Spoken long. Processing Midterm Exam - Fall 2017 Question 1: /k/ plosive, Un-Voiced /sh/ Fricative, Un-Voiced /AH/ Vowel, Voiced /M/ Nasal, Voiced (b) Pollowed by sudder explosion and aspiration When constriction · is released. Waneform e.cs. mm Affin \* Fricalives (e.g. /sh/): Non-periodic Sound (Priction) - vandom energy across wide freq. varge. Wantown e.g. Affrontalin Frequency Corresponding to Vibration rate of Vocal-folds.

Vocal Tract enhances some frequences and suppresses others.

Energy of Vowels Sounds is relativity higher them other Sounds. Waveform e.g. MMM.

21 (C) Specitio grain: Time frequency representation of speech symp It is a Concatenated spectrums of short frames (10-20 ms) and Shows the energy Sistribution over frequency and Time scale. Frg. 2 Three (5) " Spectrogram" Tompred Regoldian Windows Length treg resolution Wide Navrow - Band Wide - Band Navrow Fine Question 2 8 Source-Filter Model (KK) Vocal Track Parameters Impulse Train Glottel pulse Model generator lips Vocal radidi Tract Model Model 5(n) pitch period V(Z) R(2) Random Noise generator Filter Source

\* Impulse Frain generator corresponds to Vocal folds Vibration with frequency of to (fundamental) For Voiced Sounds in the largux Which shaps the Impulses to be like pulses (puffs) of For Un-voiced Sounds (eg. /sh/) the output of the lavynx is noise like signal. This is modeled by a random noise generator. To general, Voiced sounds have greater amplitude them to con-voiced. This is represented by Multiplying output of largues, by two constants An (Voiced) and Air (unvoiced)

An > Aur \* becal Track is modeled by a filter with sufficient unmber of poles (All-pole filter). Usually 10-14 poles for both Noices and UM-Voiced Sounds of lip radiation is modeled by a high-pass filter of a transfer function R(2). (b) Impulse Train: E(Z) = 1-ZP, P: Pitch Period.

Glothel model: (1-27)2 2 (1-27)2, for Voiced Sounds

G(Z) = 1 , For Univoiced Sounds

Vocal Trad model: 1+ Barzk, for Voiced Sound 1+ Sakzk / for Un-voiced Sound P = no. of poles L= no. of zeros. OR V(Z) = T+ & ak Zk , P= 10-14 and un-voiced speech. lips radiation & R(Z) = 1-Z, First-sider HPF.  $\frac{S(z)}{E(z)} = \frac{G}{1 + \frac{2}{5}a_{k}z^{k}} / 10.59 \leq 14$ \* Overall

