Codec2: An Open Future for Digital Voice

- David Rowe VK5DGR
- david@rowetel.com
- Codec developer.

Today's Presenter

- Bruce Perens K6BP
- bruce@perens.com
- Phone 510-DX4-K6BP
- Evangelism and systems programming.

Let's Start With The Demo

Male voice sampled at 8 Khz, 16 bits per sample, 3.75 seconds recorded.

- Raw, 48000 bytes.
- Encoded with Codec2, 1050 bytes encoded.

Both of these are being played from .wav files.

A Gift From Asia

- When JARL designed D-STAR and ICOM productized it, they used what was available. There wasn't a good Open codec, so they used AMBE+.
- It's not their fault. They used what they could get, and there wasn't much understanding of “open” at the time, and Open Source wasn't so clear a success as it is now.
- But the decision left us with some problems.

Tune In The World of Constraint

- The AMBE+ voice codec, used for D-STAR digital voice, is a proprietary, trade secret algorithm, and some aspects of it are patented. This means that Radio Amateurs can't interoperate with D-STAR voice without using an AMBE+ chip, a black box available from a single vendor, Digital Voice Systems Inc. [DVSI].
- We're *constrained* by AMBE+'s intellectual property protection to buy and use AMBE+ chips if we are to interoperate with D-STAR voice.

Limitations

- As a result, there are severe limitations on what we would otherwise *expect* to be able to do on ham radio.
- The dominant working paradigm today for digital operation in Amateur Radio is to build a software-only implementation from *input* to *modulation*, all in Open Source.
- But that won't interoperate with D-STAR.

A Dongle on Your Back

- The only solution is to use the DV-Dongle, an expensive kludge that puts *black-box* software, in the form of a programmed TI DSP chip running the AMBE+ codec with its *read-protection* fuse blown in the system.

AMBE+ Codec Chip

- But this AMBE+ chip is much less powerful than the CPU in a modern computer. Its only use is to hide an algorithm in a black box.
- DV-Dongle lists costs more than $100/unit, just to run a *subroutine.*
- None of this represents a desirable future for Amateur radio. It's time for us to take control of our digital voice future.

An Inconvenient Secret

- There are legal problems with the use of *unspecified digital codes* like AMBE+ on Amateur Radio, under 97.309(b). As far as I can tell, international communication using D-STAR is illegal unless an agreement between the two nations permits the use of unspecified codes or this particular code, and I know of no such agreements.
- Some say that because D-STAR is a voice, not data, mode, it's not covered by 97.309(b). But in this case, voice is *indisputably* digital data.

The French Disconnection

- The unspecified codec is one reason that the French government recently banned D-STAR.
- Their other reason is its capability to connect to internet. Is this a historical provision to prevent competition with phone companies, like the ones we used to have in Part 97, or is it an issue of admitting third party traffic from unuathorized operators?

QSL Matey! Pieces of Eight!

- Some of the most popular codecs used on HF digital voice today are *pirated software.*
- There are *patent thickets* around the codecs used on HF that will prevent us from ever getting legal copies.
- The operators using those codecs risk losing their license *for a lifetime,* through FCC's “character” rules.

Patently Absurd

- There are several companies that aggressively prosecute codec patents. This was especially clear in the creation of the HTML5 standard, when companies asserted that there were patents covering the Ogg Theora video format, but none of them would actually disclose what the patents were.

Secret Patents

- The threats are from companies that own credible patent portfolios, but they *never specify an actual patent number* or collection of numbers, because once they did, the *Doctrine of Laches* would force them to sue within a few years or abandon the chance to do so. That this successfully defeated the creation of an open video codec standard is troublesome.

No Sufficient Open Codecs

- While Speex was proposed for the task, it's not a good low-rate or low-latency codec.
- We'd have to make something new.
- It takes rocket science.
- TAPR, while hardly short of rocket scientists, didn't have the right one for this job.

The Project

- I proposed a *Codec2* project to solve the problem by creating a technically acceptable voice codec in an Open Source implementation, and to eventually bring its bit-stream and algorithm to an open standard.
- Performance comparable to AMBE+ was required.
- It was a goal that the codec not be encumbered by valid, unexpired patents.

