Mathematical Analysis

Transmitter (Speech Recognition):

→ Semantic Encoder: input Spectrum → text related semantic features.

- Input: speech sample $m = [m_1, m_2, \ldots m_Q]$
- It is divided into N frames.

→ Applying Hamming window to all n, as

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right), \quad 0 \le n < N$$

→ Let X be the vector of length N of FFT,

$$\therefore X(k) = \sum_{n=0}^{N-1} w(n) \cdot e^{-j \frac{2\pi}{N} kn}$$

→ After appling Hanning window, FFT, logarithmic operations and normalization, we get a spectogram that is the input to our semantic encoder.

→ The ultimate goal of the ASR task is to recover the final text transcription $\hat{t}$, as close to t.

→ Semantic Encoder is made up of ~~two~~ CNN and GRU (Bidirectional RNN).

→ The input Spectra S, are first converted to intermediate features using several convolutional layers.

→ The no. of filters, in Each CNN module is E_p, $p \in [1, 2, \ldots P]$.

→ The output of the last CNN module is

$$b \in R^{B \times C_p \times D_p \times E_p}$$

→ Then b is fed into Q BRNN modules, that gives $d \in R^{B \times G_Q \times H_Q}$, where the no. of GRU units in BRNN module is H_q, $q \in [1, 2, \ldots, Q]$.

→ Finally the text related semantic features, P are obtained from d by passing through multiple cascaded dense layers and a softmax layer.

Channel Encoder:

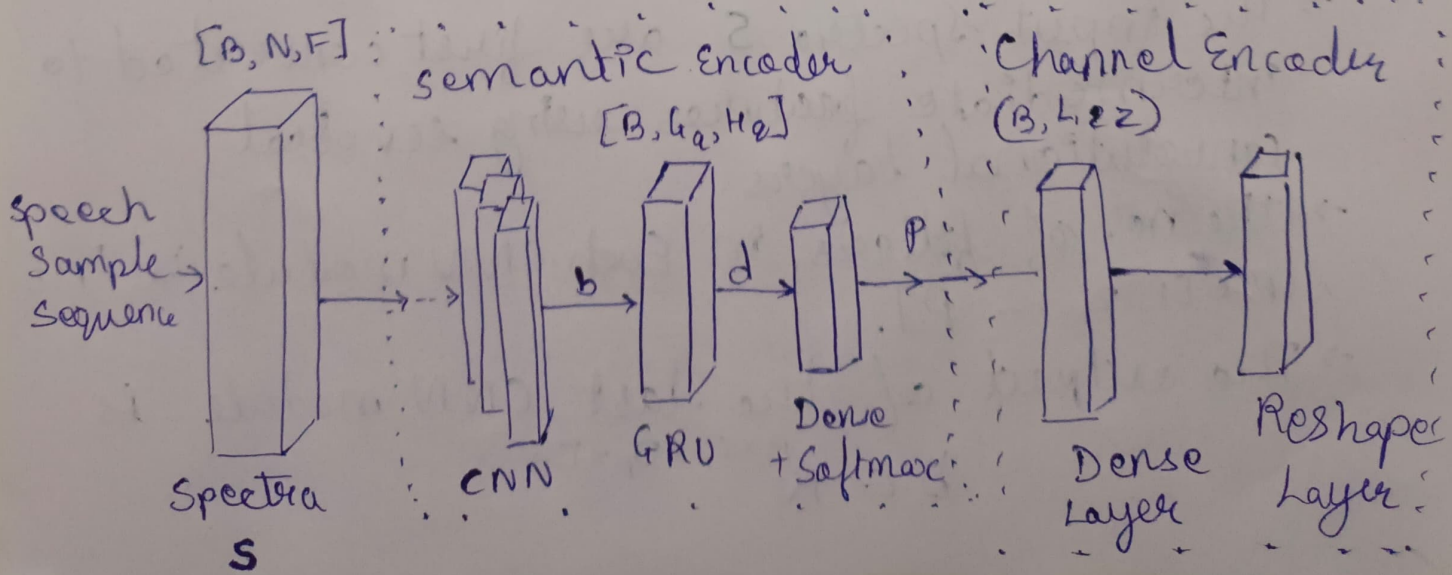
→ These features are mapped into symbols $\vec{x}$ by the channel encoder to be transmitted as

$$\vec{x} = \vec{T}_\beta^{\,c}\left(T_\alpha^{s}(\vec{s})\right)$$

(T_α^{s}: semantic encoder; $\vec{T}_\beta^{\,c}$: channel encoder.)

