Digital Speech Processing
Prof. S. K Das Mandal
Center for Educational Technology
Indian Institute of Technology, Kharagpur

Lecture – 31
Segmental and Supra-Segmental features of speech signal

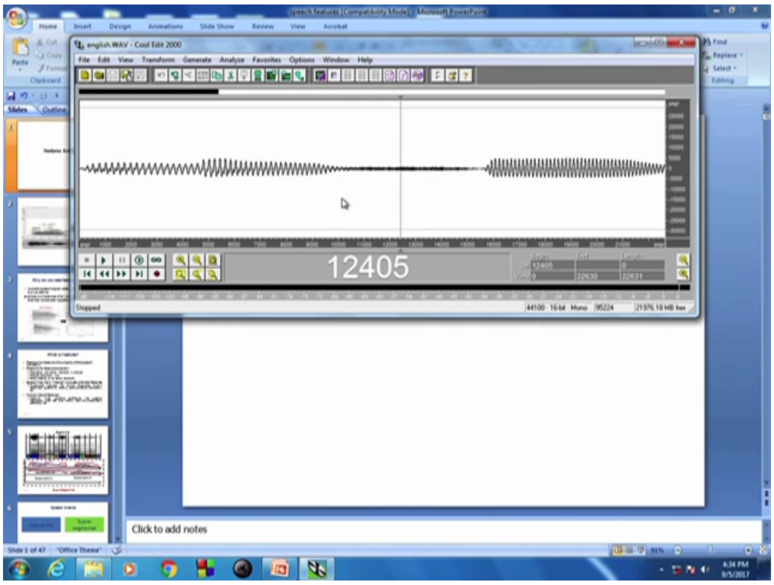
So let us start the new week which is this week we will discuss about the speech features extraction. So, all kind of that I can say that speech production systems we know. Now this week we try to find out what kind of features or why it is required the features ok.

(Refer Slide Time: 00:49)

Features Extraction

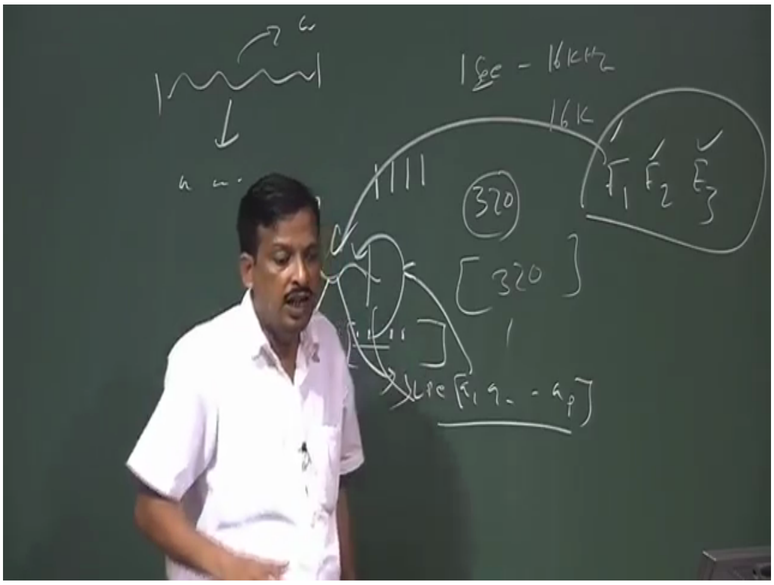
So, extraction of speech features this week title I can say the extraction of speech features. If you see when a speech signal I will saw you in cool edit softwares (Refer Time: 00:55) this is a speech signals.

(Refer Slide Time: 00:57)


english.WAV - Cool Edit 2000
File Edit View Transform Generate Analyze Favorites Options Window Help
12405
Stopped
Click to add notes

If I want to see the whole sentence let this is the whole sentence of this speech signals. So, if you see along the time line there is a different speech event. If you say the this is a speech event, this may be speech events speech events. So, I can say along the timeline speech is different or characteristics of the speech signal is different. You may say digital speech it selves is a features yes, once I do the sampling each sample can have a features x has a features.

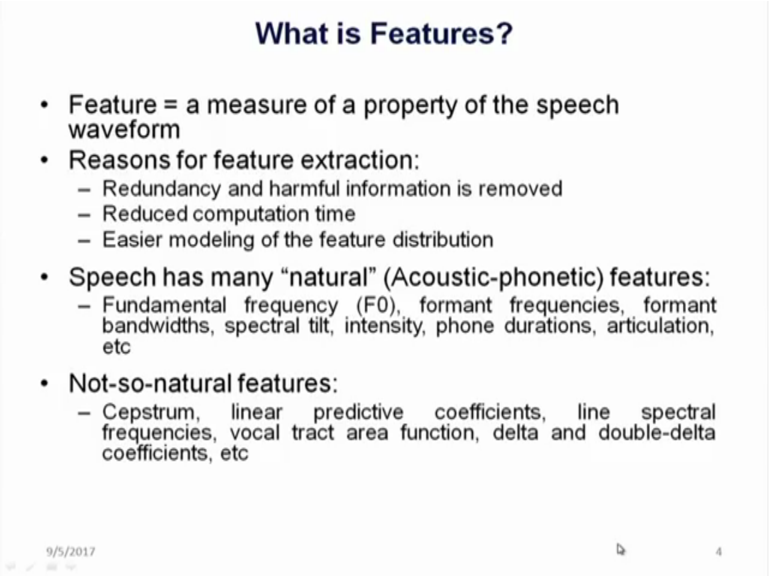
(Refer Slide Time: 01:31)



So, suppose. So, we said that speech sample can be a features. Now if you see if you consider a speech segment of one second how many samples will be there? If it is recorded with 16 kilo hertz, then I can say 16 kilo sample will be there. Even if I make a window let us 20 minute second window. So, 320 sample will be there 320 sample if compare sample by sample. Then the feature factor dimension is 320. Otherwise that if I say this time after one day or after next time 2 signals are sound like, but if I recorded those 2 signal and I try to compare I found they are different sample wise they are different. So, if I take the sampling as a feature extraction it is nots that good.

So, what will be the speech features or why it is required. So, if I say that then I can start from here.

(Refer Slide Time: 02:55)



What is Features?