David Rowe

- Jean-Marc Valin, the main author of Speex, introduced me to David Rowe VK5DGR, an Australian Open Source developer with a Ph.D. in Speech Coding. David had previously written an Open Source line echo canceler, and created an Open Hardware PBX that is commercially manufactured.
- David had the chops, and would have done the coding for a reasonable fee. But I wasn't able to raise that fee in the depth of the economic downturn.

David's Previous Work

- David had built some of the first real-time speech codecs in the late '80's, on DSP chips. In his 1999 thesis, he created a demonstrable codec upon which today's Codec2 is based.
- David's web site is http://rowetel.com/
- Today, David develops Open Source and Open Hardware full time, pursuing various grants to create and deploy communications technology.

Mr. Mesh-Potato Head

- David has created the “Mesh Potato”: a WiFi mesh networking telephony device, and a commercially-manufactured Open Hardware PBX design: the *IP04.*
- He did his own electric vehicle conversion, too.
- So, he's a lot like the very best people I've met through TAPR.

Some Great Good Luck

- Fortunately, David became re-interested in Amateur Radio after a 25-year respite, and decided to go ahead with the coding, gratis.
- David's gotten some donations through his web site, but those come in AUD$10 chunks.

Compression

- The job of this codec is not just *encoding* voice, it's data *compression*.
- We're not talking about the kind of compression that makes contesters sound louder.
- What we mean is conveying information in smaller bandwidth than it would otherwise take, through the elimination of *redundant* parts of that information.

Since Time Immemorial

- Compression has been employed since the early days of wire telegraphy. Commercial telegrams often used a code-book of 5-digit numbers to represent common phrases used in business. That was compression.
- The arrangement of Morse Code to communicate the most frequently used letters in the English language, in the smallest possible number of signals, is compression

Analog Compression

- SSB eliminates the redundant sideband of AM modulation, and the carrier. That's compression.

What's the Job?

- So, how do you compress *audio?* Well, it turns out that there are lots of ways to do that, but our job *isn't* to compress audio. We need to compress *voice.*
- Voice is to audio as a clarinet is to an orchestra. There are fewer sounds that a clarinet can produce, compared to the orchestra. If we can just *encode* everything the clarinet can do, that information should be a lot smaller than encoding an orchestra.

