Codec 2

David Rowe

Revision: 112f3b5 on branch: dr-codec2-doc

November 19, 2023

1 Introduction

Codec 2 is an open source speech codec designed for communications quality speech between 700 and 3200 bit/s. The main application is low bandwidth HF/VHF digital radio. It fills a gap in open source voice codecs beneath 5000 bit/s and is released under the GNU Lesser General Public License (LGPL). It is written in C99 standard C.

The Codec 2 project was started in 2009 in response to the problem of closed source, patented, proprietary voice codecs in the sub-5 kbit/s range, in particular for use in the Amateur Radio service.

This document describes Codec 2 at two levels. Section 2 is a high level overview aimed at the Radio Amateur, while Section 3 contains a more detailed description with math and signal processing theory. This document is not a concise algorithmic description, instead the algorithm is defined by the reference C99 source code and automated tests (ctests).

This production of this document was kindly supported by an ARDC grant [1]. As an open source project, many people have contributed to Codec 2 over the years - we deeply appreciate all of your support.

2 Codec 2 for the Radio Amateur

2.1 Model Based Speech Coding

A speech codec takes speech samples from an A/D converter (e.g. 16 bit samples at an 8 kHz or 128 kbits/s) and compresses them down to a low bit rate that can be more easily sent over a narrow bandwidth channel (700 bits/s). Speech coding is the art of "what can we throw away". We need to lower the bit rate of the speech while retaining intelligible speech, and making it sound as natural as possible.

As such low bit rates we use a speech production model. The input speech is anlaysed, and we extract model parameters, which are then sent over the channel. An example of a model based parameter is the pitch of the person speaking. We estimate the pitch of the speaker, quantise it to a 7 bit number, and send that over the channel every 20ms.

The model based approach used by Codec 2 allows high compression, with some trade offs such as noticeable artefacts in the decoded speech. Higher bit rate codecs (above 5000 bit/s), such as those use for mobile telephony or voice on the Internet, tend to pay more attention to preserving the speech waveform, or use a hybrid approach of waveform and model based techniques.

Recently, machine learning has been applied to speech coding. This technology promises high quality, artefact free speech quality at low bit rates, but currently (2023) requires significantly more memory and CPU than traditional speech coding technology such as Codec 2. However the field is progressing rapidly, and with the progress of Moore's law will soon be a viable technology for many low bit rate speech applications.

- 2.2 Speech in Time and Frequency
- 2.3 Sinusoidal Speech Coding
- 2.4 Spectral Magnitude Quantisation
- 2.5 Bit Allocation
- 3 Signal Processing Details

4 Further Work

1. How to use tools to single step through codec operation

[2]

References

- Enhancing HF Digital Voice with FreeDV, 2023. https://www.ardc.net/apply/grants/2023-grants/ enhancing-hf-digital-voice-with-freedv/.
- [2] Daniel W Griffin and Jae S Lim. Multiband excitation vocoder. *IEEE Transactions on acoustics, speech, and signal processing*, 36(8):1223–1235, 1988.

Figure 1: A 40ms segment of the word "these" from a female speaker, sampled at 8 kHz. The waveform repeats itself every 4.3ms (230 Hz), this is the "pitch period" of this segment.