- Feature = a measure of a property of the speech waveform
- Reasons for feature extraction:
  - Redundancy and harmful information is removed
  - Reduced computation time
  - Easier modeling of the feature distribution
- Speech has many "natural" (Acoustic-phonetic) features:
  - Fundamental frequency (F0), formant frequencies, formant bandwidths, spectral tilt, intensity, phone durations, articulation, etc
- Not-so-natural features:
  - Cepstrum, linear predictive coefficients, line spectral frequencies, vocal tract area function, delta and double-delta coefficients, etc

9/5/2017 4

What is features or parameters what about what about I said? So, what is speech features I can say let us take the speech signal and find out some parameters which is in lower dimension, or I can say that that there is a speech signal I take a signal segment and I represented in such dimension which represent the speech signal it selves. So, I can say the features is a measure of property of the speech signal. So, I can say some property of the speech signal will be there that may be a natural and non natural. And those property actually represent the speech signal.

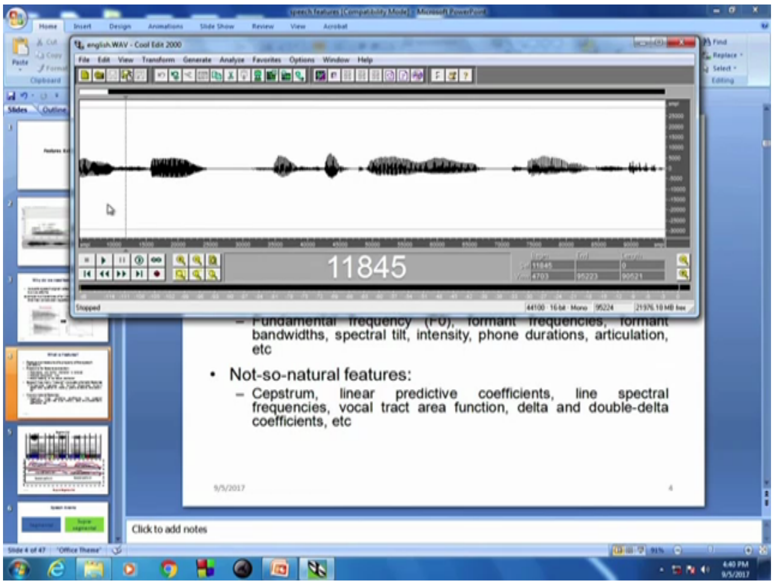
So, reason for feature extraction redone, then see if I say if I compare sample by sample there may be many sample which does not required to represent the signal. So, I can say like that suppose if you see the extract sound features by which you can remember me. So, actually we are not remembering my each and every. So, suppose I saw you my image, you are not remembering each and every pixel of my body. So, we are finding out some features which represent my body, but So they are not exactly my body. So, what I what I can say that redone then reformation which exist in my face you can discard those and find out the salient point which actually represent my face.

So, features extraction or the speech features in feature domain is nothing but they reduce the redundance information. And you can also remove the redundance information. If there is a 320 dimensional vectors, any classifier or any kind of comparison when I do it is a computationally complex every member I have to compare. So, that is the 320 member if an have simple equitant distance.

So, dimension is increase computational complexity is very large. And also modeling, suppose I want to model that using those features. So, if the features dimensionally very high the modeling is very difficult. So, actually feature extraction means that I want I want to convert this segment in a such away or such the I can say that I can represent this speech segment with a such a vector which actually represent this signals, and within redundance; that means, all redundance information is deleted and all the key informations are there to form a vector which the vector represent that speech event. So, that is call speech features ok.

So, there are many kinds of features. Some things may be natural features something may be un natural speech features. If it natural speech features if I say if you see in this speech like from here.

(Refer Slide Time: 06:28)



- Fundamental frequency (F0), formant frequencies, formant bandwidths, spectral tilt, intensity, phone durations, articulation, etc
- Not-so-natural features:
  - Cepstrum, linear predictive coefficients, line spectral frequencies, vocal tract area function, delta and double-delta coefficients, etc

english.WAV - Cool Edit 2000
File Edit View Transform Generate Analyze Favorites Options Window Help
11845
Stopped
Click to add notes
9/5/2017
Slide 4 of 47

there if you know that Acoustics phonetics So, if you see natural speech features are the phoneme different time consist of different phoneme. So, that can be a natural speech features. If you see the movement of the f 0 fundamental frequency during the whole sentence is not a constant. So, movement of the f 0 f 0 it selves can be a features phoneme can be it is features. If you see the duration if I say this is the duration of the (Refer Time: 07:03) this is the duration of the consonant to wall transition this is the transition wall duration wall to consonant transition.

So, all duration kind of speech can be a features. Then I can say articulation place of articulation in a speech event as a features ka velar consonant po belavia consonant. So, that can be a speech features. So, some features are natural some features are non natural they are abstract represent of the speech event, but some features has a meaning. Phoneme articulation, duration all has an real meaning in this speech, where from similarly formant frequency, if you if you remember we plot the f 1 f 2 plot of all havells. So, formant plot of the how all havells.

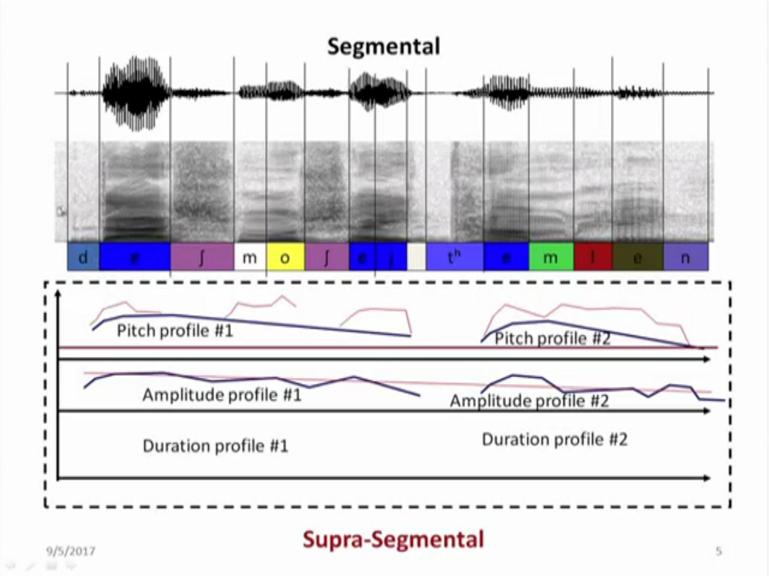
So now, using the f 1 f 2 f 3 all formant representation can be a features, by which I can identify the speech signal or by on the other hand I can say those formant frequencies represent the speech signal. So, formant may be a natural features. So, there is a some non natural features that not natural like that sitram linear predictive coefficient, there represent as be signal, but literally they does not have a meaning. So, if I say the LPC coefficient a 1 a 2 up to a p all represent that speech flame that is parameters not features, but they are non natural features .similarly sitram then MFCC, bogotrax area functions all kind of features are there which in non natural.

