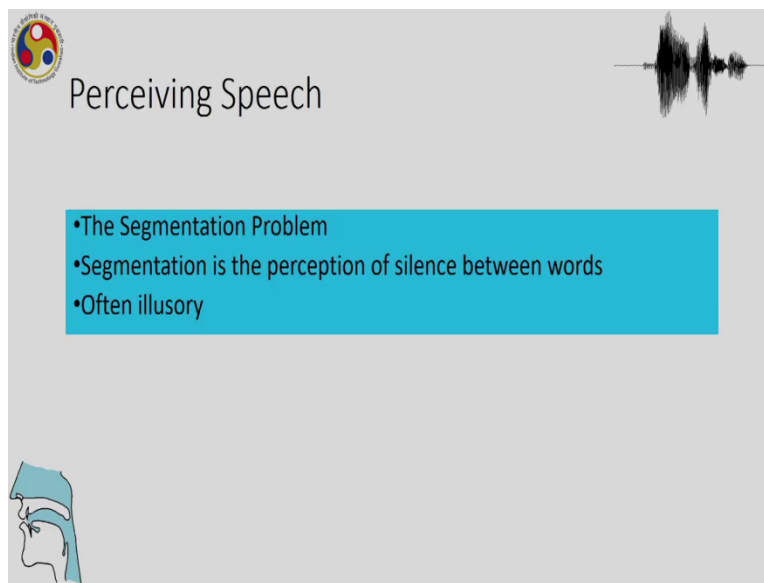


Phonetics and Phonology: A Broad Overview
Professor Shakuntala Mahanta
Department of Humanities and Social Sciences
Indian Institute of Technology, Guwahati
Lecture 13
Segmentation Problem, Categorical Perception
CP in Infants and Animals

Welcome to the NPTEL MOOC course on Phonetics Phonology, A Broad Overview. So, I am continuing with unit four that is speech perception.

(Refer Slide Time: 00:42)



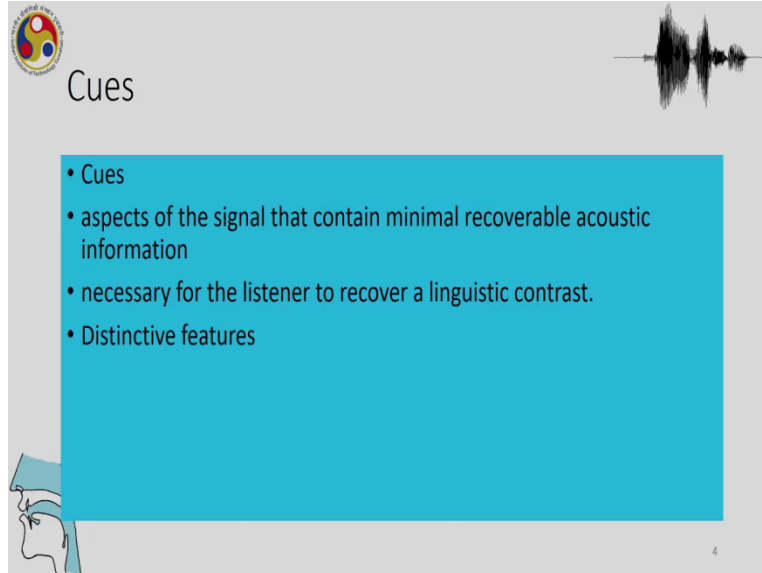
The slide features a grey background. In the top left corner is the NPTEL logo. In the top right corner is a black waveform representing a speech signal. The title 'Perceiving Speech' is centered at the top. Below the title is a blue rectangular box containing three bullet points. In the bottom left corner, there is a small illustration of a human head in profile, showing the vocal tract.

Perceiving Speech

- The Segmentation Problem
- Segmentation is the perception of silence between words
- Often illusory

And in the last class we talked quite a bit about perceiving speech and the segmentation problem, that the segmentation is the perception of silence between words.

(Refer Slide Time: 00:58)



The slide features a logo in the top left corner, a waveform in the top right, and a profile of a human head in the bottom left. The main content is a blue box with the following text:

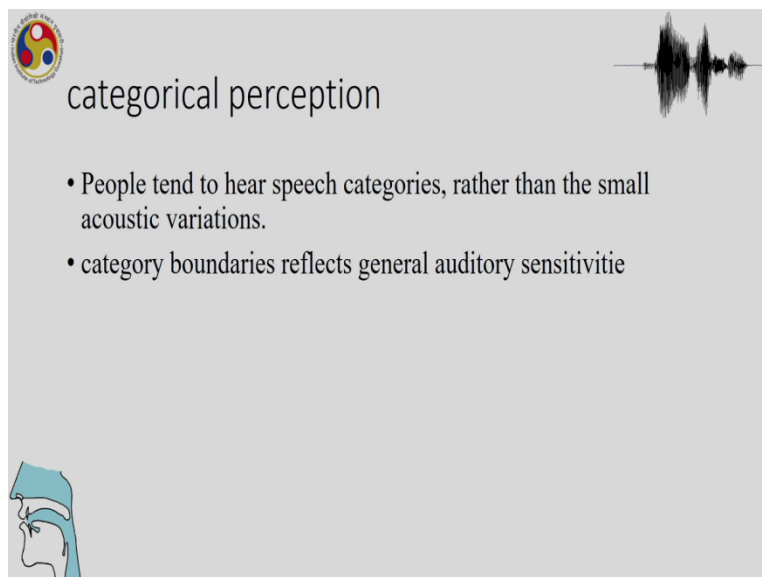
Cues

- Cues
- aspects of the signal that contain minimal recoverable acoustic information
- necessary for the listener to recover a linguistic contrast.
- Distinctive features

4

And how we use cues, how human beings use cues or aspects of the signal that contain minimal recoverable acoustic information, necessary for the listener to recover a contrast.

(Refer Slide Time: 01:13)



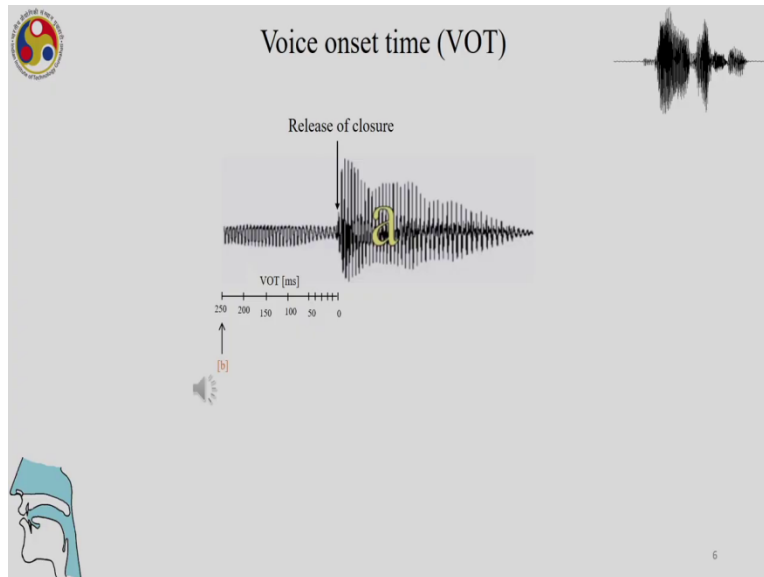
The slide features a logo in the top left corner, a waveform in the top right, and a profile of a human head in the bottom left. The main content is as follows:

categorical perception

- People tend to hear speech categories, rather than the small acoustic variations.
- category boundaries reflects general auditory sensitivities

And also this categorical perception. So, people that we hear of speech categories rather than the small acoustic variations, category boundaries, therefore, reflect general auditory sensitivities.

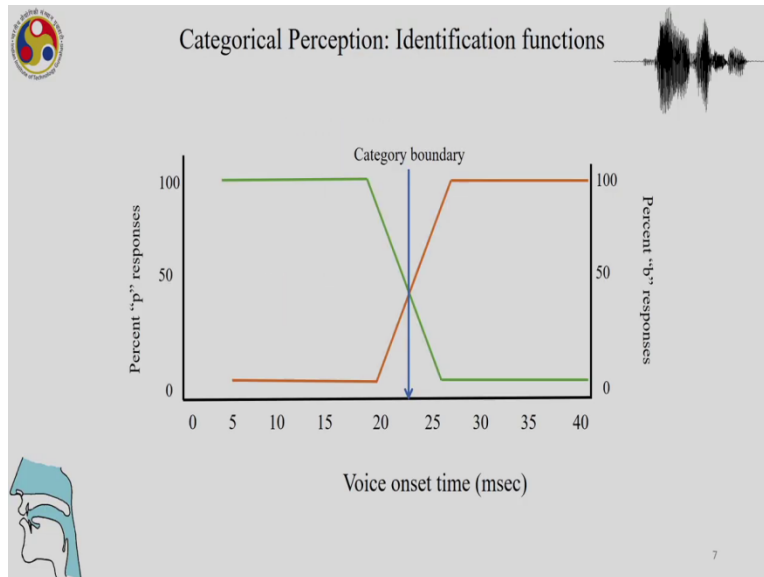
(Refer Slide Time: 01:28)



And also we looked at voice onset time, which is one of the most important things that you will study when you study categorical perception. So, here, we are showing VOT release of closure of the consonant and when voicing starts. So, depending on when the voicing will start, you will hear a consonant as voice or voiceless.