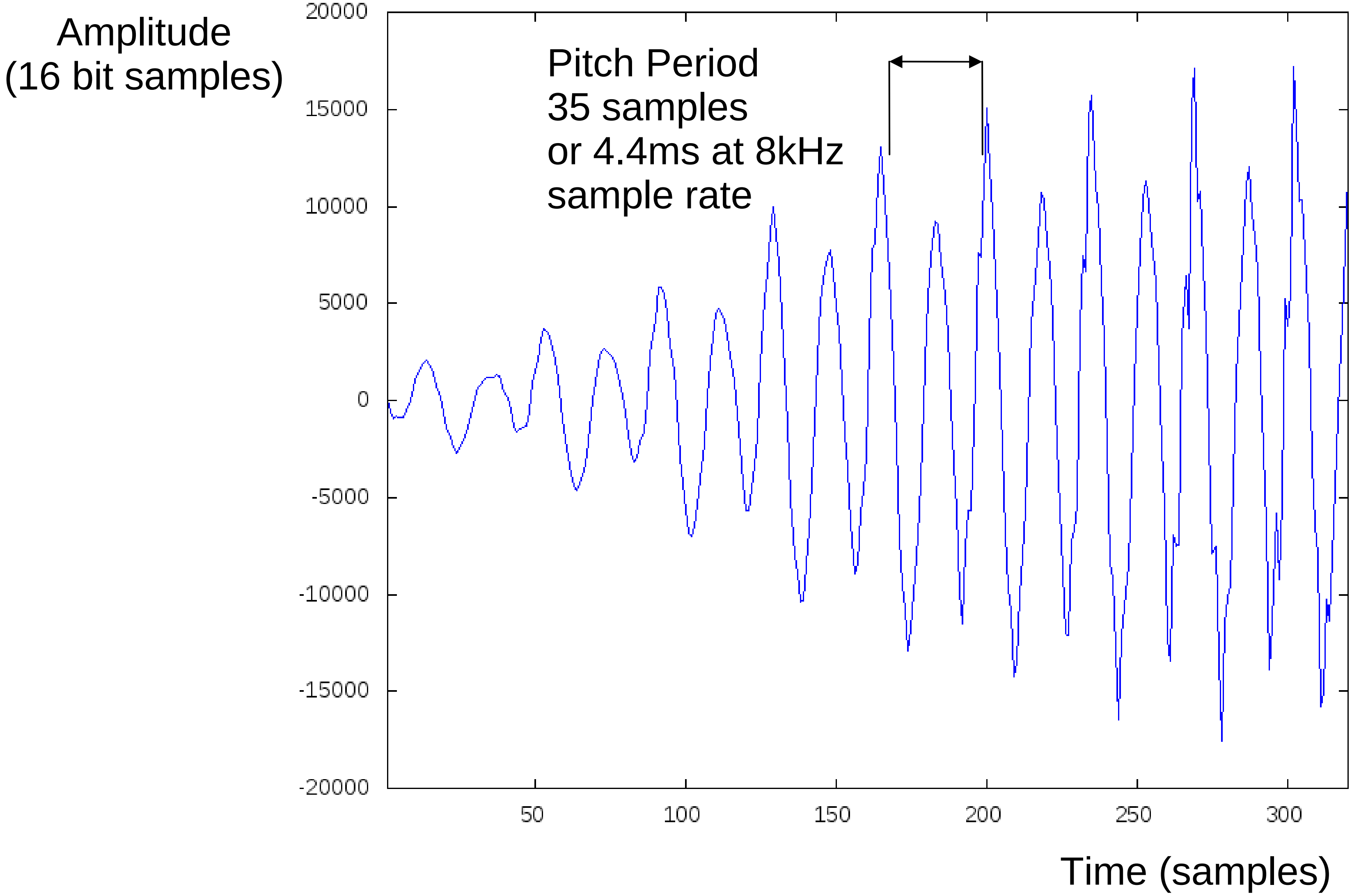
Nurturing Your Inner Clarinet

- And it turns out that your voice is a lot like that clarinet! The physical model of the human voice tract is a buzzer at the end of a pipe.
- That buzzer, your vocal chords, generates harmonic-rich buzzing, which is modulated as the shape of the pipe, your throat, changes in time. The resonant frequencies of that tube are called *formants,* a group of frequency bands, spaced about 230 Hz apart.