So, we discuss about the extraction of sound natural features and non natural features. Now the problem which feature extraction is that what about the features is there the extraction proceed your must below (Refer Time: 09:11); that means, what is (Refer Time: 09:14); that means, if this is the speech signal, if I this time I extract the features if the next time also I extract the features I will get the same features for the same speech signal.

And they should exactly represent the segment. So, sometime these features is very high and if I extracted in next time this features is very high, then there is a some problem. So, the main problem is the I should call those are the speech features which extraction procedure is robust and also computationally less complex. If it is computationally very complex then the feature extraction may required much more time. So, I cannot develop the system which is real time. Any kind of application if you say speech recognition, speech synthesis, figure identification, LPC coefficient, LPC coding or any speech coding, all are nothing but a extract the speech features and using those features I can apply any kind of classification or I can coding or I can synthesis algorithm So that I can reward back the signal.

So, what can say the speech features actually the representation of the original speech signal. Using those features I can develop different technology which can be say as the speech application. So, during this whole week we discussed about the those features extraction. Now I come to that, if you see in this slides or if you see along this speech signal if you see the spectrogram is different in different segment. Or I can saw you in here.

(Refer Slide Time: 11:03)


Segmental

Supra-Segmental

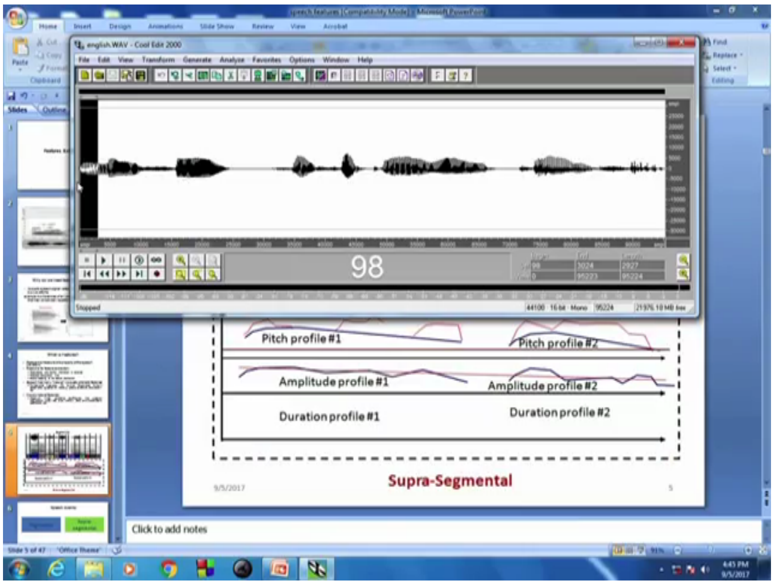
9/5/2017

5

If you see the speech segments properties or signal property of the speech segments are different in along the time. If I say this time this portion is completely silence, this time boganic, this time noising, boganic, noising, silence (Refer Time: 11:30). So, all kinds of speech enhance are there. And if I say the signals space different time the speech segment signals are different.

So, if I extract the features for this small segment, then I can say those are call segmental features as you understand. Or not if I extract the features for the segment then I call segmental features.

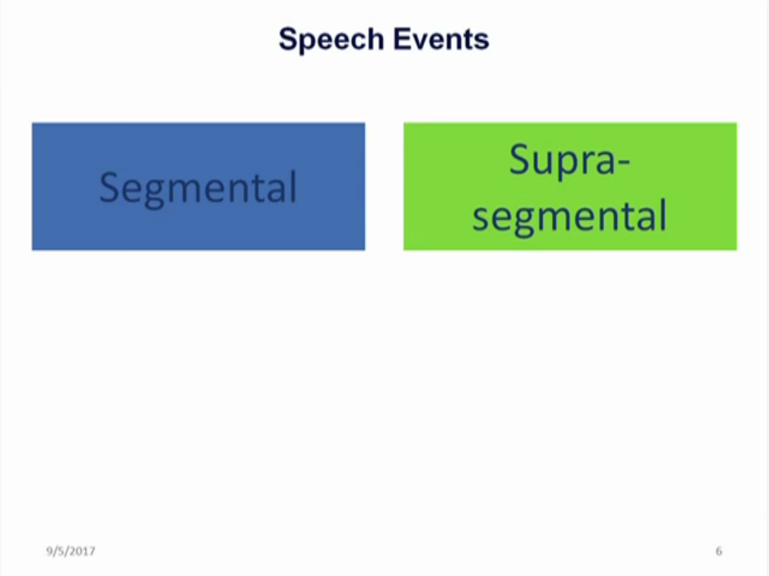
(Refer Slide Time: 12:05)



There may be a some features which very across the segment. Like that f 0 movement, duration of the speech event. So, they are very across the segment. So, I can say those are called supra segmental features. So, so I can say during the best on the extraction I can say 2 types of features, one is call segmental features another is call supra segmental features. So, segmental features which related to particular segment may be a consonant to wall transition is a segmental features, pure occlusion period is a segmental features, bass is a segmental information.

So, if you know features is extracted segment by segment or for a particular segment of the speech signal then I call this is a segmental speech features, if the features is extracted across the segment then I call it is supra segmental speech features. So, if I say segmental speech features all the speech event phoneme may be part of the phoneme also segmental speech features occlusion period (Refer Time: 13:20) consonant to hall transition. So, all are segmental features. If it is supra segmental features movement of the fundamental frequency change of duration of phonemes or syllable, and if you say the duration profile amplitude profile change of amplitude across the whole speech is also a supra segmental speeches ok.

