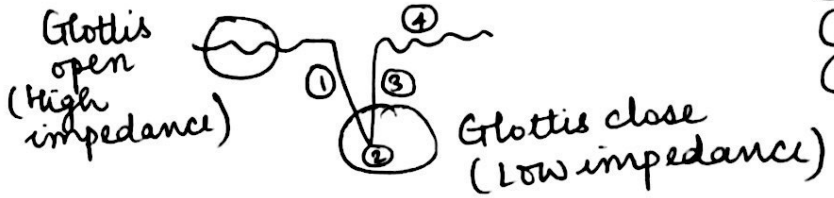
Assignment 3

Harshita Sharma
20171099

Quest: Epoch Extraction using EGG and ZFF. Voice activity detection using ZFF signal.

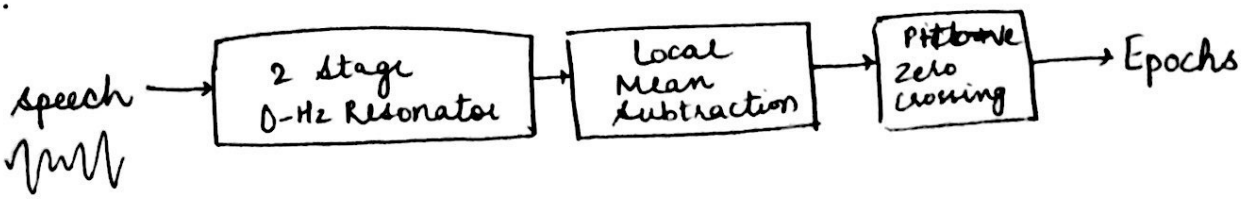
EGG: Electroglottography (EGG) is a noninvasive method of measuring the vocal fold contact during voicing without affecting speech production. EGG measures the variation in impedance to a very small electrical current between a pair of electrodes placed across the neck, as the area of contact of the vocal folds changes during voicing. The demodulated impedance signal is referred to as EGG signal.

① - closing
② - closed
③ - opening
④ - open



Hence, the four distinct phases of a glottal cycle can be identified.
This EGG signal is differenced to enhance the epochs — wherein locations of negative peaks correspond to the instants of glottal closure.

ZFF:



The discontinuities in the excitation signal caused by the sharp closure of the glottis can be approximated by ~~the~~ a sequence of impulses of varying amplitudes. The effect of impulse-like Epoch-based analysis of speech signals excitation is reflected across all frequencies, including the zero-frequency (0-Hz). The effect of the discontinuity due to the impulse-like excitation is clearly visible in the output of narrowband filtering of the speech signal. Since the vocal-tract system has resonances at much higher frequencies than at the zero-frequency — a zero-frequency resonator is chosen.

① The speech signal is differenced to remove any slowly varying component.

② The differenced signal is passed through a cascade of two ideal zff — to provide sharper cut-off to reduce the effect of resonances of the vocal tract system.

③ The average pitch period is computed after which the trend in the final signal is removed by subtracting the local mean computed over the average pitch period.

The resulting signal is called zero-frequency filtered (ZFF) signal

Now, the +ve to -ve zero crossing correspond to the epochs found.

Voiced activity detection using ZFF:

~~In the about~~ In the absence of vocal fold vibration, the vocal tract system may be excited by random noise. The energy of the random noise excitation is distributed both in time and frequency domains. Whereas the energy of an impulse is highly concentrated in the time domain, but is distributed uniformly in the frequency domain. The ZFF signal exhibits significantly lower amplitudes for random noise excitation compared to the impulse-like excitation and hence can be used to detect the regions of glottal activity. The unvoiced regions in the speech signal have very low amplitude in the ZFF signal. Hence, the short-term energy of the ZFF signal ~~computed over~~ can be used for glottalic activity detection.

The accuracy of the zero-crossings derived from the filtered signal of speech and the robustness of the zero-crossings derived from the HE are used in conjunction to obtain an accurate and robust estimate of the instantaneous fundamental frequency.

Ques 2: Spectral Estimation using ZTW technique.
Relation between ZTL and ZFF.

ZTW (zero time windowing) extracts spectral features of the vocal tract system from very short segment of speech, at every instant of time.

