Analog Transmission of Vocoder Features

October 24, 2023

Given a vector of vocoder features $\mathbf{f}$, use an autoencoder E to map them to a dimension D vector $\mathbf{g}$ where D is even. The elements of $\mathbf{g}$ are mapped to $D/2$ complex symbols $\mathbf{s}$. The magnitude of each symbol is constrained to a maximum of 1, but unlike digital modulation is not constrained to a discrete set of points on the complex plane.

The complex symbols $\mathbf{s}$ are updated every T_s seconds. The total symbol rate over the channel is therefore $R_s = D/(2T_s)$ Hz. If a spectrally efficient form of transmission is used the bandwidth over the analog (e.g. radio) channel will also be R_s.

The channel is AWGN with a noise power spectral density N_0 Watts/Hz. We assume perfect synchronisation between the transmitter and receiver such that the received symbols:

$$\hat{\mathbf{s}} = \mathbf{s} + \mathbf{n} \tag{1}$$

where $\mathbf{n}$ is a vector of complex (single sided) Gaussian noise samples.

The Signal to Noise ratio (SNR) is the ratio of signal power S to noise power N (both in Watts) given by:

$$\begin{aligned} \frac{S}{N} &= \frac{E_s R_s}{N_0 B} \\ &= \frac{E_s D}{2 N_0 T_s B} \end{aligned} \tag{2}$$

where E_s is the mean energy per symbol, and B is the bandwidth we measure SNR in. This is often $B = R_s$ but when comparing to another communications system it is useful to use a common B.

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

Tasks:

1. Collate training databases, extract LPCNet features.

2. Design and code up a suitable autoencoder network.

3. How to simulate channel (add complex noise to complex symbols) during training?

4. We could simplify for a first pass by treating the channel as D real symbols, and adding real noise.

5. How to constrain symbol magnitude during training?

6. Train at a candidate SNR.

7. Test over a channel at the candidate SNR. and evaluate quality

8. If we start with just LPCNet spectral (cepstrum) features we can use spectral distortion as a early objective indication of performance.

9. It would be interesting to contrast with a more traditional digital model, e.g. same autoencoder network but VQ or entropy encode to a bit stream and use QPSK (or just the known BER at a given SNR) over the channel.

10. It would be interesting to determine SNR performance at different D. Small D will make each symbol more sensitive to noise, however the power will be concentrated in less symbols. The Shannon–Hartley theorem suggests wider bandwidth (larger D) will be better.