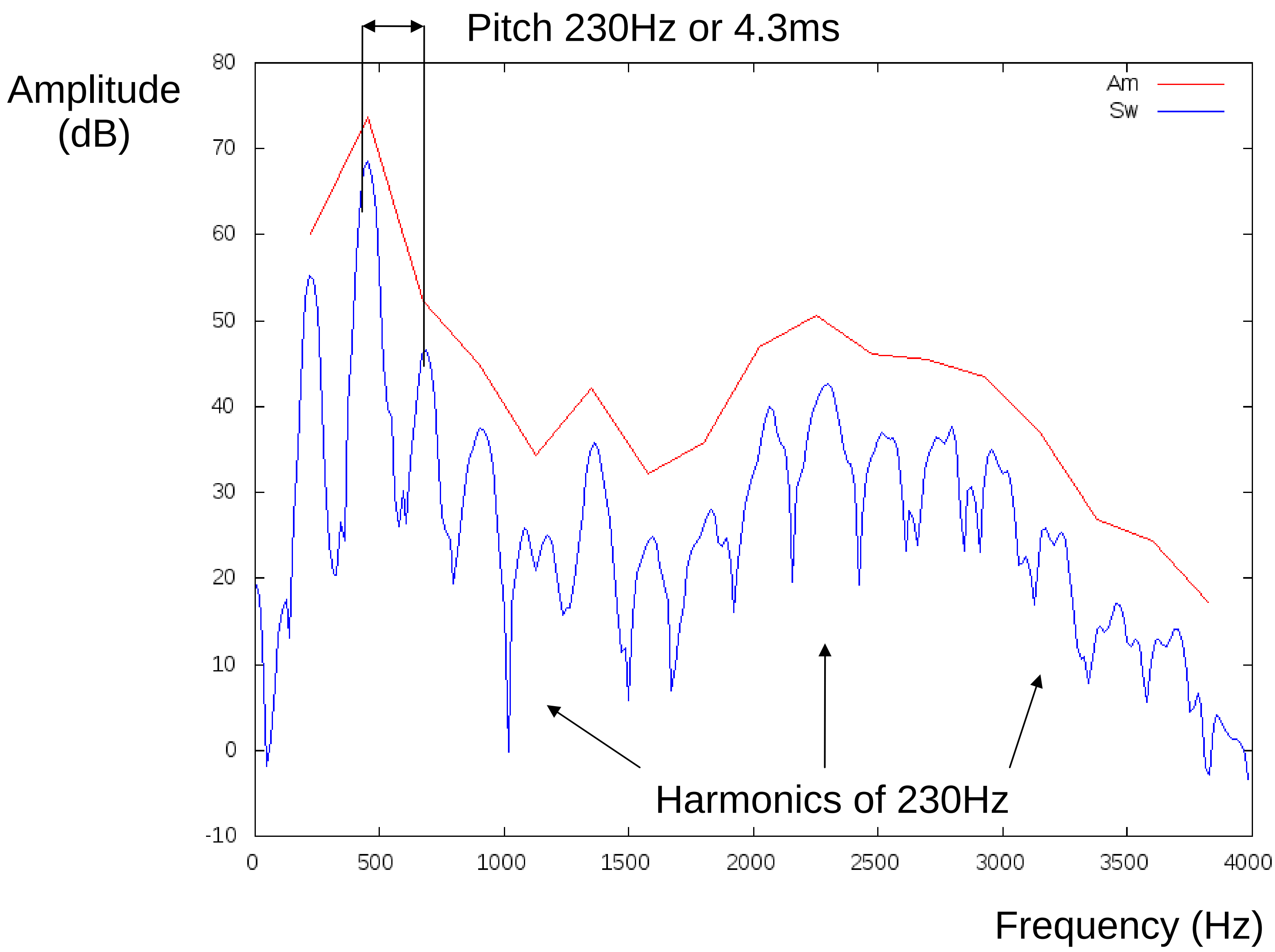
Get Sinusoidal!

- So we start with *Sinusoidal Speech Coding* to encode those formants.
- And here's where I start showing David's slides, and reading his notes.