The operation of passing a signal through a 0 Hz resonator is equivalent to multiplying the spectrum of the signal with a window function given by the frequency response of the resonator. By choosing a similar function in time domain which gives more weightage to samples around zero-time — we are performing an operation which is closer to integration in the frequency domain and thereby not smearing the spectral information as much as any arbitrary window would. This window function is given by —

$$w_1[n] = \begin{cases} 0 & n = 0 \\ 1/(4\sin^2(\pi n/2N)) & n = 1, 2 \ldots N-1 \end{cases}$$

where N is the window length. The window fn $\frac{1}{4\sin^2\left(\frac{\pi n}{2N}\right)}$ can be approximated to $1/n^2$ for smaller values of n; if $N >> M$ where M is the length of the segment after apending with zeros. Hence the chosen function provides an approximation to integration in frequency domain.

Relation b/w ZFF and ZTW

The effect of zero-time windowing operation in time domain is equivalent to smoothing the spectrum by successive integration in the frequency domain. This is analogous to ZFF ~~propos~~ for extraction of the excitation source features.

Ques 3: Speech enhancement using spectral subtraction and its drawbacks. Any improvisation to overcome them.

Spectral subtraction is one of the first algorithms proposed for enhancement of single channel speech. In this method, the noise spectrum is estimated during speech pauses and is subtracted from the noisy speech spectrum to estimate the clean speech. Consider a noisy signal $y[n]$ which consists of clean speech $s[n]$ degraded by statistically independent additive noise $d[n]$—

$$y[n] = s[n] + d[n]$$

This representation in short-time Fourier Transform is given by—

$$Y(w,k) = S(w,k) + D(w,k)$$

where k is frame number. Since the speech is uncorrelated to bg noise (assumpⁿ) —

$$|Y(w)|^2 = |S(w)|^2 + |D(w)|^2$$

The speech can be estimated by subtracting a noise estimate from the recived signal —

$$|\hat{S}(w)|^2 = |Y(w)|^2 - |\hat{D}(w)|^2$$

Now, the estimation of noise spectrum can be achieved by averaging recent speech pauses frames.

Finally,

$$|\hat{S}(w)| = \max\{|Y(w)| - E\|D(w)\|, 0\}$$

$$\angle\hat{S}(w) = \angle Y(w)$$

Drawbacks: The effectiveness of spectral subtraction is heavily dependent on accurate noise estimation, which is a difficult task. When the noise estimation is less than perfect, two ~~prob~~ major problems occur:

- Remnant noise with musical structure and
- Speech distortion

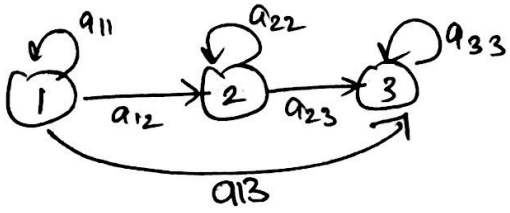
Improvisation: A number of variants of this method have been developed to address the mentioned drawbacks which form a family of spectral subtractive-type algorithms. ~~The~~ (such as spectral over-subtraction, Wiener filtering etc.) The principle of these algorithms is to estimate the short-time spectral magnitude of the speech by subtracting estimated noise from the noisy speech spectrum or by multiplying the noisy spectrum with gain functions and to combine it with the phase of the noisy speech.

Assignment 4

Harshita Sharma
20171099

Ques 1: Discrete Markov Process

Ans: Markov chain is discrete-time stochastic process such that each random variable takes place in a discrete set. It is a stochastic model describing a sequence of possible events in which probability of each event depends only on the state attained in the previous event. Markov Model is a sequence of states. Each state is associated with probability density function. Consider a system with N no. of States —



Markov chain with 3 states

At regularly spaced discrete times, the system undergoes a change of state, according to a set of probabilities associated with the state. Time instant associated with state change is denoted as $t = 1, 2 \cdots$ & state at time t as q_t. A full probabilistic description of the system, would in general require specification of the current state and predecessor states. If we consider the process to be independent of time — we get a set of state transition probabilities $a_{ij} = P[q_t = S_i \mid q_{t-1} = S_j] \quad 1 \le i, j \le N$

The above process is called as an observable Markov Model.

