Codec 2 – Open source speech coding at 2400 bit/s and below

David Rowe, VK5DGR
Rowetel
9 Nelson Avenue
Flinders Park, 5025
SOUTH AUSTRALIA
david@rowetel.com

Abstract

Codec 2 is an open source, low bit rate codec for speech over HF/VHF digital radio. Most low bit rate codecs are proprietary, closed source, and require licensing fees. Codec 2 is unique in that it is open source, allowing experimentation and modification. It fills a gap in open source, free-as-in-speech voice codecs beneath 5000 bit/s and is released under the GNU Lesser General Public License (LGPL).

This paper describes the motivation for Codec 2, status, and future plans. The Codec algorithm is described using figures, avoiding high level maths and Digital signal Processing (DSP) theory.

Key words: low bit rate speech coding, open source, patent free

Motivation and History

Codec 2 [3] was inspired by a need for an open source, patent free speech codec for modern digital modes like DSTAR [1]. An experimental service like Ham radio needs a codec that Hams can understand, explore, modify and experiment with. An open source codec is license and royalty free, and can therefore be incorporated at zero cost in modern Software Defined Radio (SDR) designs.

In contrast proprietary codecs come in hardware chip form or require software licensing schemes. They cannot be modified, understanding of how they work is discouraged, and modification is forbidden.

The author has a background in speech codec design, development, and implementation. A baseline algorithm was developed in the 1990s [2] has been significantly developed over the past two years into the current alpha version of Codec 2 [3], operating at 2550 bit/s.

Patents and Speech Codecs

The authors of proprietary and patented codecs have borrowed heavily from the public domain. Perhaps 5% of the algorithms they use are original and patented, the remaining 95% of the algorithms in these codecs are public domain algorithms. To build a codec with equivalent performance, we simply need alternatives for the 5% that are patented. Unfortunately this means an open source codec is unlikely to be compatible with a proprietary codec.

Speech Coding 101

Figure 1 illustrates how Codec 2 fits into a typical digital radio system.

The codec converts a speech signal sampled at 8 kHz to a 2400 bit/s bit stream for formatting and transmission over the radio channel. The goal of speech coding is to throw away as much of the speech signal as possible while retaining intelligible and natural sounding speech. Codec 2 uses a model based speech coding approach. Instead of sending the speech waveform, model parameters are extracted from the input speech, and these model parameters are transmitted to the decoder. To track the evolving

speech signal the model parameters are updated regularly, for example every 20ms.

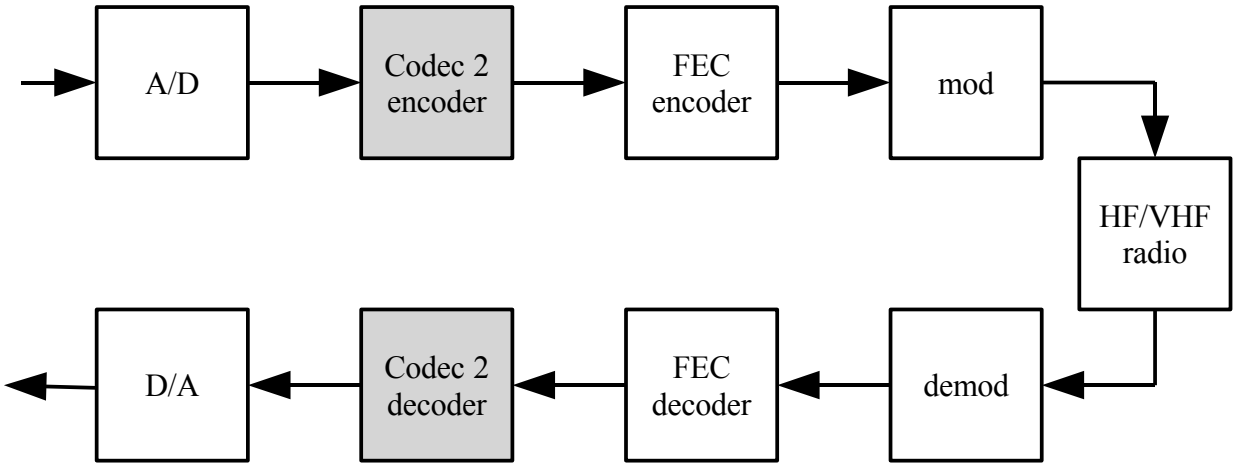

Figure 1: Codec 2 and a Digital Radio System

A typical model parameter is the pitch, which for humans is in the range of 50 to 500 Hz. This can be quantised with about 7 bits. If the pitch is updated every 20ms, the bit rate for the pitch information is 7/20ms = 350 bit/s.

