

# Analog Transmission of Vocoder Features over Radio Channels

March 28, 2024

Given a vector of vocoder features  $\mathbf{f}$ , use an autoencoder  $E$  to map them to a dimension  $d$  latent vector  $\mathbf{z}$  where  $d$  is even. Unlike digital modulation, each element  $z_i$  of  $\mathbf{z}$  is continuously valued and not constrained to a discrete set of points. For bandwidth efficient transmission over the channel the elements of  $\mathbf{z}$  are mapped to  $d/2$  complex symbols  $\mathbf{q}$ . Compared to classical digital modulation, the elements of  $\mathbf{z}$  can be considered BPSK symbols (continuously valued, analog bits), and the elements of  $\mathbf{q}$  analog QPSK symbols.

Our goal is to determine if reasonable speech quality can be obtained over a channel of bandwidth  $B < 3000$  Hz and SNR (measured in  $B = 3000$  Hz) of between 0 and 6dB, roughly the lower limit of Single Side Band (SSB) - a common power and bandwidth efficient form of analog radio communication.

## 1 Simulation of AWGN Channels

The autoencoder output  $\mathbf{z}$  is updated every  $T_z = 1/R_z$  seconds, giving a BPSK symbol rate of:

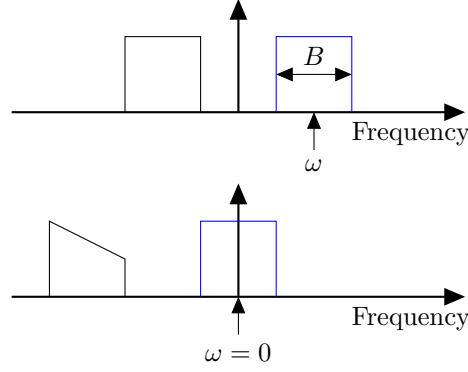
$$R_b = d/T_z \tag{1}$$

For example with  $T_z = 0.04$ ,  $d = 80$ ,  $R_b = 2000$  symbols/s. The QPSK symbol rate is given by:

$$R_q = \frac{d}{2T_z} \tag{2}$$

For example with  $T_z = 0.04$ ,  $d = 80$ ,  $R_q = 1000$  symbols/s.

Figure 1: Real sampled off-air signal. We are interested in the blue bandpass interval of bandwidth  $B$ , which is single sided and hence complex valued. After shifting to baseband, it's power is unchanged, and it remains complex valued.



We wish to simulate an AWGN channel with a user-defined  $E_b/N_0$ , where  $E_b$  is the energy of each BPSK symbol, and  $N_0$  is the noise power per unit bandwidth. Consider a real valued signal sampled off air (Figure 1). We will follow convention and define signal and noise power in the “single sided” bandpass interval of the frequency spectrum with bandwidth  $B$  centered on  $\omega$ . As the interval is single sided, we must use complex valued quantities to represent it.

We wish to simulate a bandpass AWGN channel at baseband ( $\omega = 0$ ). This implies a frequency shift of the complex valued signal, but the signal remains complex valued and it's power is unchanged. The negative frequency component on the LHS of Figure 1 is redundant and after frequency shifting can be removed by filtering.

Note that even at baseband we must use complex valued quantities for the signal and noise to represent a bandpass signal of bandwidth  $B$ . For example, given a fixed sample rate  $B$  and noise power  $N$ , a real valued noise sequence can only represent a bandwidth of  $B/2$  which results in doubling the noise density  $N_0 = N/(B/2) = 2N_0$  compared to a complex valued noise sequence with the same power.

The energy of each BPSK symbol  $E_b$  is the signal power  $S$  divided by the symbol rate  $R_b = 1/T_b$ . The noise per unit bandwidth is the total noise power  $N$  divided by the bandwidth  $B$  of the system. If we are simulating at one sample

per symbol,  $B = R_b$ :

$$\begin{aligned}\frac{E_b}{N_0} &= \frac{S/R_b}{N/R_b} \\ &= \frac{S}{N} \\ &= \frac{A^2}{\sigma^2}\end{aligned}\tag{3}$$

where  $A$  is the amplitude of each BPSK symbol and  $\sigma^2 = N$  is the variance of the complex valued noise (mean noise energy per sample). Given a set point  $E_b/N_0$ :