Ques 2: Extension of discrete Markov process to HMM

Ans: If a probability is not associated with a state, it is called Markov model.
If a probability density function is associated with state it is called HMM. It is statistical markov model in which system being modelled is assumed to be a markov process with unobserved (hidden) states. Hidden refers to state sequence through which the model passes but the observation are seen. To extend the ideas of HMM, consider the "Urn and Ball model":
Consider N glass urns in a room. Each urn has large number of coloured balls. Let 'm' be the number of distinct coloured balls. According to some random process, someone chooses an initial urn — and from this urn — a ball is chosen at random, its colour is recorded as observation. The ball is then replaced in the urn from which it was selected. A new urn is then selected according to the random selection process associated with current urn and ball selection process is repeated.

Ques 3: Elements of HMM.

Ans 3: ① Number of states: N
② Set of states: $Q = \{q_1, q_2 \dots q_n\}$
at time instant 1, system is in q_1 state.

③ No. of distinct symbol per state = M

④ Symbol probability or observation prob. distribution

$$B = \{b_j\} \quad 1 \le j \le N$$

⑤ State transition probability matrix A

$$A = \{a_{ij}\} \text{ where } a_{ij} = \{P(q_{t+1} = j \mid q_t = i)$$

Ques 4: Mathematical formulation of three basic problems of HMM.

Ans: Problem 1: Given an observation sequence $O = O_1, O_2 \cdots O_T$ and trained model $\lambda = (A, B, \pi)$ how to efficiently compute the likelihood $P(O|\lambda)$ — likelihood of the model generating the observation sequence O. This is called the Evaluation Problem.

Goal is to compute $P(O_1, O_2 \cdots O_T|\lambda)$. There are many state sequences. Consider one state seq.

$q = q_1, q_2, q_3 \cdots q_T$

If we assume that observations are independent

$$P(O|q,\lambda) = \prod_{t=1}^{T} P(O_t|q_t,\lambda)$$

$$P(O|q,T) = b_{q_1}(O_1)\, b_{q_2}(O_2) \cdots b_{q_T}(O_T)$$

and

$$P(q|\lambda) = \pi_{q_1} a_{q_1 q_2} \cdots a_{q_{T-1} q_T}$$

We know,

$$P(O|\lambda) = \sum_{T} P(O,q|\lambda)$$

Then,

$$P(O|\lambda) = \sum_{T} P(O|q;\lambda) P(q|\lambda)$$

$$= \sum_{q_1 \cdots q_T} \pi_{q_1} a_{q_1 q_2} \cdots a_{q_{T-1} q_T} b_{q_1}(O_1) \cdots b_{q_T}(O_T)$$

→ Order $2TN^T$

NOT EFFICIENT!

Problem 2: Optimal Path / Decoding Problem

Given O and λ, how to find the optimal state sequence ($Q = q_1, q_2 \dots q_T$) or best state sequence.

One way to do so is to find out at each state maximum probability —

$$\gamma_t(i) = P[q_t = i \mid O, \lambda]$$

$$= \frac{P[O, q_t = i \mid \lambda]}{P(O \mid \lambda)}$$

And calculate $\max \gamma_t(i)$ at each instant. But this may sometimes result in invalid state sequences because the above mentioned solution finds maximum probability without regard to the probability of occurence of states.

Problem 3: Training Problem

Adjust $\lambda = (A, B, \pi)$ to best maximise $P(O \mid \lambda)$ i.e. given training data how to estimate parameters of $\lambda = (A, B, \pi)$ so as to maximise the probability of representation of training data by the model. — $P(O \mid \lambda)$.

Ques 5: Solutions to the three problems of HMM.

Ans: Solution to Problem 1: Forward Algorithm.

① Define forward variables as:

$$\alpha_t(i) = P(O_1 \dots O_t, q_t = S_i \mid \lambda)$$

The probability that the state at position t is S_i and of the partial observation $O_1 \dots O_t$

given the model λ.

forward step:

1) Initialisation:

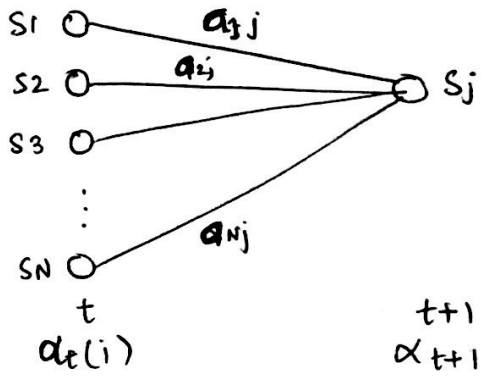
$$\alpha_1(i) = \pi_i b_i(O_1) \qquad 1 \le i \le N$$