(Refer Slide Time: 13:50)



Speech Events

Segmental

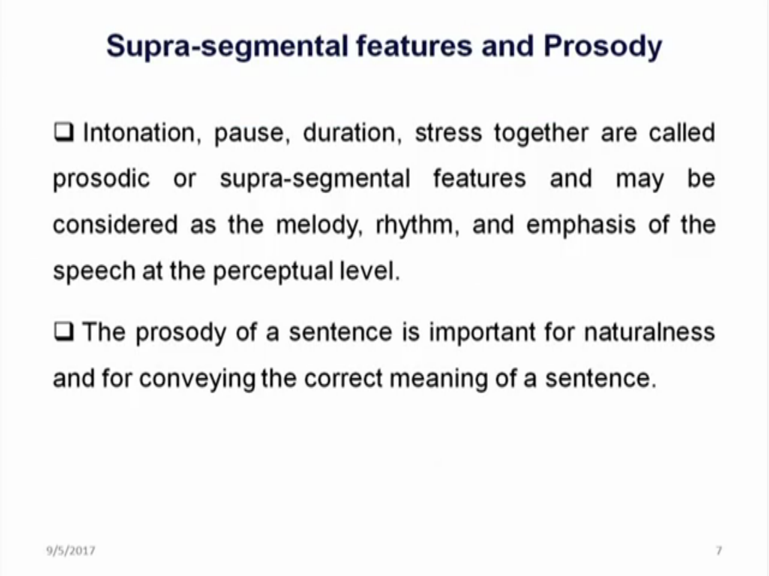
Supra-segmental

9/5/2017

6

So, speech features are basically 2 type, one is call segmentals features another is called supra segmentals features. So, f 0 duration, amplitude profile, are the supra segmental features segmental features I can say phonemes or part of a phoneme is also a segmental features, like occlusion period (Refer Time: 14:05) if I able to extract the (Refer Time: 14:07) occlusion period BOT all are segmental information.

(Refer Slide Time: 14:15)



Supra-segmental features and Prosody

❑ Intonation, pause, duration, stress together are called prosodic or supra-segmental features and may be considered as the melody, rhythm, and emphasis of the speech at the perceptual level.

❑ The prosody of a sentence is important for naturalness and for conveying the correct meaning of a sentence.

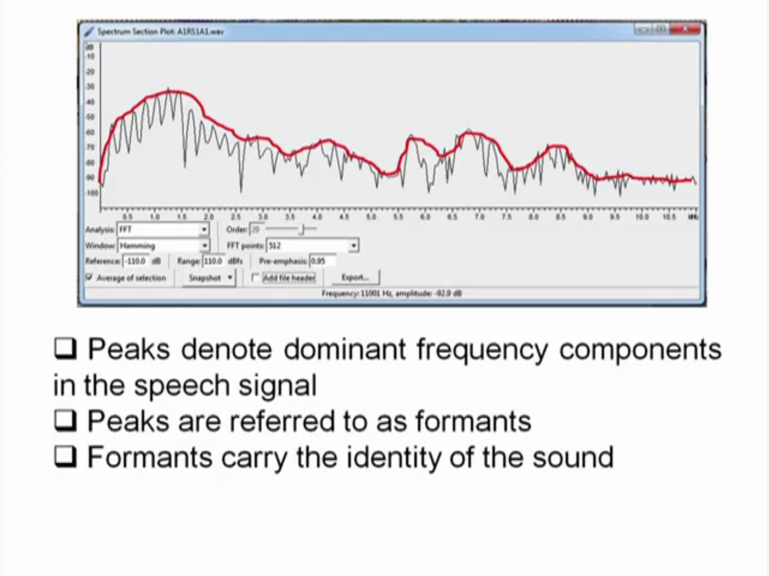
9/5/2017

7

Now, how do we extract the I will come supra segmental features I will come later on the intonation pause duration stress all are call supra segmental features how they extracts.

So, will discuss about the 2 kinds of methodology, one is segmental features extraction another is supra segmental features extraction. And supra segmentals features are use to model speech prosody now consolidated segmental features ok.

(Refer Slide Time: 14:40)

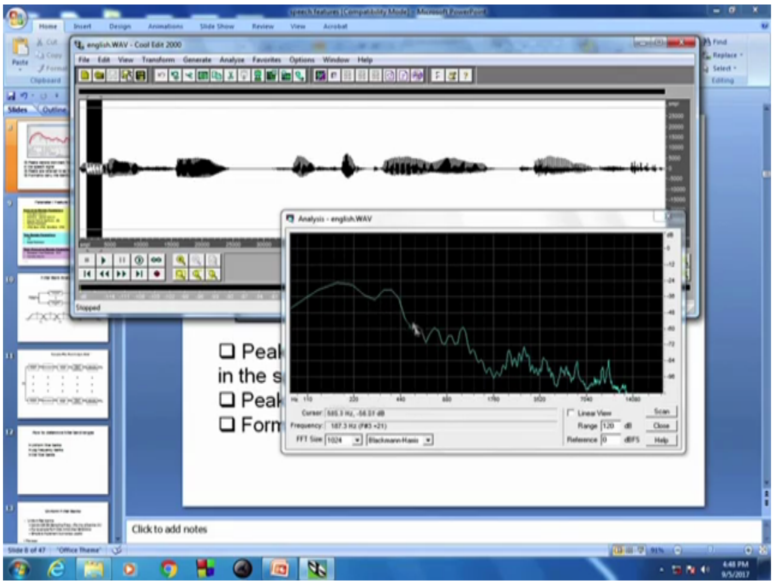

Spectrum Section Plot: A1RS1A1.wav
dB
-10
-20
-30
-40
-50
-60
-70
-80
-90
-100
0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0 6.5 7.0 7.5 8.0 8.5 9.0 9.5 10.0 10.5 kHz
Snapshot
Export...