$$\sigma = \frac{A}{\sqrt{E_b/N_0}}\tag{4}$$

The complex noise sample  $r_i$  can be generated as:

$$r_i = \frac{\sigma}{\sqrt{2}}(\mathcal{N}_{2i}(0, 1) + j\mathcal{N}_{2i+1}(0, 1))\tag{5}$$

where  $\mathcal{N}_i(0, 1)$  is the  $i$ -th sample of a unit variance, zero mean, real Gaussian noise source. Note the noise power is split evenly between the real and imaginary arms. Our symbols passing through an AWGN channel can be simulated at complex baseband as:

$$\begin{aligned}\hat{z}_i &= z_i + r_i \\ \hat{q}_i &= q_i + r_i\end{aligned}\tag{6}$$

If the noise is zero mean, we can estimate  $\sigma^2$  over  $K$  noise samples  $r_i$  as:

$$\sigma^2 = E[|r_i|^2] = \frac{1}{K} \sum_{i=0}^{K-1} |r_i|^2\tag{7}$$

## 1.1 SNR Measurement

In order to compare with other methods of speech communication that have varying bandwidths  $B$ , it is useful to formulate expressions for estimating SNR from the BPSK and QPSK symbols. The Signal to Noise ratio (SNR) is given by:

$$\begin{aligned}\frac{S}{N} &= \frac{E_b R_b}{N_0 B} \\ &= \frac{E_q R_q}{N_0 B}\end{aligned}\tag{8}$$

A noise bandwidth  $B$  needs to be selected; common choices are  $B = R_b$ , in which case  $S/N = E_b/N_0$ ; for HF radio  $B = 3000$  Hz to compare with existing analog and digital voice waveforms; or  $B = 1$  to obtain a normalised  $C/N_0$  carrier

power to noise density ratio - useful for comparing waveforms with different bandwidths.

At one sample per symbol, the power, the mean energy of each QPSK symbol over a window of  $K$  samples is given by:

$$E_q = E[|q_i|^2] = \frac{1}{K} \sum_{i=0}^{K-1} |q_i|^2 \quad (9)$$

Note the variance function should not be used to calculate  $E_q$ , as we cannot guarantee  $q_i$  is zero mean. As each QPSK symbol contains 2 BPSK symbols, the energy is split evenly:

$$E_b = E_q/2 \quad (10)$$

For example if the symbol amplitude is  $A = 1$ ,  $E_b = A^2 = 1$ , then  $E_q = 1+1 = 2$ .

To model transmission over multipath channels using OFDM we arrange the QPSK symbols as  $N_c$  parallel carriers, each running at a symbol rate of  $R_s = R_q/N_c$  symbols/s, where  $R_s$  is chosen based on delay spread considerations. Typical values for HF modems are  $N_c = 20$  and  $R_s = 50$  Hz. However the OFDM carriers are arranged such that the total symbol rate over the channel remains constant. So for a given signal power  $E_q$  and  $E_b$  remain constant (Table 1).

Waveform	$N_c$	$R_s$	$R_q$	$R_b$	$E_q$	$E_b$
Single Carrier BPSK	1	-	-	2000	-	$S/2000$
Single Carrier QPSK	1	-	1000	2000	$S/1000$	$S/2000$
OFDM QPSK	20	50	1000	2000	$S/1000$	$S/2000$

Table 1:  $E_b$  and  $E_q$  examples for single and multi-carrier OFDM waveforms for constant carrier power  $S$

## 1.2 Calibration and Testing

In order to evaluate the ML system early in the development process it is important to ensure the noise is correctly calibrated. The expressions above can be used to check the noise injection process:

1. Set a target  $E_b/N_0$  for the simulation run, and calculate  $\sigma$  using (4).
2. Establish the equivalent target SNR from (8) evaluated using the target  $E_b/N_0$ .
3. After the simulation run measure  $E_q = E[|q_i|^2]$  over a sample of transmitted symbols. Note that in general  $E_q \neq 2$  as the encoder outputs continuous values.
4. Calculate measured SNR using (8) and compare.