2) Induction:

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij} \right] b_j(O_{t+1}) \qquad \begin{matrix} 1 \le t \le T-1 \\ 1 \le j \le N \end{matrix}$$

3) Termination

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$$



Solution to Problem 2: Viterbi Algorithm

1) Initialisation:

$$\delta_1(i) = \pi_i b_i(O_1) \qquad 1 \le i \le N$$

$$\psi_1(i) = 0$$

2) Recursion:

$$\delta_t(j) = \max_{1 \le j \le N} [\delta_{t-1}(i) a_{ij}] b_j(O_t) \qquad 2 \le t \le T;\ 1 \le j \le N$$

$$\psi_t(j) = \operatorname*{argmax}_{1 \le i \le N} [\delta_{t-1}(i) a_{ij}] \qquad 2 \le t \le T;\ 1 \le j \le N$$

3) Termination:

$$p^* = \max_{1 \le i \le N} (\delta_T(i))$$

$$q_T^* = \operatorname*{argmax}_{1 \le i \le N} (\delta_T(i))$$

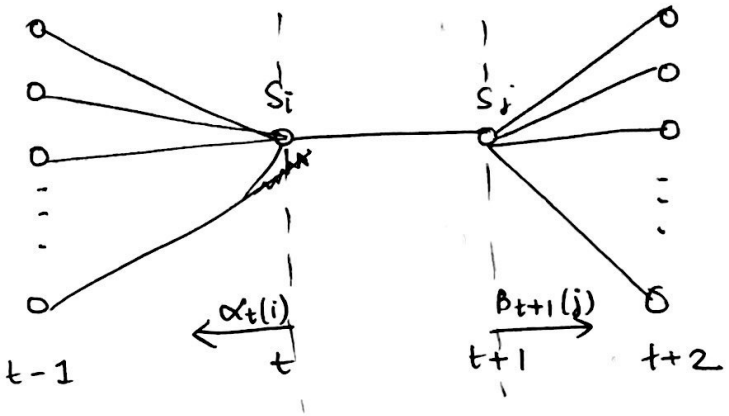
4) Path (state-sequence) backtracking

$$q^*_t = \Psi_{t+1}(q^*_{t+1}) \qquad t = T-1, T-2, \ldots 1$$

Hence –
- best state individually likely at position i
- best state given all the previously observed states and observations

Solution to Problem 3 : Baum-Welch algorithm

To re-estimate (iteratively update and improve) HMM parameters A, B, π by –



① Define $\xi_t(i,j) = P(q_t = S_i, q_{t+1} = S_j \mid O, \lambda)$

② Putting forward and backward steps, we can write $\xi_t(i,j)$ as –

$$\xi_t(i,j) = \frac{\alpha_t(i)\, a_{ij}\, b_j(O_{t+1})\, \beta_{t+1}(j)}{P(O|\lambda)}$$

$$= \frac{\alpha_t(i)\, a_{ij}\, b_j(O_{t+1})\, \beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_t(i)\, a_{ij}\, b_j(O_{t+1})\, \beta_{t+1}(j)}$$

③ Define $\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j)$

④ We get —

$\sum_{t=1}^{T-1} \gamma_t(i)$ = expected number of transitions from S_i

$\sum_{t=1}^{T-1} \xi_t(i,j)$ = expected number of transitions from S_i to S_j

Using above formulas, a set of reasonable reestimation formulas for π, A and B are —

$\bar{\pi}_i$ = expected frequency in state S_i at time $t=1$ is $\gamma_1(i)$

$\bar{a}_{ij}$ = $\frac{\text{expected no. of transitions from } S_i \text{ to } S_j}{\text{expected no of transitions from } S_i}$

$= \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}$

and

$\bar{b}_j(k)$ = $\frac{\text{expected no. of times in } j \text{ observing } V_k}{\text{expected no. of times in state } j}$

$= \frac{\sum_{t=1}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)}$ s.t. $O_t = V_k$