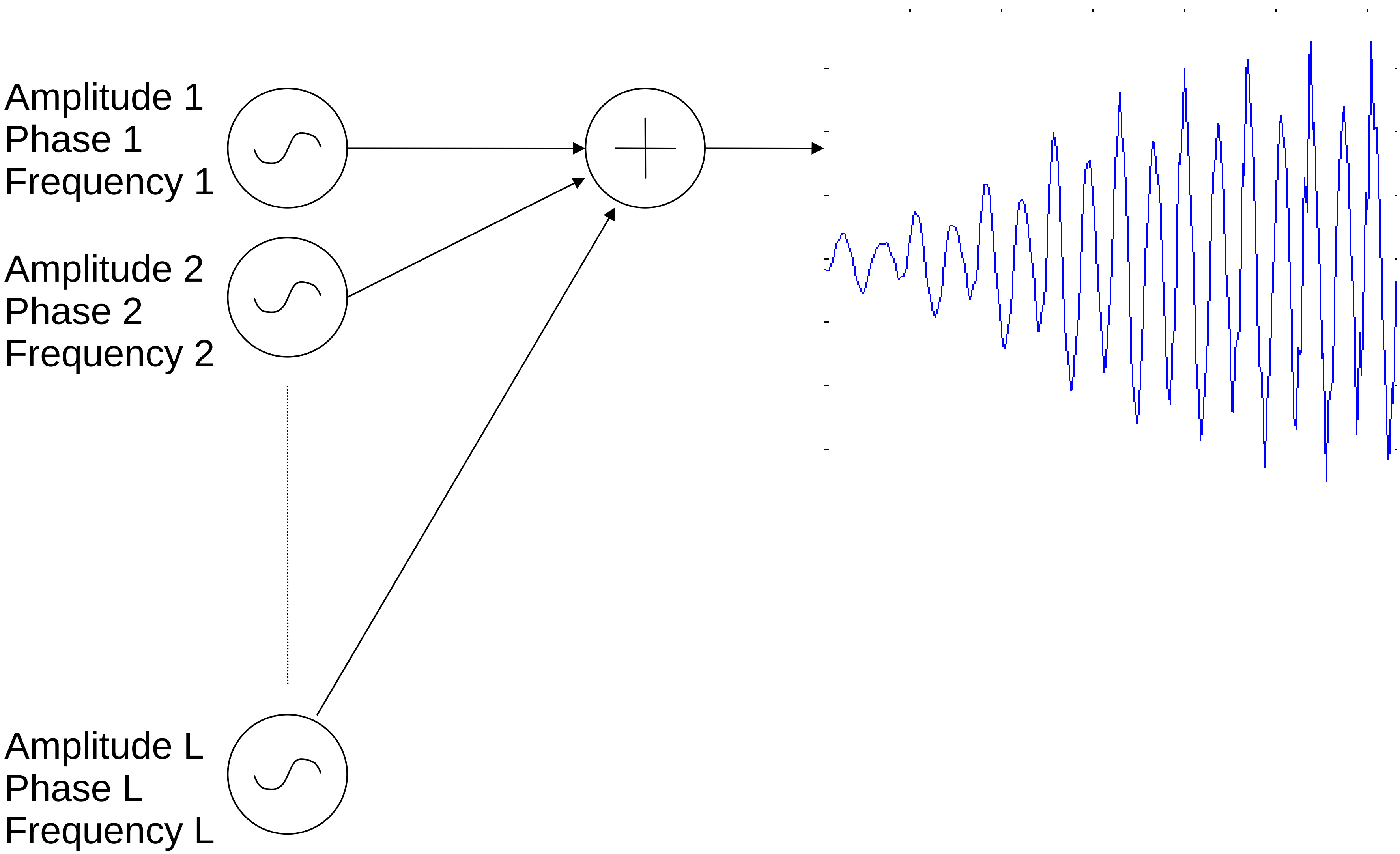
Sinusoidal Speech Coding



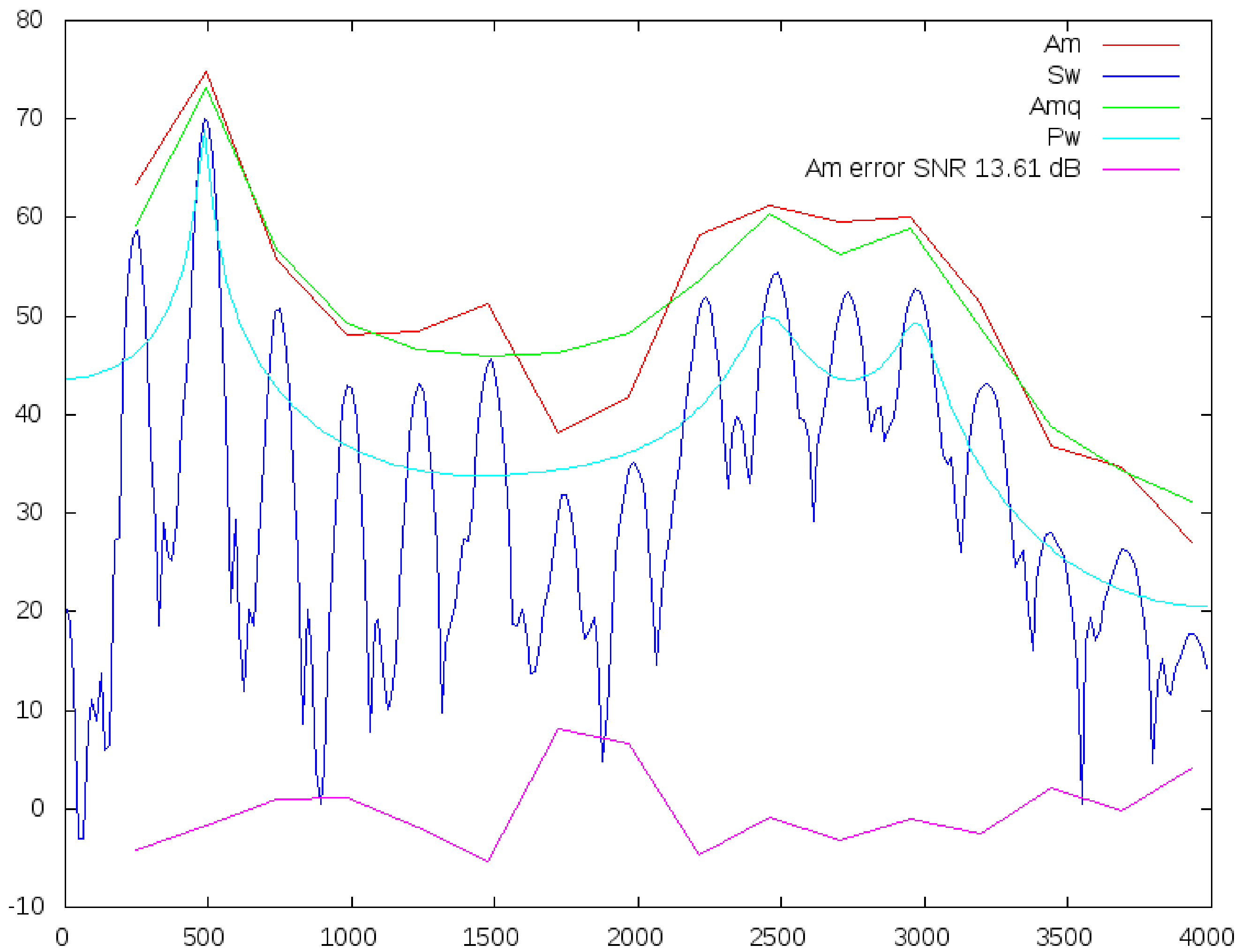
Sinusoidal Speech Coding



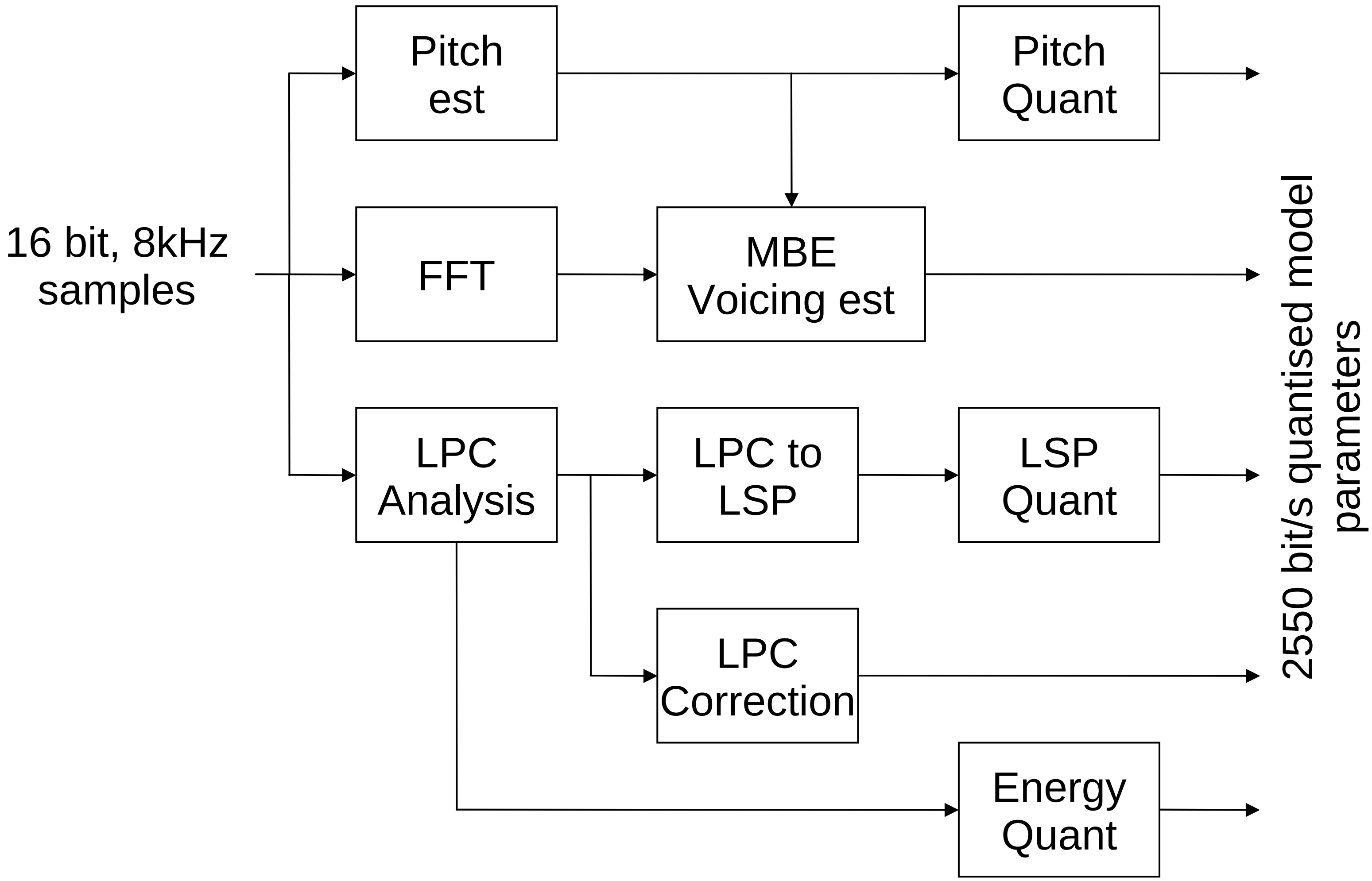
Sinusoidal Speech Model



Amplitude Modelling



Encoder Block Diagram

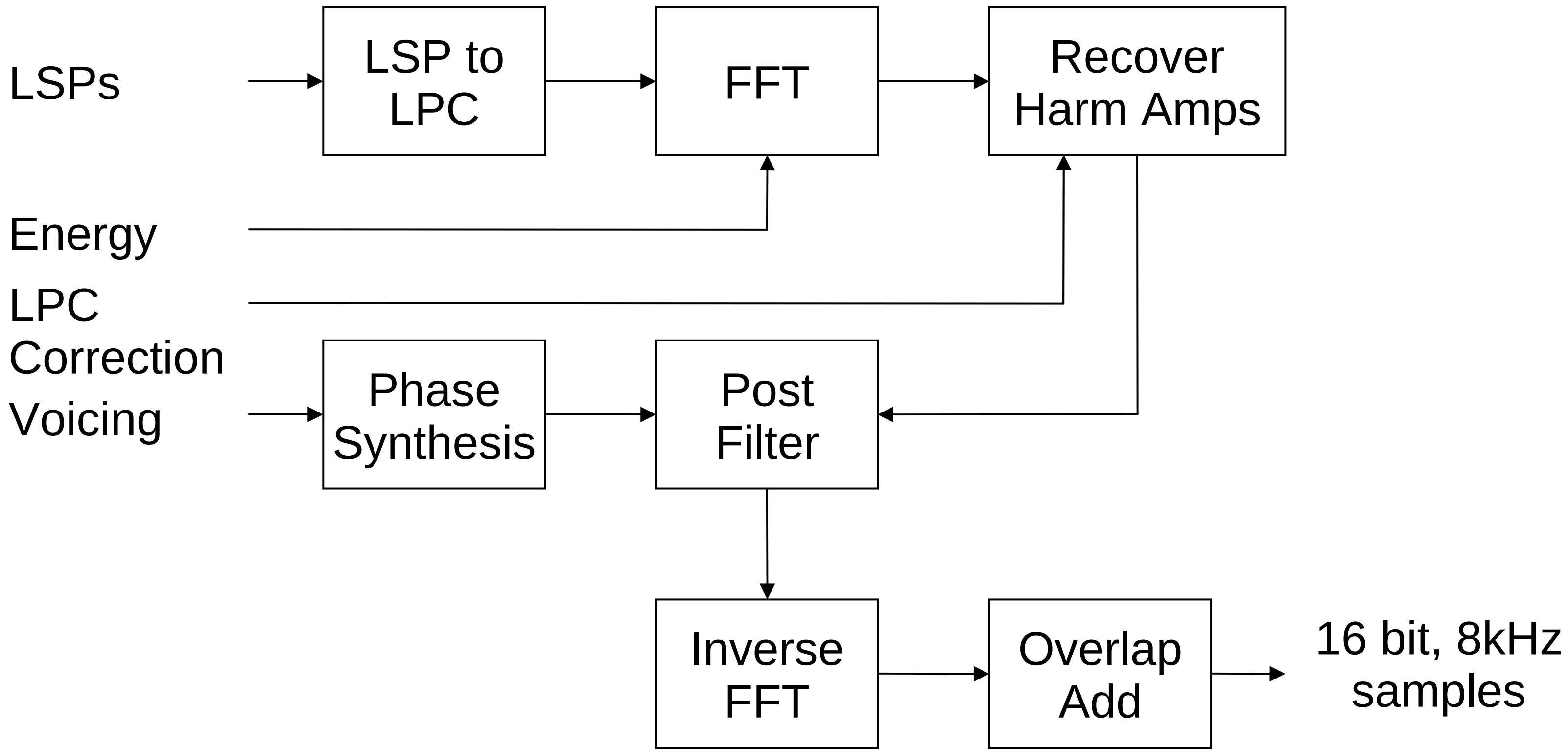


Bit Allocation

- Alpha V0.1 codec, subject to rapid change
- 51 bits per 20ms frame, or 2550 bit/s

Parameter	Bits/frame
Spectral magnitudes (LSPs)	36
Low frequency LPC correction	1
Energy	5
Voicing (updated each 10ms)	2
Pitch	7
Total	51

Decoder Block Diagram



Prior Art Summary

- Sinusoidal Coding, Mcaulay & Quatieri, 1984
- Linear Predictive Coding, Makhoul, 1975
- Line Spectrum Pairs, Itakura, 1975
- MBE Voicing, Griffin & Lim, 1988