Analysis: FFT
Order: 20
Window: Hamming
FFT points: 512
Reference: -110.0 dB
Range: 110.0 dBfs
Pre-emphasis: 0.95
☑ Average of selection
☐ Add file header
Frequency: 11001 Hz, amplitude: -92.9 dB

- ❑ Peaks denote dominant frequency components in the speech signal
- ❑ Peaks are referred to as formants
- ❑ Formants carry the identity of the sound

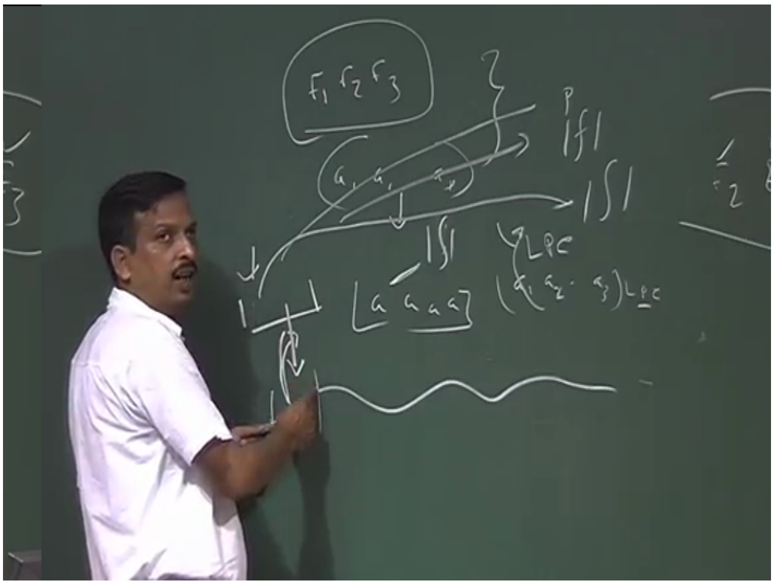
If you see if I show a segment of this speech signal that I this segment and if I analyze the frequency see this is the frequency spectrum of that segment. So, in frequency government the speech is represented in a frequency government of this segment.

(Refer Slide Time: 15:05)


english.WAV - Cool Edit 2000
File Edit View Transform Generate Analyze Favorites Options Window Help
Analysis - english.WAV
❑ Peak
in the s
❑ Peak
❑ Form
Click to add notes

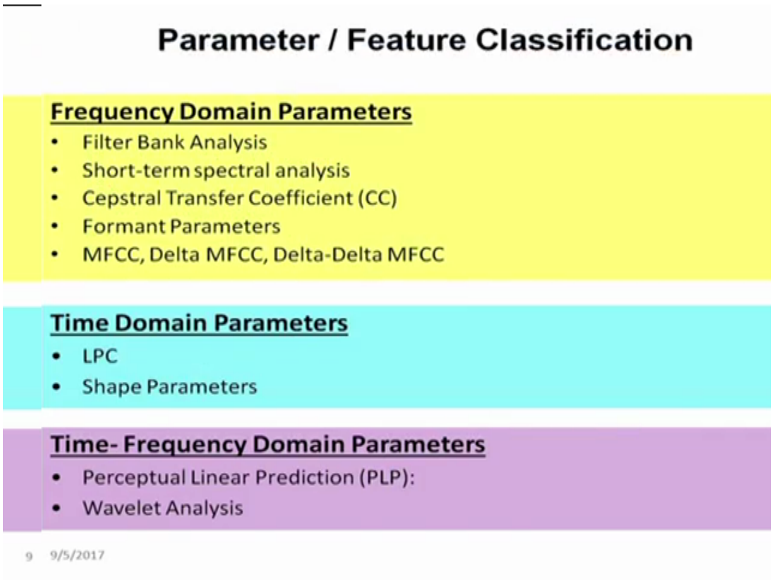
Now, if you see there is a movement of the formant, if I choose small segment, there is a movement of the formant. So, you can say the peak denotes the frequency component of the signals. So, if you see this slides there is a this is the spectrum information. And if you see the if the red line is the envelope of the spectrum. And those peak? Peak of the envelope denotes the formant. So, if I say formant is my features. So, I am to extract those formant, how do I find out those formant? So, I have to find out an algorithm by which I can robustly extract the formant frequency ok.

(Refer Slide Time: 16:09)



So, there is a I required an algorithm by which I can extract all the formant frequency f 1 f 2 f 3 f 4. One of the methods is that you know that if there is LPC coefficient a 1 to a p if I take the frequency task form of those LPC coefficient they are actually represent the formant position of the signal. So, formant frequency are extracted using the LPC formant LPC spectrum analysis also. So, I want an algorithm by which I can extract the formant frequency then is call formant frequency extraction, and those formant frequency for a particular speech segment.

(Refer Slide Time: 16:57)



Parameter / Feature Classification

Frequency Domain Parameters

- Filter Bank Analysis
- Short-term spectral analysis
- Cepstral Transfer Coefficient (CC)
- Formant Parameters
- MFCC, Delta MFCC, Delta-Delta MFCC

Time Domain Parameters

- LPC
- Shape Parameters

Time- Frequency Domain Parameters

- Perceptual Linear Prediction (PLP):
- Wavelet Analysis

9 9/5/2017

Now, if you see in parameter domain there is a lot of speech parameters or features one is call filter bank. So, features parameter extraction methods lot of like LPC we have discuss LPC parameter extraction method we have discussed. So, there is a frequency domain parameters there may be a time domain parameters. So, features can be extracted from the time domain representation of the speech signal.

So, if this speech signal is this. So, I from the time domain representation I can extract the features then I call time domain features from the frequency domain representation from this representation. If I extract the features then I call frequency domain parameters. So, best on this analysis we can say there is 3 types of parameter frequency domain. Parameter extraction or 3 types of parameter extraction methods one is call frequency domain parameter extraction methods time domain parameter extraction method time frequency both parameter extraction methods and extracted parameter are call parameters.

So, if the extract LPC parameter extraction extracted using LPC, LPC analysis then all coefficient a 1 a 2 a 3 all are cal LPC parameters. So, those represent the speech signal for a particular speech segment. So, if the feature is extracted segment wise. So, it is a call features or the extracted segment wise, that I will come later on why it is. Then that is a call frequency domain parameter filter bank analysis short term spectrum analysis spectral analysis formant parameters MFCC delta MFCC all kinds of parameters are call frequency parameter time. Domain parameter LPC say parameter frequency time domain parameter perceptual linier prediction and oblate analysis ok.