The calibration of the noise injection can be checked by replacing the encoder output  $z_i$  with discrete PSK symbols to create a digital modem, then measuring the BER at  $E_b/N_0$  points. The theoretical BER over an AWGN channel is:

$$BER = 0.5 \operatorname{erfc}(\sqrt{E_b/N_0}) \quad (11)$$

For a multipath channel:

$$BER = 0.5 \left( 1 - \sqrt{\frac{E_b/N_0}{E_b/N_0 + 1}} \right) \quad (12)$$

## 2 Radio Frequency Tests

Impressive results have been obtained from the symbol rate simulations of an OFDM modem. These assumed ideal synchronisation. We would like to verify these results using real radio signals in Over The Cable (OTC) and Over the Air (OTA) tests. This requires building up a rate  $F_s$  system, and synchronisation subsystems. For a first pass, the choice was made to use classical DSP pilot symbol based synchronisation, although we acknowledge potential for ML based synchronisation in future iterations.

The goal is to compare speech quality to SSB at  $E_b/N_0 = 0\text{dB}$  (approx -3dB SNR in a 3000Hz BW), and work through any issues that prevent the system working over real radio channels. PAPR optimisation will be ignored for the first iteration, as our initial goal is to verify the low  $E_b/N_0$  results suggested by the symbol rate simulations.

Pilot symbols have been inserted into each OFDM carrier at a rate of one pilot every 4 data symbols. From [1], the SNR loss from injecting the pilot symbols is given by:

$$L_p = 10 \log_{10} \frac{N_s}{N_s - 1} \quad (13)$$

For  $N_s = 5$ , the loss is 0.96dB. To maintain the same payload symbol rate over the channel, the bandwidth increases:

$$R'_b = R_b \frac{N_s}{N_s - 1} \quad (14)$$

For  $N_s = 5$ ,  $R'_b = 2500$  symbols/s, or a OFDM QPSK bandwidth of around 1250 Hz plus some guard bandwidth.

A pilot based sync system was built in PyTorch, and is used for coarse and fine timing, phase and amplitude equalisation. Unlike classical PSK, the ML network is likely to be sensitive to amplitude variations. Phase equalisation also allows small frequency offsets ( $\pm 2$  Hz) to be tracked, sufficient for tests with short samples.

Several phase estimators were prototyped, and evaluated using BER measurements. Maintaining low loss synchronisation at low  $E_b/N_0$  is challenging. Using per-carrier phase estimation makes the system less dependant on fine

timing accuracy and gives us the ability to handle multipath, but has higher loss than algorithms that consider all carriers at the same time. As further work a lower latent dimension  $d$  and higher  $E_b/N_0$ , would allocate more power to pilots, and result in less carriers.

Figure 2: OFDM pilot based synchronisation algorithm performance, tested by measuring the BER obtained using discrete PSK symbols on an AWGN channel. With ideal sync, the autoencoder produces intelligible speech at  $E_b/N_0 = -6$  dB which corresponds to BER=0.24. Several algorithms, combined with gain and frequency offsets were simulated.