- Overlap Add, Tribolet & Crochiere, 1979
- NLP Pitch Estimation, Rowe, 1999
- LPC Amplitude Recovery (algorithm used here), Rowe, 1991, 1999, 2009
- Post Filter, Rowe, 2009

Further Work

- Better phase model and voicing estimator
- Toll quality at 2000 bit/s
- Lower bit rate, 2400, 1200 bit/s
- Better background noise performance
- FEC and non-redundant error correction
- Integration with modem and test over radio channels
- Fixed point and DSP chip implementation

Issues

- This is only a voice codec. It's not supposed to handle music.
- Background noise is an issue for all voice codecs, and is a big problem for emergency communications.
- Best way to address is software noise reduction at the encoder.

Early Stage Implementation Plan, Internet

- Library-ize the code. Currently loads tables at run time, that needs to be fixed. Make interface look like speex?
- Dumb demo programs.
- Add Codec2 to *mumble* client. Mumble is a full-duplex speakephone group conference system with low latency, developed for gamerz but used by CW operators, etc.

Early Stage Implementation Plan, Radio

- Standardize a way to identify Codec2 in D-STAR data, so that devices can switch automatically. Must encode version numbers, as the codec is still in development.
- Run as D-STAR data using unmodified D-STAR Radios and all of the various D-STAR access points and software-only implementations.
- HF voice. Can we get the FDMDV modem opened or must we write our own?

Middle Stage

- Clone ICOM digital voice daughter cards, UT-122 etc., to hold both DVSI and Open codec chip, and switch appropriately.
- Doesn't address IC-92AD, which has AMBE+ on main board.
- Build codec gateways in D-STAR repeaters.
- Can we get ICOM interested, or must we do this without them? Other vendors?
- Will DVSI attempt to play hardball with ICOM, others?

Later Stage

- D-STAR is not the VHF/UHF data-link layer we'd build today.
- Open Hardware is making tremendous progress, provides the non-RF part of the platform today.
- Open, programmable platform in HT, no hardware codec, SDR for modulation and demodulation.
- IPV6, unique global IP encoding callsign (Naoto Shimazaki, 1998)
- Spread spectrum?