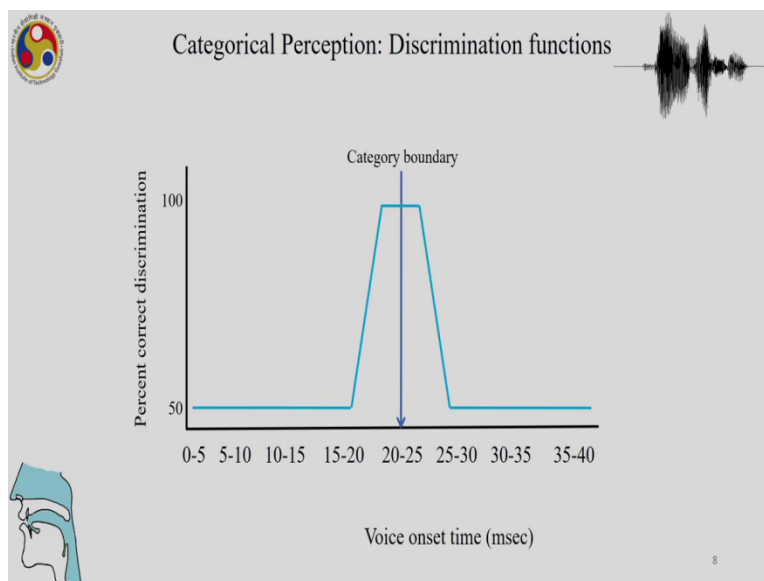
And you saw in the last class, that if it is, there could be between 20 and 25 milliseconds in this experiment, there could be a category change, and that is, and the experimental results shows that almost 100 percent of the subjects gave a positive response within that category boundary, and they will never hear an in between category, and they would hear only one category or the other category and nothing in between.

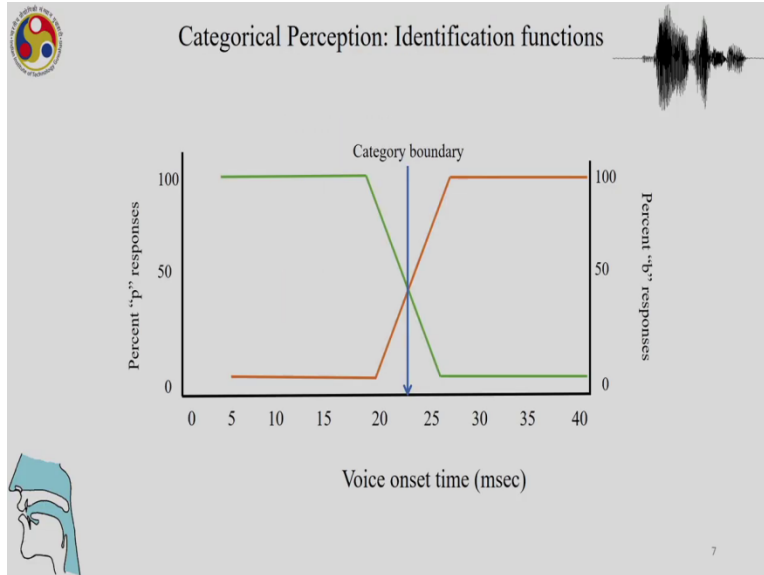
(Refer Slide Time: 02:36)



So, this is what we saw in the last class.

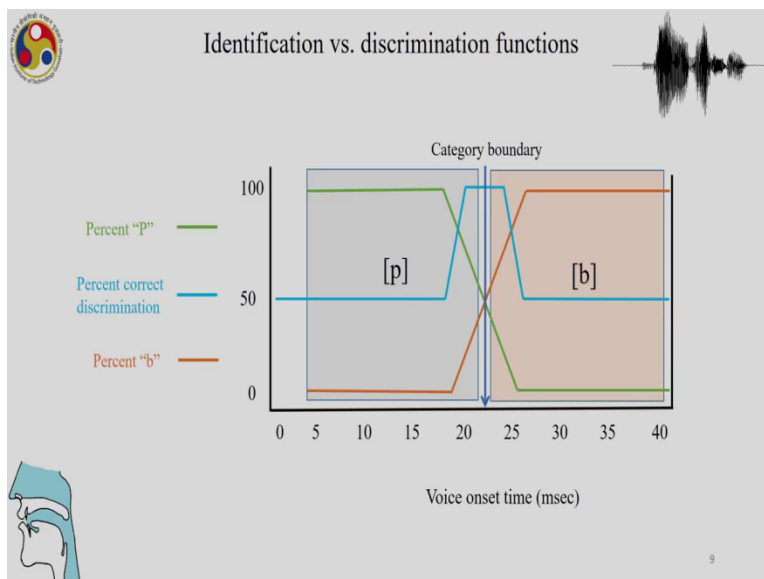
(Refer Slide Time: 02:41)