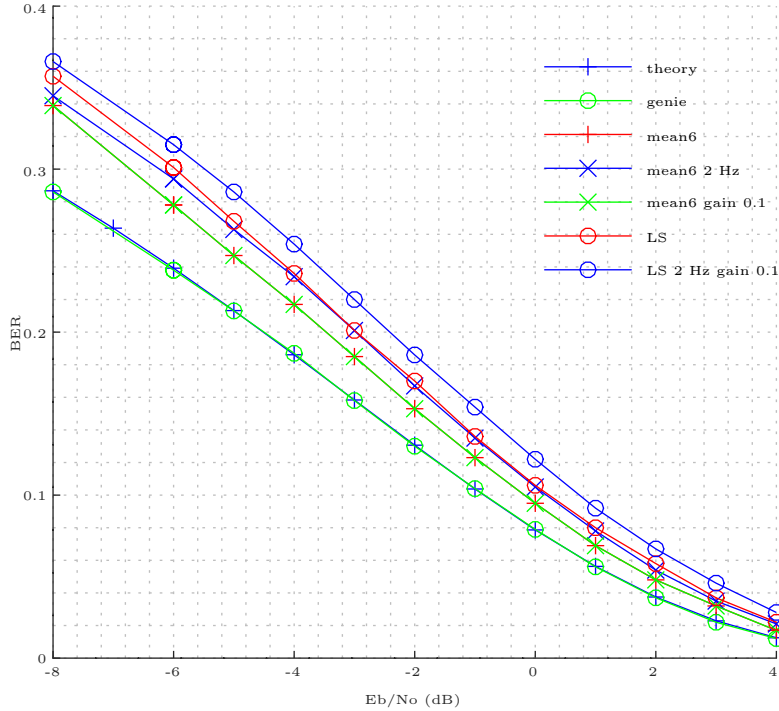
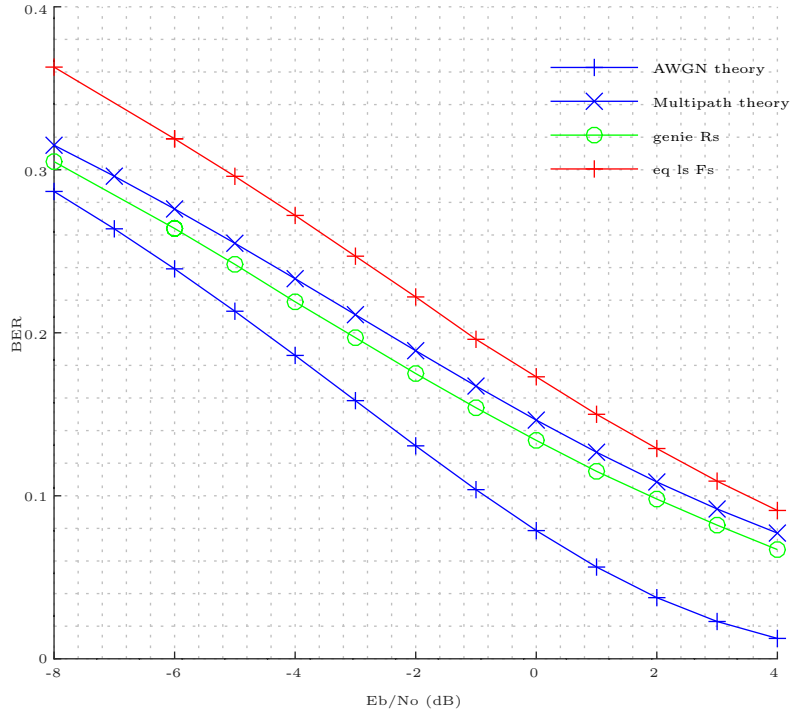


Figure 2 plots the simulated performance of the OFDM pilot based synchronisation system on an AWGN channel. The *genie* curve is the baseline rate  $F_s$  OFDM system with ideal sync, and matches the theoretical BPSK BER curve. Using this baseline system to send ML symbols, we obtain intelligible speech at  $E_b/N_0 = -6$ , which corresponds to BER=0.24 in a digital modem. We can use the BER=0.24 line to estimate the synchronisation loss at this operating point, which for the *mean6* and *LS* algorithms under realistic conditions is around 2

to 2.4 dB. The  $LS$  works better on fast fading multipath channels.

Combined with  $L_p$ , we estimate a total sync loss of 3dB for this first pass of the ML system combined with classical pilot based synchronisation. However we note that no cyclic prefix has been added at this time.

Figure 3: OFDM pilot based synchronisation algorithm performance, tested by measuring the BER obtained using discrete PSK symbols on an multipath (MPP) channel.



TODO Figure of OFDM frame. TODO Block diagram of OTC tests  
Take aways:

1. We have combined ML vocoder, ML autoencoder and classical DSP OFDM to build a system capable of sending speech over radio channels. It is robust to AWGN and multipath channel impairments. Our initial simulation, VHF OTC, and HF OTA tests suggest performance highly competitive with the incumbent analog SSB at the same  $C/N_0$ .
2. The ML algorithms run at a 40ms frame rate (with some DSP at the

OFDM 20ms symbol rate), resulting in modest CPU requirements. The PyTorch simulation code runs several times faster than real time in inference mode.

3. Simulation and initial HF OTA results show surprisingly good performance on multipath channels where the period of the fading (100's of ms) is large compared to the 40ms analysis window of the ML code. Classical DSP would require a 1000-2000 ms interleaver (introducing a algorithmic delay of the same order) for similar robustness to fading on these channels.
4. Unlike regular PSK, we require magnitude estimation and correction, due to the limited dynamic range of ML systems and the wide dynamic range of radio signals.
5. The *Candidate 2* OFDM frame design [2] contains three latent vectors  $z$ , so introduces an algorithmic delay of 120ms, due to OFDM framing considerations. This amount of additional delay is tolerable in a PTT radio system, but ideally should be reduced.
6. Substituting classical PSK digital symbols and measuring BER is a very useful way to test sync, and to allows us to verify  $E_b/N_0$  using BER measurments during development.
7. Much of our development effort was concerned with the “is this too good to be true” question; significantly more effort was put into noise calibration and testing than the actual ML.