So, all parameter extraction methodology let us which try to explain one by one. So, learn how to extract the all kinds of speech parameters. Now if I say this is the whole speech signal, I say the parameters are extracted segment wise. Now if I say the phoneme or I say phoneme is a segment. How do we define the phoneme boundary in a continuous speech. Manually I can listen and I can find up to here to here it is a [FL] kind of signal. But think about I want the extraction method by which automatically I want to extract what kind of phoneme it is. So, instead of extract that is very difficult. So, instead of extracting that what kind of phoneme it is, I want the some kind of representation of the signal which is unique in certain domain.

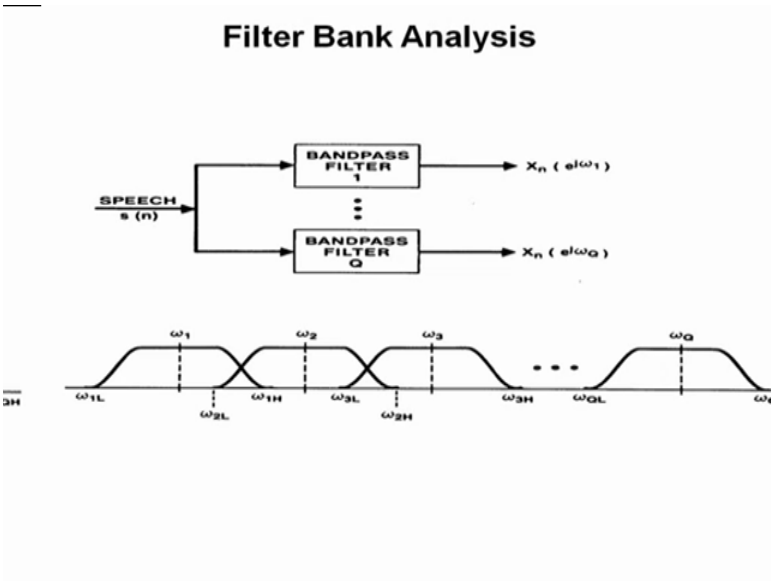
So, I can say probably if this is this kind of parameter extraction, and those parameters belong to particular phoneme may be (Refer Time: 20:18). So, although the extraction of phoneme or part of a phoneme is a meaningful features or you can natural features of the speech signal, but detecting the boundary of a phoneme is very difficult in continuous speech. So, how to I can find out a phoneme (Refer Time: 20:41) or how do I can find out phoneme (Refer Time: 20:43).

So, instead of finding out phoneme sh, let us blindly analyze the speech signal frame by frame, and assign each frame to a particular phoneme. Or I can say the extracted parameter represent this frame later on I try to classify whether this frame belongs to (Refer Time: 21:08) or (Refer Time: 21:08) or (Refer Time: 21:09). So, that is my job. So, I can say the features are extracted segment wise, and they are the representation of that segment either it can time domain or it can be frequency domain. So, if those features are extracted frequency representation of that segment then I call frequency domain parameters. If they are extracted from the time domain representation of the signal then I call it is a time domain parameters ok.

So, speech features natural features like phoneme we are not extracted really. So, what we are doing we try to design an algorithm, by which segment or you can say the framing I done the framing of the speech signal frame by frame speech signal will be analyze. And either time domain or frequency domain parameter will be extracted which will represent exactly that frame of the speech signals ok.

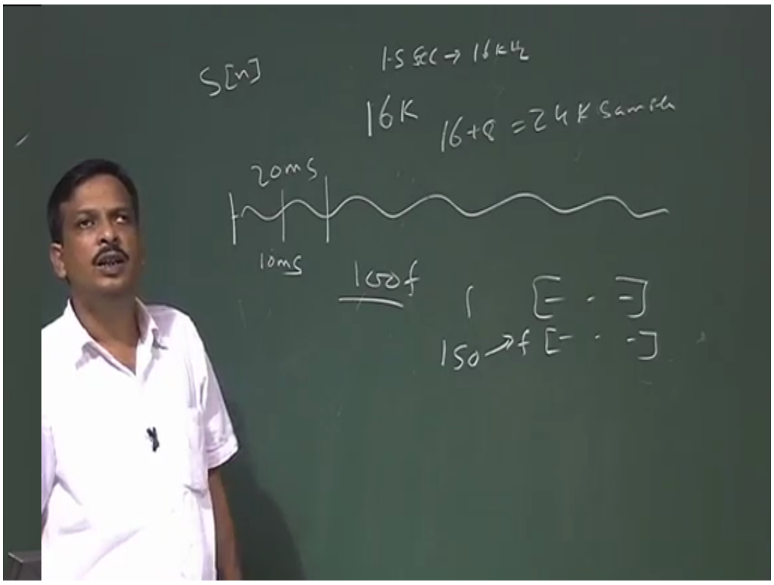
So, that is my job. Suppose I have a speech of fast discuss about the filter bank analysis. What is filter bank analysis? Come here. So, where discuss about the filter bank analysis.

(Refer Slide Time: 22:44)



Filter Bank Analysis

(Refer Slide Time: 22:48)

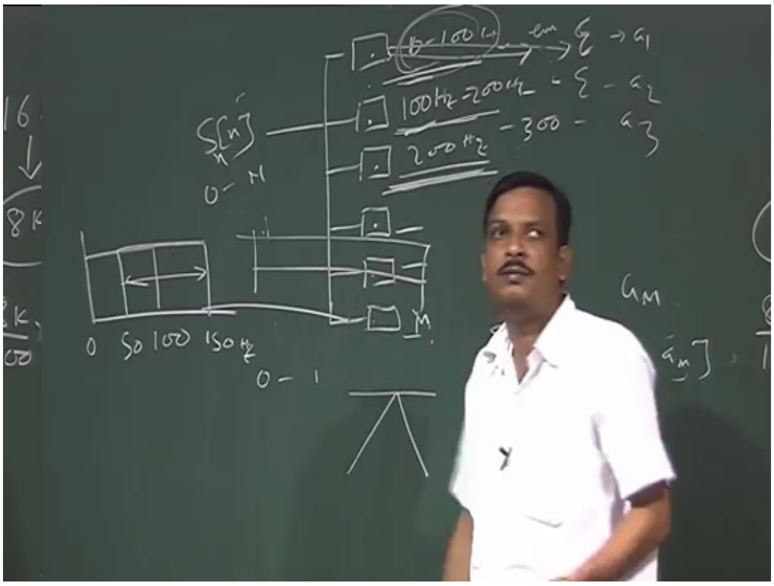


So, suppose forget out the slides let us s n is my signal speech signal recorded speech signal. So, let us I give real example let us I recorded my name and my name has consist of the signals of let us 1.5 second and that signal is recorded 16 kilo hertz. So, I can say in one second there is a I can say there is a 1 second 16 k sample is there. So, 1.5 second or I can say one and half second. So, there will be a 16 plus 8 is equal to 24 k sample and k sample. If you see if I recorded my name. So, different speech event has different kind of speech signal.