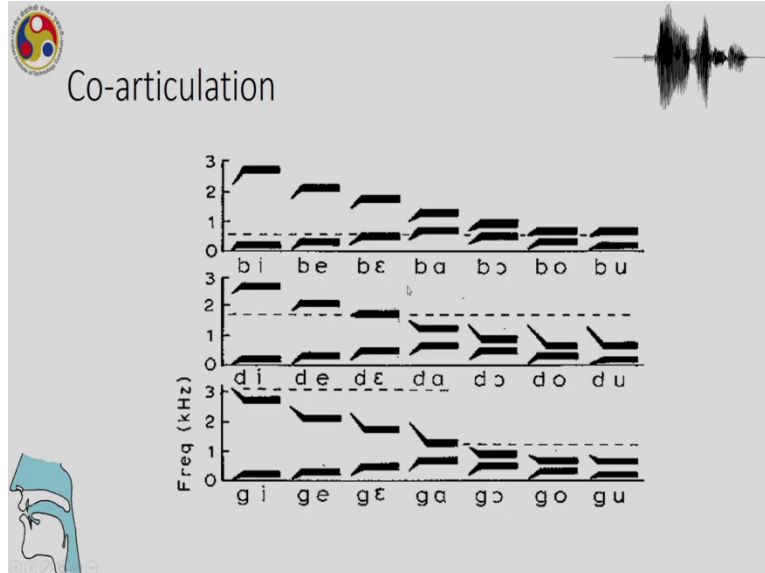
So, we saw identification, and also saw discrimination. So in identification, as we said, In the last class, this is false identification. So, participants are asked to identify based on orthography, whether something is ba or pa, but in discrimination functions, it is not based on orthography. So, speakers hear a sound, and would have to identify something as that sound, or another sound.

(Refer Slide Time: 03:17)



And also, these, both these experiments merge in the findings of a category boundary. So, it is either ba on one side and pa on the other side, but speakers never hear an in between category.

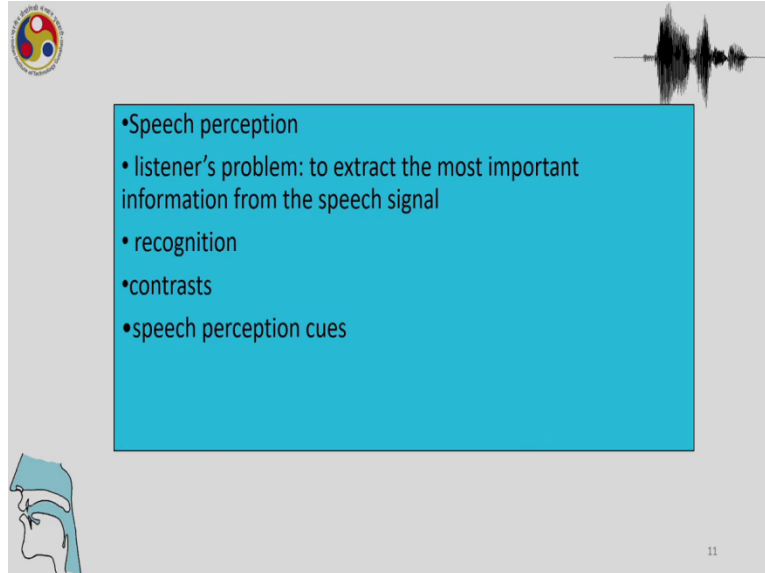
(Refer Slide Time: 03:39)



And apart from voice onset time and categorical perception, we also saw co articulation how, form and transitions can change based on the following vowels. And these are three consonants. And we saw that irrespective of whether it is the same consonant or not, depending on the following vowel transitions could be different. So, a ba has different transition from a b, and similarly, so across consonants also, b is different from gi and gi is different from di.

So, both across vowels and across consonants, there are differences in form and transitions. So, we have a lot of variability in the signal. And it seems people are sensitive to these variations.

(Refer Slide Time: 04:31)

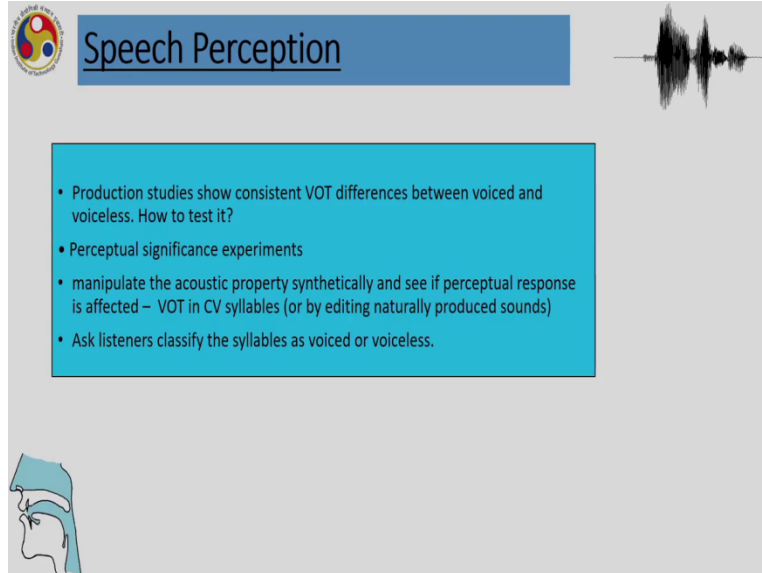


- Speech perception
- listener's problem: to extract the most important information from the speech signal
- recognition
- contrasts
- speech perception cues

