, background noises, and acoustic conditions. The datasets are anonymized and manually reviewed for accuracy.

Webex also allows customers and users to enable and control use of the feature. Please contact Cisco if you have concerns or feedback about your experience with Webex AI Codec.

Fairness

Webex uses diverse datasets to train and develop the Webex AI Codec to address a wide range of demographics, languages, accents, noises, and acoustic conditions. Webex continuously evaluates datasets and labeling techniques for these attributes, and we expect the Webex AI Codec to perform similarly across Webex platforms and endpoints for the intended purpose and identified user base, regardless of demographics.

Privacy and Security

Webex addresses processing personal data, including retention periods, in its Privacy Data Sheets, found on the Cisco Trust Portal. The Webex AI Codec does not use personal data for training, inference, or evaluation purposes.

webex by CISCO

Updates and Maintenance

Cisco is not aware of any known biases associated with the Webex AI Codec or derived features. Cisco routinely monitors and rebalances the training dataset that powers the Webex AI Codec to address any issues that are brought to our attention. The improvements Cisco makes based on those identified cases, including objective and subjective scores that measure the accuracy and performance of the model, are well documented. Webex communicates ongoing significant product updates and improvements in user experience via updates to the product release documents, customer email notifications, and Help Center documentation.

License

The Webex AI Codec is proprietary and not publicly available. Cisco uses Webex AI Codec to implement a real-time communication system that is resilient to many adverse network conditions in the context of Cisco devices, Webex applications, or with Webex SDKs.

References

RAI Principles

RAI Framework

Cisco Trust Portal

DRED: Deep REDundancy Coding of Speech Using a Rate-Distortion-Optimized Variational Autoencoder

Webex AI Codec Technical Paper

BS.1534 Method for the subjective assessment of intermediate quality level of audio systems

Perceptual Objective Listening Quality Analysis (POLQA)

Packet Loss Concealment Mean Opinion Scores (PLCMOS)

webex by cisco