So, I am not analyzing whole speech signal at a time and represent by a single vector. So, instead of doing that if I do that then time resolution, I completely loss, as we discussed in the frequency analysis. So, what I will do instead of analyzing that I will hold that this is my whole signal and I frame the signal. So, I can say I window the signal. So, I can say let us take 20 minute second window, and shifted the window by 10 minute second. So, I can say in a 1 second I will get 100 frame 100 such frame 10 minute second 10 minute second 10 minute second. So, 100 such frame and each frame from each frame I get a representative feature vector that is my job? So, in 1 second or we can 1.5 second I can get at least 150 frame and each frame is represented by the some feature vector ok.

Let us those features are let us those features are extracted using filter bank analysis. So, what is filter bank analysis? If I say so, whole speech signal let us s n is there ok.

(Refer Slide Time: 25:13)



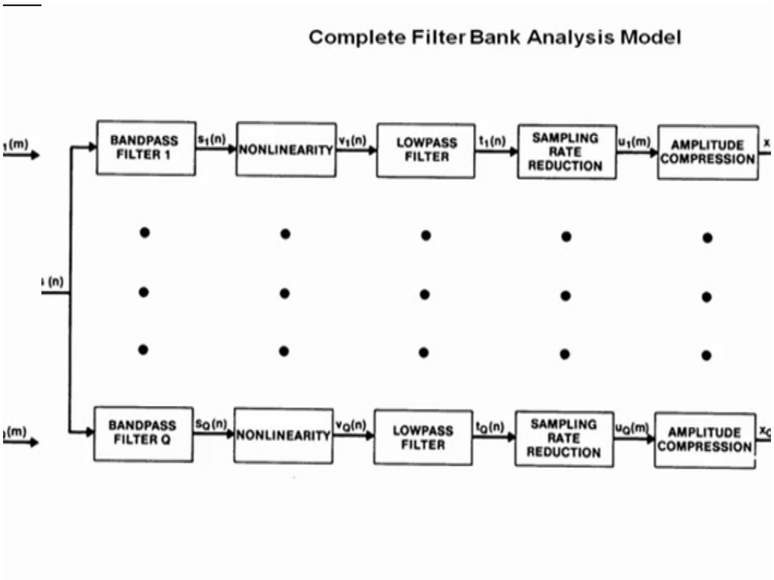
Let this is by S n n which is the n at frame of the speech signal. This would be pass through a some parallel filter, I can pass that signal with a parallel filter, some parallel filter and each filter has a passband. So, let us this is 100 0 to 100 hertz. This is again I said 100 hertz. So, there is a n1 overlapping filter there may be let us n1 overlapping 100 hertz to 200 hertz, then 200 hertz to plus 300 hertz. So, each filter has a particular passband. So, from the each passband I will get the power if that particular frequency component is there in the speech signal.

So, if there is a any component 0 to 100 hertz. So, output of the filter is nothing but a 0 to 100 hertz representation of the signal. That is 100 hertz to 200 hertz signal, that is 200 hertz to 300 hertz signal. So now, if I quantize the let us the power let what about the frequency component is there let us sum their power and take as a parameter a 1. So, I can take those sum as a parameter of a 2 those sum a 3 let us there is a m number of filters. So, I can say up to a m. So, this a 1 a 2 a 3 and a m represent the feature vector which is extracted using the filter bank analysis, I am I clear.

Now, how to design those filters are particular nothing but a bank was filter because there particularly particular band is pass if the whole signal passes through this filter. So, 0 to 100 hertz band will be pass, and what is the collected the passband signal? I find out the some energy. Energy of 0 to 100 hertz signal. So, I can say 0 to 100 hertz signal what about the energy is there I can sum them or I can take the average energy of the 0 200 hertz. So, that energy represent one parameter. So, a m parameter a m number of parameters are there which is designed by a filter bank analysis is clear. Now those filters design of those filter has a different methodology. So, I can design them either overlapping I can design them n1 overlapping (Refer Time: 28:13) mirror filter not kind of design of those filter can be done.

And whatever I design the output is call filter bank analysis of the speech signal.

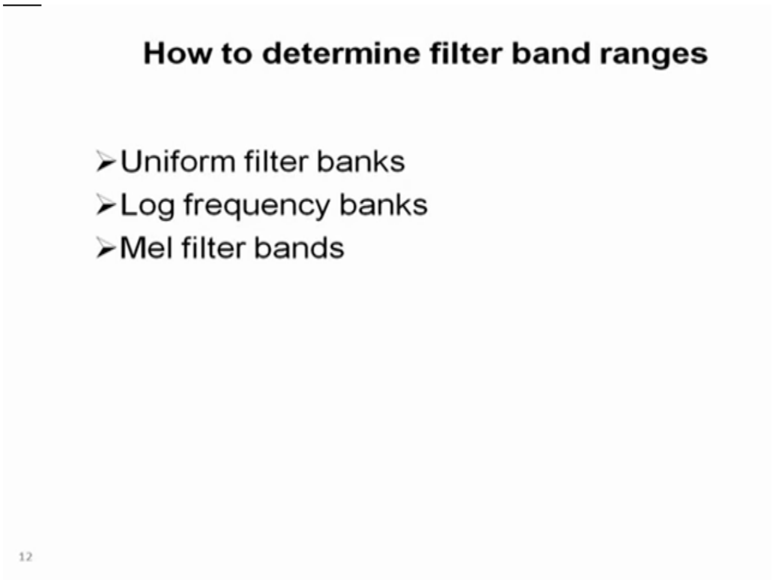
(Refer Slide Time: 28:29)


Complete Filter Bank Analysis Model

So, each of the filter out put if this is the complete filter bank analysis model, if you see the filter pass through a bandpass filter. So, this is the passband second band third band upto q number of band is there. So, if I say I have a signal call let us I have a 16 kilohertz sampling frequency. So, what is the basement frequency. Basement frequency signal is 8 kilo hertz. So, my speech signal has a component up to 8 kilo hertz. So, if I say I am to design 100 hertz bank filter. So, how many filter will be there every filter is 100 hertz n1 overlapping condition. So, I can say 8 k divided by 100 hertz. So, I can say it is nothing but a 8000 divided by 100 hertz. So, I can say there is a 80 filter will be there.