So, speech perception essentially is, it is a problem, it can be called, the listener's problem. So, how do you extract the most important information from the speech signal, from the acoustic signal how, what is the process which is responsible for extracting the most crucial information. And it involves recognition of words, and because that extraction will lead to the differences between words recognition of the words and also help in making the contrast between two sounds, ba pa da ga da pa da, dhe tha etcetera.

And a lot of the speech research, speech perception research has shown that speech perception cues are the ones which help people to make those decisions.

(Refer Slide Time: 05:25)



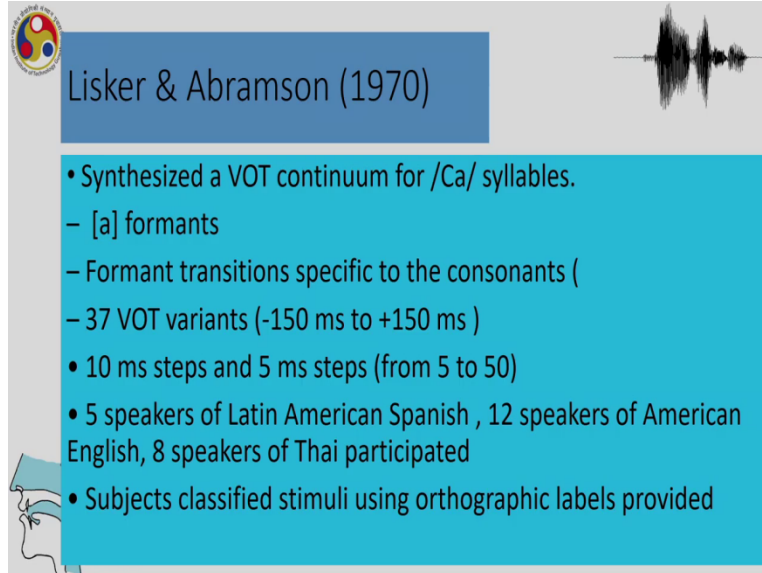
The slide features a logo in the top left corner, a waveform in the top right, and a profile of a human head in the bottom left. The main content is a blue box with a list of bullet points.

Speech Perception

- Production studies show consistent VOT differences between voiced and voiceless. How to test it?
- Perceptual significance experiments
- manipulate the acoustic property synthetically and see if perceptual response is affected – VOT in CV syllables (or by editing naturally produced sounds)
- Ask listeners classify the syllables as voiced or voiceless.

So, production studies show consistent VOT differences between voiced and voiceless. And as we saw in the first few slides, that between 20 to 25 milliseconds and that is also shown in production studies. So, how is it tested? Perceptual significance experiments manipulate the acoustic property synthetically or edited and see if perceptual responses affected. So, VOT, suppose VOT in CV syllables and it can also be done by naturally editing produce sounds and listeners are asked to classify the syllables as voiced or voiceless.

(Refer Slide Time: 06:11)



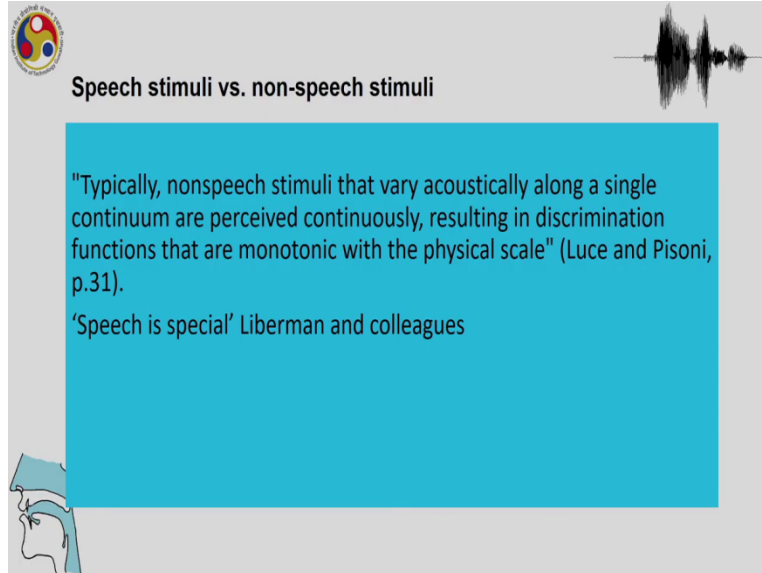
Lisker & Abramson (1970)