- Assumption: $E\|\vec{x}\|^2 = 1$
- System model assuming perfect CSI:

$$\vec{y} = \vec{h} * \vec{x} + \vec{w} \; ; \; \vec{w} \sim CN(0, \sigma^2 I)$$



System Model Assuming Imperfect CSI.

→ Imperfect CSI at Rx is more realistic assumption

System model:

$$y = \hat{h}x + n$$

$n \sim CN(0, N_0)$

Received signal at R_x : $(-\sqrt{E_s}, \sqrt{E_s})$

Let us assume the imperfect CSI to $\hat{h}$. and the correlation between h and $\hat{h}$ be ρ.

i.e. $E[h^* \hat{h}] = \rho$, $0 \le \rho < 1$.

Assumption:

① Imperfect CSI is denoted as $\hat{h}$, is available at the receiver for detection (R_x)

② $h \sim CN(0,1)$, ρ = correlation between h & $\hat{h}$.

$$E[h^* \hat{h}] = \rho , \quad 0 \le \rho < 1$$

By Maximum Likelihood detection rule:

$$d = Re\{\hat{h} * y\}$$

$$= Re\{\hat{h} * (hx + n)\}$$

If $d \ge 0$ the detected symbol is 1.

$d < 0$, the detected symbol is 0.

h can be expressed in terms of $\hat{h}$ and Random error δ as follow:

$$h = \rho\hat{h} + \sqrt{1-\rho^2}\,\delta,$$

$$\delta \sim CN(0,1).$$

$$d = Re\{\hat{h}^*(\rho\hat{h} + \sqrt{1-\rho^2}\,\delta)x + n\}$$
$$= Re\{(\rho|\hat{h}|^2 + \sqrt{1-\rho^2}\,\hat{h}^*\delta)x + \hat{h}^*n\}$$
$$= Re\{\rho|\hat{h}|^2 x + \sqrt{1-\rho^2}\,\hat{h}^*\delta x + \hat{h}^*n\}$$
$$= \rho|\hat{h}|^2 x + Re\{\sqrt{1-\rho^2}\,\hat{h}^*\delta x + \hat{h}^*n\}$$

$$z = \frac{d}{|\hat{h}|} = \underbrace{\rho|\hat{h}|x}_{\text{signal part}} + Re\Big\{\underbrace{\sqrt{1-\rho^2}\,\frac{\hat{h}}{|\hat{h}|}\delta x}_{\text{Not a noise part but due to } \delta \text{ (Random error)}} + \underbrace{\frac{\hat{h}^*}{|\hat{h}|}n}_{\text{noise part}}\Big\}$$

$\hat{h}$ and δ both should be independent

Nakagami-m Fadding Channel:

Gray coded discrete (square and rectangular) M-ary QAM ~~modules~~ mods with $M = 2^n$ $(n = 2, 3, \ldots, N)$ are used for the channel adaptive scheme.

→ The transmissions are assumed to be over a slowly flat-fading channel model assumed to follow a Nakagami-m distribution

Received signal $y = ax + gn$

where $a \rightarrow$ Nakagami - m - fading coefficient with pdf as

$$P_a(a) = \frac{2}{\Gamma(m)} \left(\frac{m}{\Omega}\right)^m a^{2m-1} \exp\left\{-\frac{ma^2}{\Omega}\right\}$$

where m is the Nakagami fading parameter and $\Gamma(m)$ is the Gamma function defined by $\Gamma(m) = \int_0^\infty y^{m-1} e^{-y}\, dy$.

and $\Omega = E[|a|^2]$

Expectation

By simplifying the instantaneous SNR will be

$$\gamma = \frac{|a|^2 E_s}{N_0}$$

for Nakagami - m fading channel, Pdf of γ is given by

$$P_\gamma(\gamma) = \left(\frac{m}{\gamma}\right)^m \left(\frac{\gamma^{m-1} \exp\left\{-\frac{m\gamma}{\gamma}\right\}}{\Gamma(m)}\right)$$

$\gamma \geq 0$.

$$\bar{\gamma} = E[\gamma] = \frac{\Omega E_s}{N_0}$$

The received SNR is split into (N+1) fading regions (bins) with region n having a corresponding mode m_n.

$\{r_n^m\}_{n=2}^{N}$ contains lower Threshold for N fading regions

r_1^m is set to 0dB and r_{N+1}^m to ∞.

Thus Received SNR, r, falls within region n ($r_n^m \leq r^m < r_{n+1}^m$)

$r_1^m \leq r^m < r_2$ (outage)