Sinusoidal Speech Coding

Figure 2 show is a plot of 20ms of a female vowel signal. The signal is periodic, i.e. there is a regular structure to the waveform that repeats itself across the frame. The rate which the waveform repeats is called the pitch which in this example has a period of 4.3ms or a frequency of 230Hz. The amplitude of the waveform is also slowly increasing across the segment.

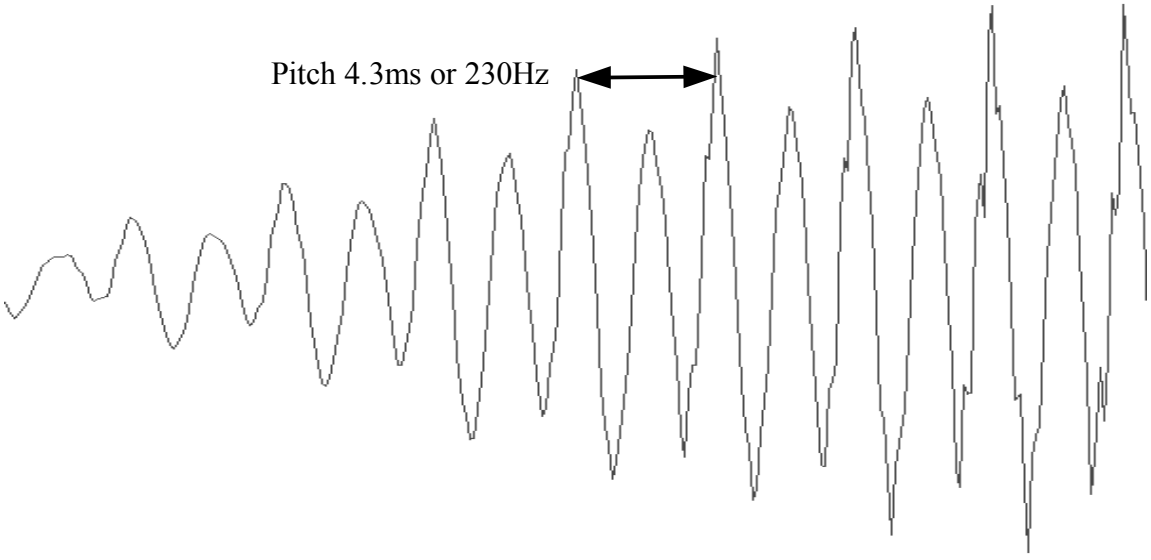

Figure 2: 20ms sample of female speech

Figure 3 is the spectrum of the signal in Figure 2, i.e. a frequency domain version of the 20ms time domain speech segment. The horizontal axis is 0 to 4000Hz, and the vertical axis is in dB. Because the signal is periodic in time it is also periodic in frequency. It has a comb-like structure, with each spike

representing a harmonic of the pitch. Each of the spikes can be viewed as a sine wave oscillator, with it's own frequency, phase, and amplitude. In the Figure 3 there are 16 spikes, so we could model the signal with 16 sine wave oscillators.