- Synthesized a VOT continuum for /Ca/ syllables.
 - [a] formants
 - Formant transitions specific to the consonants (
 - 37 VOT variants (-150 ms to +150 ms)
 - 10 ms steps and 5 ms steps (from 5 to 50)
 - 5 speakers of Latin American Spanish , 12 speakers of American English, 8 speakers of Thai participated
 - Subjects classified stimuli using orthographic labels provided

So, the Lisker and Abramson experiment first showed that synthesized VOT continuum for consonant and a syllables and were in a stable state formant region for a was selected and formant transitions related to the consonants were added and 37 VOT variants were given to participants to hear. So, there were 10 milliseconds steps as well as 5 milliseconds steps and speakers of Latin American Spanish and 12 speakers of American English and 8 speakers of Thai participated.

And subjects classified stimuli using orthographic labels as we had discussed. The wave force identification experiments are conducted based on orthographic labels where participants are asked whether a sound is ba or a pa.

(Refer Slide Time: 07:29)



The slide features a grey background. In the top left corner is a circular logo with a yin-yang symbol and the text 'SCHOOL OF PSYCHOLOGY'. In the top right corner is a black waveform representing a speech signal. The main content is a large blue rectangular box containing text. In the bottom left corner of the slide, there is a small profile illustration of a human head with the ear area highlighted in blue.

Speech stimuli vs. non-speech stimuli

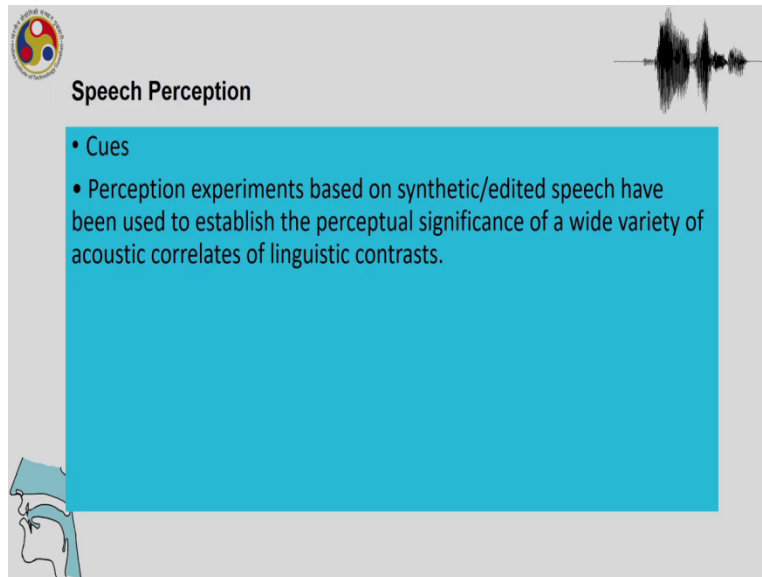
"Typically, nonspeech stimuli that vary acoustically along a single continuum are perceived continuously, resulting in discrimination functions that are monotonic with the physical scale" (Luce and Pisoni, p.31).

'Speech is special' Liberman and colleagues

So, categorical perception is important because it has been shown that typically non speech stimuli vary acoustically along a single continuum and are perceived acoustic continuously resulting in discrimination functions that are monotonic with a physical scale. However, as Liberman and colleagues had shown speech, especially a lot of research has ensued in perception, research h and has shown a perception, speech perception can also be continuous.

However, this lecture does not cover the greater details of the perception research.

(Refer Slide Time: 08:06)

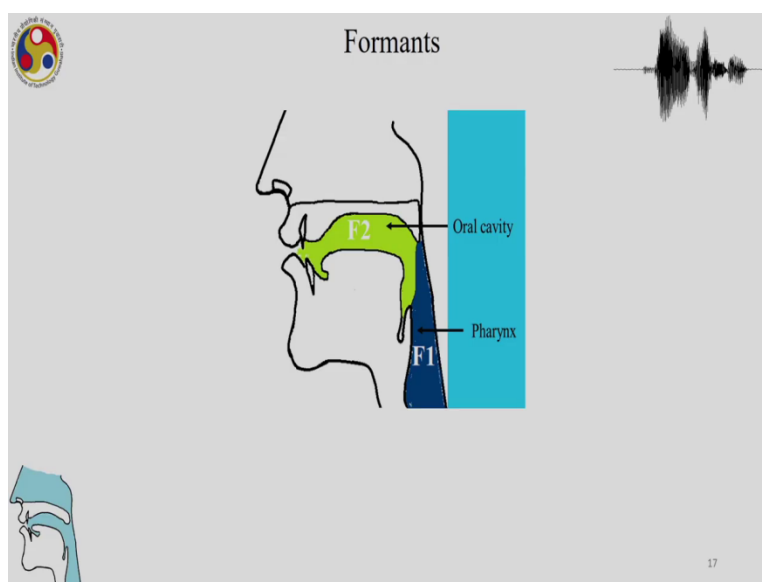


Speech Perception

- Cues
- Perception experiments based on synthetic/edited speech have been used to establish the perceptual significance of a wide variety of acoustic correlates of linguistic contrasts.