Further work:

1. Try a low dimension latent vector, e.g.  $d = 40$ , and see if similar speech quality can be obtained at 3dB higher  $E_b/N_0$ . This would result in lower sync losses and RF bandwidth for the same channel SNR or  $C/N_0$ , as the  $E_b/N_0$  of the pilots would be increased. Does the encoder output still resemble BPSK, or is it training to a higher order constellation?
2. There is significant synchronisation loss from the classical DSP OFDM wabeform and associated sync algorithms. Attempt to use the ML network to perform frequency, phase and amplitude equalisation, with or without passing the pilots to the decoder. An initial attempt (model07) without pilots showed some ability to correct phase, frequency, and magnitude offsets, but resulted in some performance degredation. However this may be acceptable if comparable to the pilot based sync losses. Some pilot or unique word injection may still be required to perform coarse and fine timing estimation using classical DSP running at the sample rate.
3. We may be able to improve on algorithmic delay using ML sync techniques that are less dependant on traditional OFDM framing considerations.
4. Some form or interleaving, or a long time window for the ML network, may improve performance.



5. Work to improve the current classical DSP sync, e.g. a feedback loop to track out frequency offsets is worth 1 dB.
6. Include PAPR optimisation and rate  $F_s$  multipath channels in the training.
7. To better model SSB, locate a better analog compressor, a 3rd party reference implementation in software form would be useful.
8. For formal subj evaluation, how about radio style 1-5 readability scale across a panel of listeners.
9. Definition of lower limit link closure for this use case, for example “CQ CQ, and callsign, enough to produce a QSO report”.
10. For a practical implementation, coarse amplitude and timing need to be updated regularly. As we are testing short samples, a single block estimate is used at present.

### 3 Comparison with Other Speech Waveforms

We wish to compare our radio autoencoder with existing waveforms used for speech transmission over radio channels. This section assumes the *radac* system can send intelligible speech at an  $E_b/N_0$  of 6dB, with a PAPR of 1dB, both results have been demonstrated in simulation (but not at the time of writing together over real world radio channels). We include a 3dB synchronisation overhead.

We start with the assumption that we have a transmitter of  $C$  watts, and an AWGN channel with a spectral noise density of  $N_0$  watts/Hz. As the speech waveforms being considered vary in bandwidth we will choose  $C/N_0$  as the SNR metric.

The  $C/N_0$  (in dBHz) at the demodulator input of a terrestrial radio receiver is given by:

$$\frac{C}{N_0} = P - PAPR - L_{path} - NF + 174 \quad (15)$$

where  $P$  is the maximum output power of the transmitter power amplifier,  $PAPR$  is the Peak to Average Power Ratio of the waveform,  $L_{path}$  is the path loss,  $NF$  is the noise figure of the receiver. For example consider a 400 MHz FM hand held radio over a 1km urban (non line of site) path. The radio has a 1W (30 dBm) power output,  $L_{path} = 120$  dB, with noise dominated by ambient EMI such that  $NF = 10$  dB.  $C/N_0 = 30 - 0 - 120 - 10 + 174 = 74$  dBHz, sufficient for good quality speech (Table 5).

Note that  $C/N_0$  at the demodulator is a function of the waveform PAPR. With all other link properties (e.g. peak PA power, noise figure, path loss) being equal, a high PAPR reduces the  $C/N_0$  available at the receiver. We effectively “back off” the transmitter power from the maximum  $P$  by the PAPR. We assume the PA is capable of sustaining  $P$  watts indefinitely, i.e. it is only the waveform choice that lowers the average power. As PAPR varies by waveform and has an

impact on the  $C/N_0$  available the receiver, it should be included in any metric for comparison of waveforms. We define  $P/N_0$  as:

$$P/N_0 = C/N_0 + PAPR \quad (16)$$

A waveform that delivers intelligible speech at a low  $P/N_0$  is the target. A low PAPR waveform has other desirable properties, such as greater PA efficiency, longer battery life, and the use of low cost semiconductors in the radio power amplifier electronics.