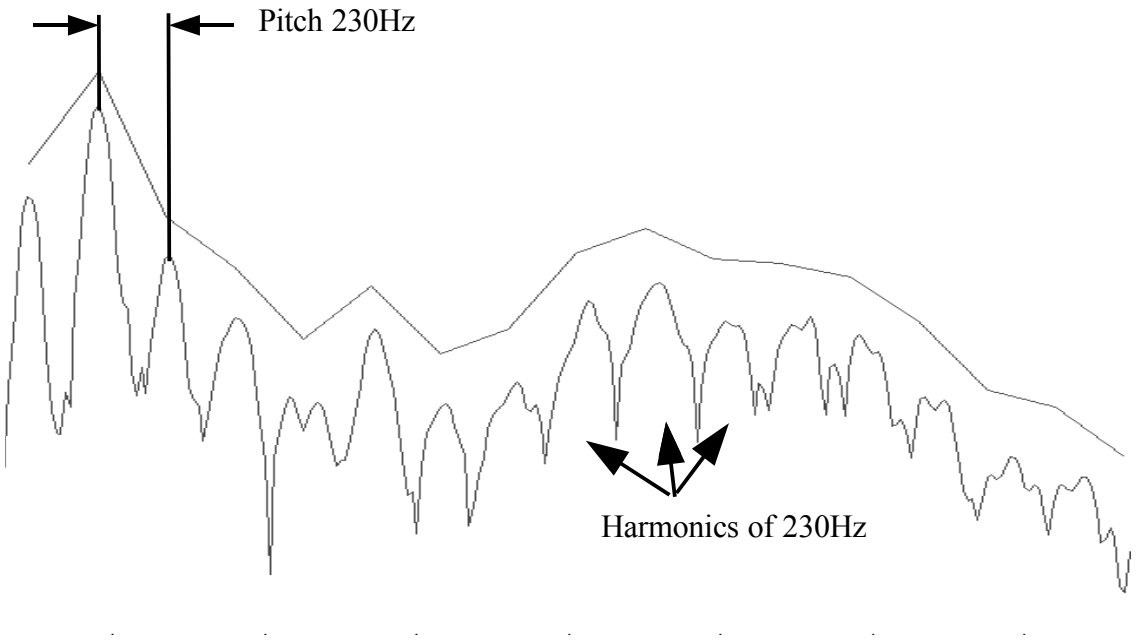

Figure 3: 20ms sample of female speech

This leads us to the sinusoidal model of speech in Figure 4. To synthesise speech, we sum the output from a bank of sine wave oscillators. The frequency of each oscillator is a harmonic of the pitch, for example 230Hz, 460Hz, 690Hz, etc. The frequency, phase, and amplitude change over time and are updated every 20ms.

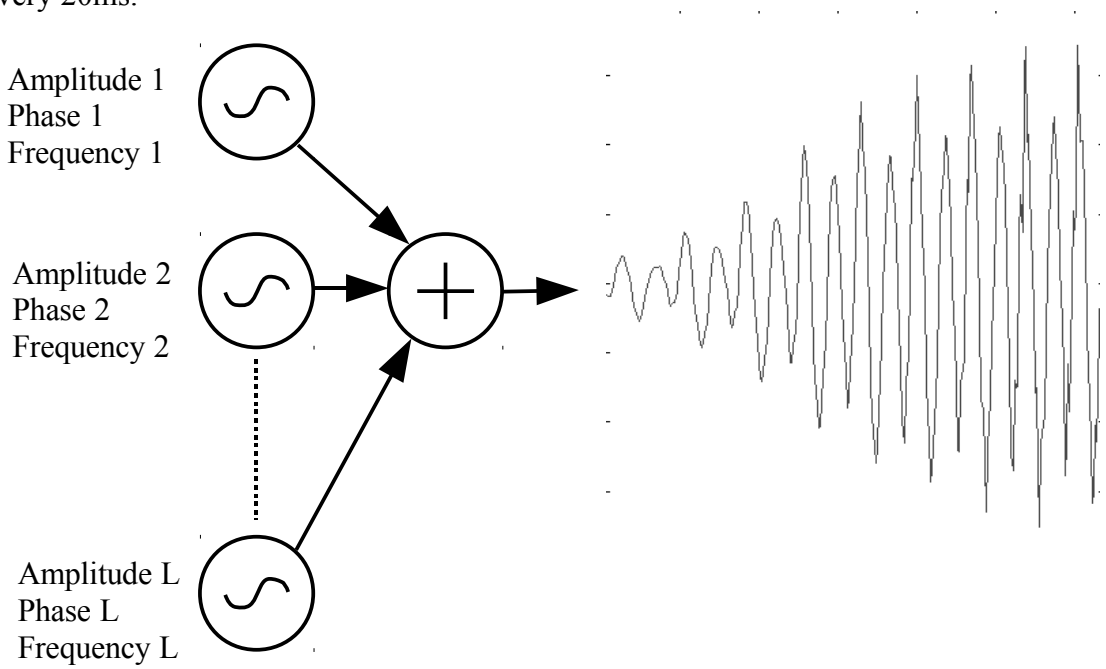

Figure 4: Sinusoidal Speech Model

In practice, the amplitudes of adjacent oscillators tend to be closely related (solid line at top of Figure 3), which leads to coding efficiencies. At low bit rates, there are not enough bits available to send the phase of each oscillator so we synthesise them at the decoder using a rule based approach.