So, speech perception research and ability to perceive speech has been shown in the research to be based on cues as we had already said, cues, that is the information which helps us to extract the most important information, so that we can discriminate contrast and so, that we, the word recognition process is optimum. And speech perception experiments based on synthetic edited speech have been used to establish the perceptual significance of a wide variety of acoustic correlates of linguistic contrasts.

(Refer Slide Time: 08:47)



Formants

F2

Oral cavity

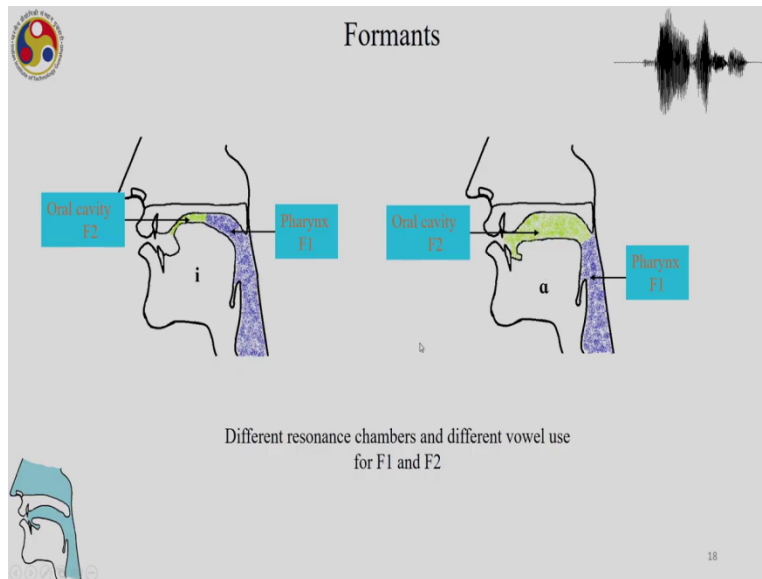
F1

Pharynx

17

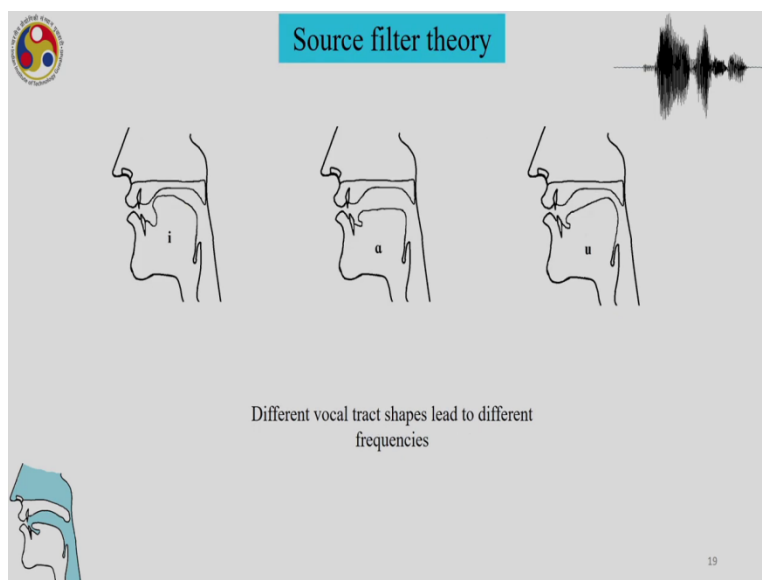
So, recall our lectures on formants, which showed that oral cavity and the pharyngeal cavity, so, changes modulates the speech spectra which is generated from the glottal region.

(Refer Slide Time: 09:04)