So, q is equal to 80. So, 80 bandpass filter will bet there. And each output filter then there is nonlinearity I can pass through a nonlinearity and low pass filter and then sample reduction, and then amplitude completion and I get a parameter x 1. So, that each filter output I can defined is filter output if the signal is there will filter output will be 0 to n number of let us see though signal content each frame at 0 to n number of sample each filter output will be 0 to n number of sample. So, that output I can quantize or I can find out the average energy. So, that is my parameters. So, this is the complete filter bank analysis.

(Refer Slide Time: 30:15)



How to determine filter band ranges

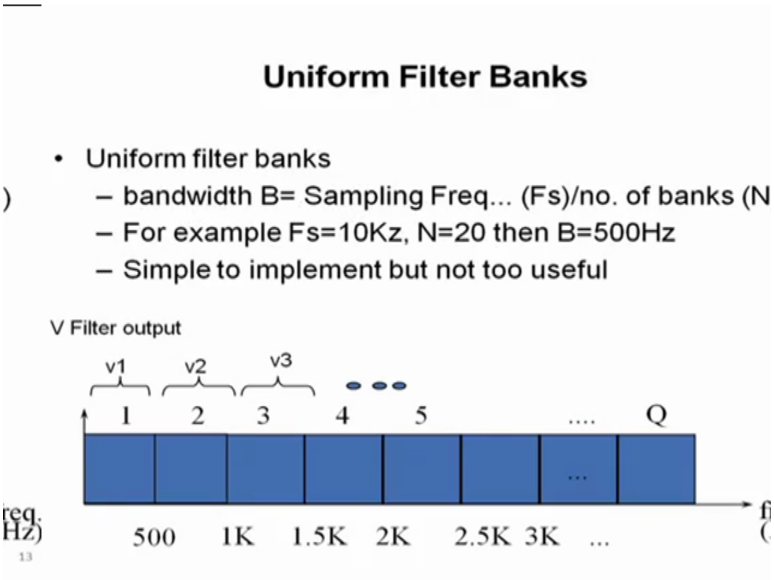
- ➢Uniform filter banks
- ➢Log frequency banks
- ➢Mel filter bands

12

Now, how to design those filter? I can uniformly uniform filter when design I can say every filter is 100 hertz and there is a no overlapping region. I can say every filter band width will be determine by a long frequency scale. Or I can say every filter band width will be determine by a mel scale, which is the perceive frequency by human with the scale of perceive of human human being frequency. So, how the human being perceive frequency in mel scale. So, I can design those filter bank filter band width in a mel scale, or I can say those are the overlapping filters. So, I can say the filters are design using 50 percent overlap. So, if this is my fastband 0 to 100 hertz next till there will be 50 hertz to 150 hertz ok.

So, then I can say there 50 percent overlap filter bank. Then I can say the bandwidth of the filter is not linier. So, not always 100 hertz in the low frequency there everywhere 100 hertz, but the high frequency bandwidth may be larger, and that bandwidth may be determine by mel frequency mel scale. So, all kind of things we can use and we can design the filter and find out the filter bank parameters.

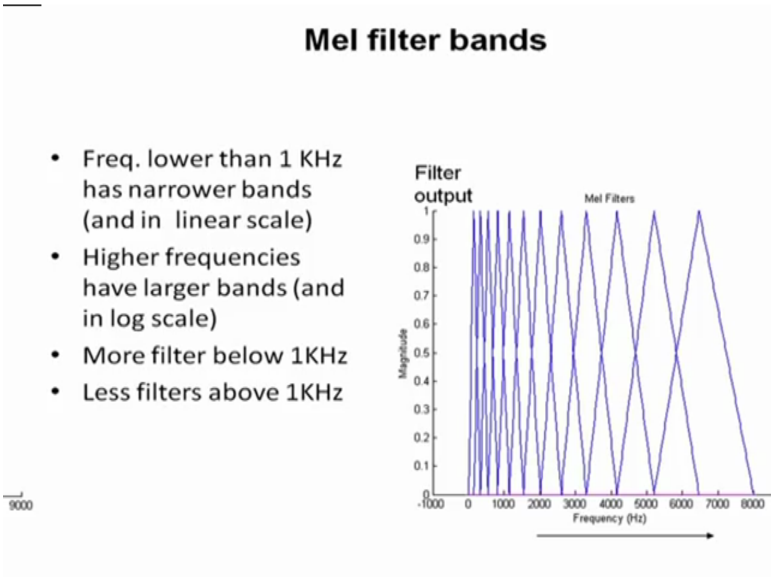
(Refer Slide Time: 31:52)



Uniform Filter Banks

- Uniform filter banks
  - bandwidth B= Sampling Freq... (Fs)/no. of banks (N
  - For example Fs=10Kz, N=20 then B=500Hz
  - Simple to implement but not too useful

(Refer Slide Time: 31:59)



Mel filter bands

- Freq. lower than 1 KHz has narrower bands (and in linear scale)
- Higher frequencies have larger bands (and in log scale)
- More filter below 1KHz
- Less filters above 1KHz

So, like this uniform filter bank then I can go for the non linier log frequency filter bank, then I go for the mel scale filter bank. And structure of the filter also I do not want this is the flack, I can want there will be triangular filter or instead of rectangular filter I can say there a rectangular filter, I find out the energy within that triangle. So, what about I design here which is call filter bank analysis. So now stop here. So, next lecture will go for the another methods for parameter extraction ok.

Thank you.