Waveform	Threshold
Single Sideband	0dB SNR in 3000Hz noise BW, 2400Hz audio bandwidth, Tx speech compressor with 6dB PAPR
Frequency Modulation	-120 dBm quoted for many NBFM radios, 54dB above -174dBm/Hz noise floor
FreeDV 700D	10% PER threshold at -2dB SNR in 3000Hz noise BW
Radio Autoencoder	Intelligible speech at $E_b/N_0 = -6$ dB, $R_b = 2000$ symbols/s, 3dB sync overhead, a PAPR of 1dB

Table 2: Thresholds for speech link closure for each waveform. The link is considered closed when the speech is barely intelligible to a trained listener.

TODO: include 1st gen VHF/UHF digital voice - I think they go down to -123 dB (5%) BER, but speech quality is sub FM.

Waveform	Threshold $C/N_0$ calculations (dBHz)
Single Sideband	$0 + 10\log_{10}(3000) = 35$
Frequency Modulation	$-120 + 174 = 54$
FreeDV 700D	$-2 + 10\log_{10}(3000) = 33$
Radio Autoencoder	$-6 + 10\log_{10}(2000) + 3 = 30$

Table 3: Threshold  $C/N_0$  calculations.

Waveform	Abbr	RF BW	PAPR	$C/N_0$	$P/N_0$	$\Delta$
Single Sideband	SSB	2400	6	35	41	-10
Frequency Modulation	NBFM	16000	0	54	54	-23
FreeDV 700D	700D	1100	4	33	37	-6
Radio Autoencoder	radAE	1400	1	30	31	0

Table 4: Comparison of link closure by waveform over AWGN channels.

Waveform	Audio BW	$C/N_0$	$P/N_0$	$\Delta$
Radio Autoencoder	8000	36	37	0
Frequency Modulation	3000	64	64	-27
Single Sideband	3000	55	61	-24

Table 5: Comparison of good quality “arm chair copy” by waveform over AWGN channels. They are ranked in terms of maximum achievable speech quality. FreeDV 700D has been omitted because of its low speech quality even at high SNR. Even at 20dB SNR there is noticeable noise in received SSB, although DSP based noise reduction may help. Only the radio autoencoder delivers wideband (8000 Hz) speech.

## 4 Glossary

Symbol	Explanation	Units
$B$	noise or signal bandwidth	Hz
$C$	Carrier (transmitter) power $C = S$ for this study	
$C/N_0$	Carrier power/spectral noise density	
$d$	dimension of latent vector $\mathbf{z}$	
$E_b/N_0$	energy per BPSK symbol on spectral noise density	
$E_q/N_0$	energy per QPSK symbol on spectral noise density	
$N$	total noise power	Watts
$N_c$	Number of carriers	
$N_0$	Noise power in 1 Hz of bandwidth	
$P/N_0$	Peak trasnmmitter power/spectral noise density	
$\mathbf{q}$	vector of QPSK symbols	
$q_i$	single QPSK symbol, element of $\mathbf{q}$	
$R_b$	BPSK symbol rate	symbols/second
$R_q$	QPSK symbol rate	symbols/second
$R_s$	OFDM per carrier QPSK symbol rate	symbols/second
$R_z$	latent vector update rate	Hz
$SNR$	signal to noise Ratio	
$S$	total signal (carrier) power	Watts
$T_b$	BPSK symbol period	seconds
$T_q$	QPSK symbol period	seconds
$T_s$	OFDM per carrier QPSK symbol period	seconds
$T_z$	time between latent vector updates	seconds
$r_i$	noise sample	
$\mathbf{z}$	Autoencoder output latent vector	
$z_i$	single latent vector element of $\mathbf{z}$ , a BPSK symbol	

Table 6: Glossary of Symbols

## 5 References

- [1] Low SNR FreeDV Modes. [https://github.com/drowe67/misc/blob/master/freedv\\_low/freedv\\_low.pdf](https://github.com/drowe67/misc/blob/master/freedv_low/freedv_low.pdf).
- [2] David Rowe. FreeDV-032 Radio Autoencoder Waveform Design (spreadsheet).