Figure 5 is the block diagram of the Codec 2 Encoder. The Pitch Estimator is an algorithm that measures the pitch of the current frame. This is then quantised and sent to the decoder. A Fast Fourier Transform (FFT) converts the time domain samples to a frequency domain signal. This information is then used to determine if the frame is voiced or unvoiced. Voiced speech (like vowels) has a regular structure with continuous phase between frames. Unvoiced speech (e.g. consonants) is synthesised using random phases. A single bit (voicing) is sent to the decoder.

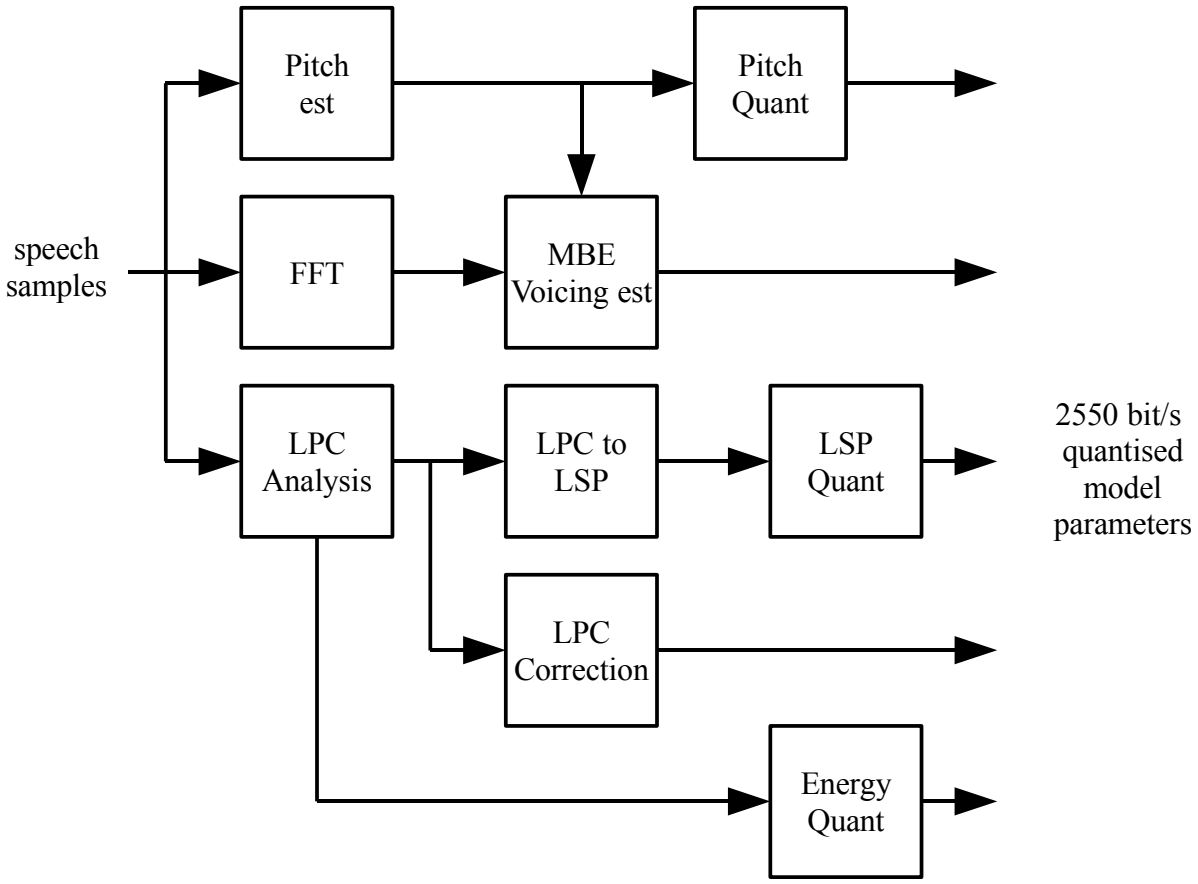

Figure 5: Codec 2 Encoder

Parameter	**Bits/frame**
Sine wave amplitudes (encoded as LSPs)	36
Low frequency LPC correction	1
Energy	5
Voicing (updated every 10ms)	2
Pitch	7
Total	51