# Lett hand graph (Nocal tract) is the envelop of the  Speech spectrum, or low Enqueury Composed of the Speech  Spectrum.  Source en Tract > speech signe  Source en Tract > speech signe  S(n) = E(n) × N(n)  S(2) = E(2) · N(2).   * This losts like a voiced sound because the power is highier in the (ow frequency Components (i.e. power at F1) power at F2) no.  (d) Two benefits of using DCT in the MFCC technique.  1. Un correlate features; so diagonal Covariance Midvix  Can be used in the Modelling.  Le separate source signal (exidence) from the Nocal tract filler response by truncoling DCT Coefficients to the filler response, and high order Coefficients Consponds to Source Signal.  1 of power Coefficients Source Source Source Source  Spectra	Question 2
Spectrum.    Source   em   Tract   > speech signed   Mn   S(n)	(c) Right hand graph is spectrum of speech Frame Left hand graph is the response of the Jocal tract
Source em Tract > speech signed  S(n) = eCn) × N(n)  S(1) = ECn) × N(n)  S(n)  First looks like a voiced Sound be cause the power is highier in the low frequency components (i.e. power at Fi ) power at Fi ) power at Fi ) now at fi ) no.  (d) Two bonefits of using DCT in the Marce technique,  I. Un correlete features; so disagonal Covariance Medrix  Can be used in the Modelling.  2. separate source signed (existed in) from the Vecal tract  Filter response by trunching DCT coefficients to the filter response; and high order Certificients Consponds to Source signed.  The filter response, and high order Certificients Consponds to Source signed.  In the filter response and high order Certificients Consponds to Source signed.  Source Signed.  In the filter response and high order Certificients Consponds to Source signed.  Source Signed.  In the filter response and high order Certificients Consponds to Source signed.  In the filter response and high order Certificients Consponds to Source signed.  Source Signed.  In the filter response and high order Certificients Consponds to Source.	of Left-hand graph (Nocal tract) is the envelop of the special spectrum, or low Frequency Componed of the special
John John Low frequency Components i.e. power at Fi power is highier in the low frequency Components i.e. power at Fi power of file.  (d) Two benefits of using DCT in the MFCC technique,  J., Un-Correlete features, so diagonal Covariance Mutrix  Can be used in the Modelling.  2. separate source signal (exidence) from the local track  filter response by truncoling DCT Coefficients to the  The low order coefficients (or first 12) Corresponds to Vocal  track filter response and high order Coefficients Consponds to  Source signal.  In the many power of the many power of the speaking the source of the second of the	
(d) Two benefits of using DCT in the MFCC technique,  1. Un-Correlete features, so singul Covariance Medvix  Can be used in the Modelling.  2. separate source signal (existation) from the Vocal tract filter response by trunching DCT Coefficients to the  The how order coefficients (or first 12) corresponds to Vocal  tract filter response, and high order coefficients componed to  source signal.  In Mannenan Spectrum  Log-power  Spectra  DCT  Source  Source	
I. UN-Correlete features, so Singenel Covariance Medvix  Can be used in the Modelling.  2. Separate source signal (existed on) from the Vocal track filter response by truncoling DCT Coefficients to the  The few order Coefficients (or First 12) Corresponds to Vocal  track filter response, and high order Coefficients Consponds to  Source signal.  The power of the Manney Spectrum  Tog power of the Source  Source Source  Source  Source  Source	in the low frequency components i.e. power at Fi > power at fr
The how order coefficients (or First 12) Corresponds to Vocal track filter response, and high order Coefficients Corresponds to Source Synd. In Manual Spectrum Vocal Track log power ( DCT )	(d) Two benefits of using DCT in the MFCE technique.  1. UN-Correlete features, so singonal Covariance Medvix  Can be used in the Modelling.
Source Synd. Johnson Spectrum Vocal Tract log power ( DCT Source Source	2. separate source signal (existation) from the Vocal tract filter response by truncating DCT Coefficients to the
Spectra DCT Source	The from order coefficients (or first 12) Corresponds to Vocal track filter response, and high order Coefficients Corresponds to Source signal. I Many specific spectrum / Vocal Track
	Speilra DCT Source

Question 3 8

$$=\frac{1}{8}\left[\left(-0.35\right)^{2}+\left(6\right)^{2}+\left(3.15\right)^{2}+\left(-2.5\right)^{2}+\left(-3.5\right)^{2}+\left(2.8\right)^{2}+\left(2$$

(ii) 
$$\frac{1}{N} = \frac{1}{N} \left[ \frac{1}{Sign(S(n))} - Sign(S(n-1)) \right]$$

$$=\frac{1}{3}\cdot(4)=\frac{1}{2}$$

(iii) 
$$R(k) = \sqrt{\sum_{n=0}^{N-1} S(n-k)}$$

We compute R(2), R(3) and R(4) => The pitch period corresponds to k which gives R(R) max. Value.

23. (9) (cii) Continue --

$$S(n)$$
 = 0.35 0 3.15 = 2.5 = 3.54 2.8 0 = 0.28  $S(n-3)$  0 0 = 0.35 0 3.15 = 2.5 = 3.54 2.8  $S(n-3)$  0 0 0 = 0.35 0 3.15 = 2.5 = 3.54  $S(n-3)$  0 0 0 0 0.35 0 3.15 = 2.5

R(2) = = (3.15)(-0.35) + (3.15)(-3.54) + (2.8)(-2.5) + (2.8)(-0.28) = -2.673

R(4) = \frac{1}{5}(-3.54)(-0.35) + (-0.28)(-2.5) = 0.2425

So, R(3) is the max. Value => k=3 Corresponds to Pitch pariod

Q3 (b)

High energy + low 2cc + nonzero Fo =) Voiced speech low energy + High 2cc + Zero fo =) Unvoiced speech

This speech segment is voiced be cause Fundamental frequency (fo) is Not Zero (fo = 20 Hz).

(Z)

## Question3 (c)

- (1) No because min. Sound pressure level for 100HZ is around 25 &B
- (2) atone of 1000 HZ at 2018 because human ear is more sensitive
- (3) B 2 kHZ 5 kHZ.