Table 1: Codec 2 Bit Allocation

Linear Predictive Coding (LPC) is used to represent the sine wave amplitudes. LPC models use a fixed number of parameters which makes quantisation and transmission simpler compared to a time varying number of amplitude samples. The LPCs are converted to Line Spectrum Pairs (LSPs) for transmission over the channel. The energy of each frame is transmitted using a log quantiser.

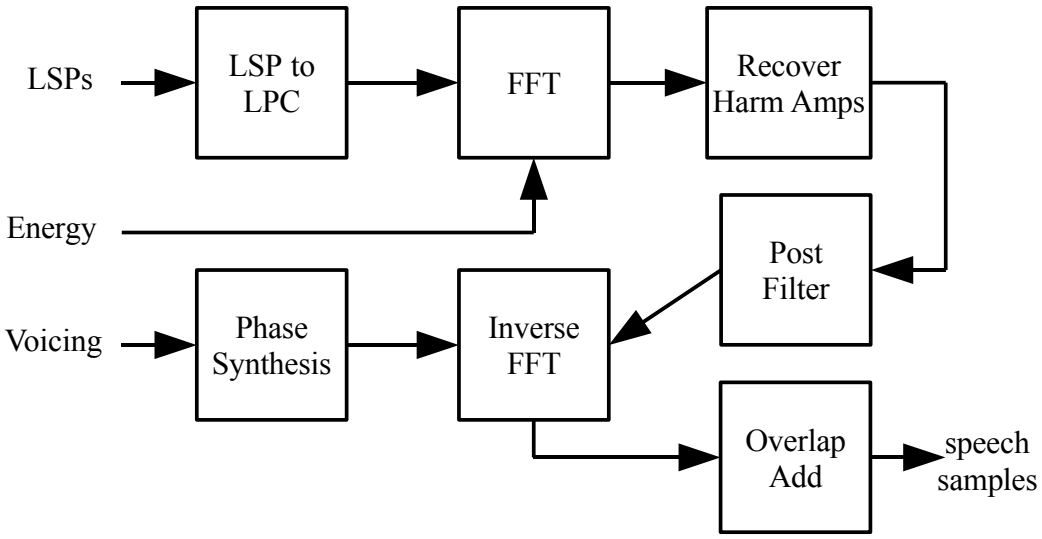

Figure 6: Codec 2 Decoder

Figure 6 illustrates the Codec 2 decoder. The harmonic sinusoidal amplitudes are recovered from the LSPs. The phase of each oscillator is synthesised at the decoder using a rule based approach. For voiced speech the phase of each oscillator is continuous from frame to frame. For unvoiced frames the phase is random, which creates a noise like signal. Finally the time domain output speech is created using an inverse FFT and overlapped with previous frames to create a continuous output signal.

One problem with model based speech coding is a tendency to break down for non-speech signals like background noise. For example engine noise may seem strange when passed through the codec. A Post Filter has been developed that improves the background noise performance of Codec 2. The Post Filter monitors the background noise level and randomises the phase of any harmonic beneath this level.

Status and Future Work

In late 2010 an Alpha 2550 bit/s version of Codec 2 was released and has been used for several experimental over the air and VOIP calls. Much can be done to improve the algorithm and lower the bit rate. Future work includes a better phase and voicing model to improve speech quality, 2400 and 1200 bit/s versions, inclusion of FEC and non-redundant error correction, and integration with modem software to build a complete HF and VHF open source digital communications system.

References

[1] Perens, Bruce, http://codec2.org/historic/

[2] Rowe, David, PhD Thesis, “Techniques for Harmonic Sinusoidal Coding”, http://www.itr.unisa.edu.au/~steven/thesis/dgr.pdf

[3] Rowe, David, “Codec 2 Home Page”, http://www.rowetel.com/codec2.html