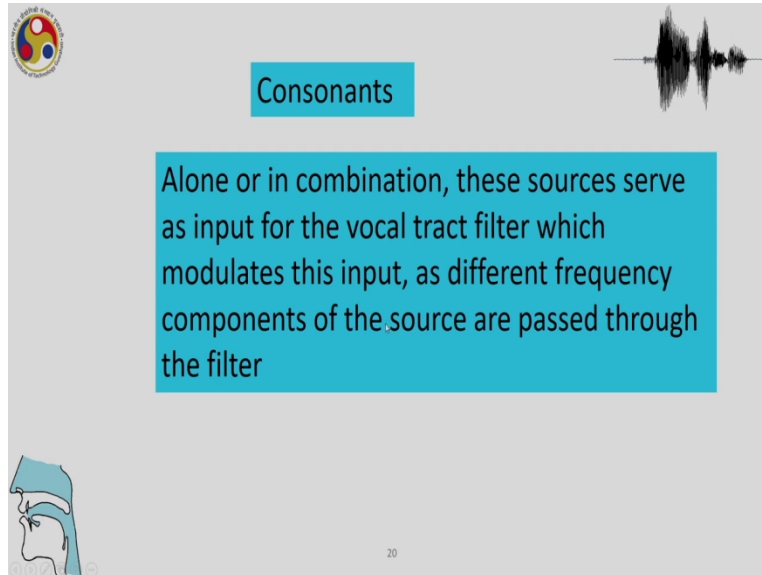
And we have seen all these resonance chambers and how the spectrum is modulated by the vocal tract.

(Refer Slide Time: 09:16)



So, this is the source filter theory, this is how sound is produced and filtered in the vocal tract.

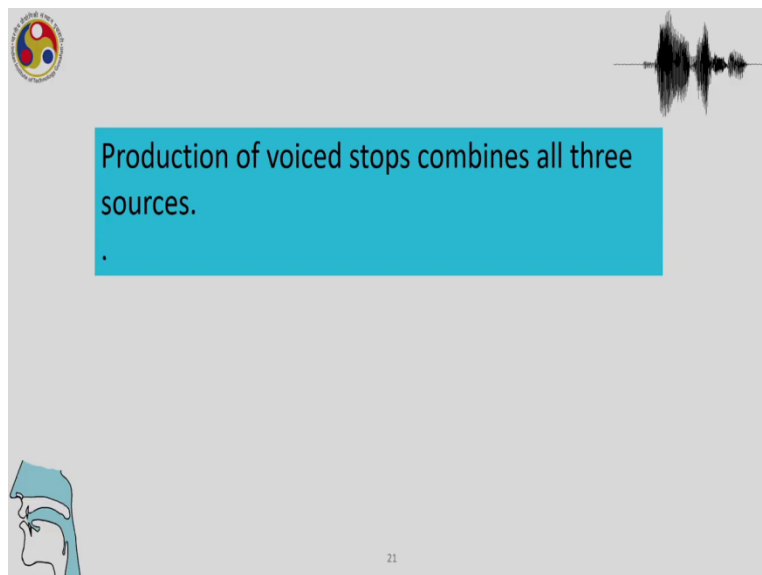
(Refer Slide Time: 09:29)



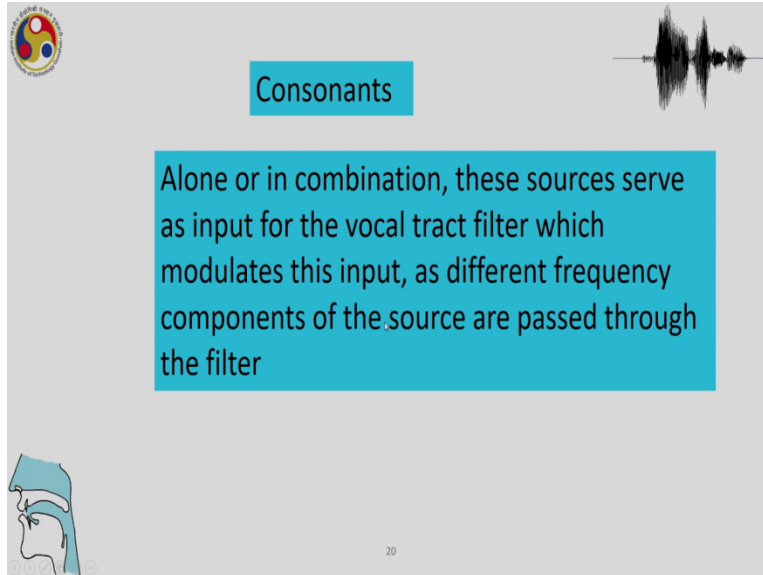
Slide 20 features a logo in the top-left corner, a waveform in the top-right, and a profile of a human head in the bottom-left. The main content is a blue box with the text: "Consonants" and "Alone or in combination, these sources serve as input for the vocal tract filter which modulates this input, as different frequency components of the source are passed through the filter". The number 20 is centered at the bottom.

And after we looked at these things, we talked about how the sources serve as input for the vocal tract filter, which modulates input as different frequency components of the source are passed through the filter.

(Refer Slide Time: 09:39)



Slide 21 features a logo in the top-left corner, a waveform in the top-right, and a profile of a human head in the bottom-left. The main content is a blue box with the text: "Production of voiced stops combines all three sources." and a small dot below it. The number 21 is centered at the bottom.



Slide 20 features a logo in the top left corner, a waveform in the top right, and a profile diagram of the human head in the bottom left. The main text is centered in a blue box.

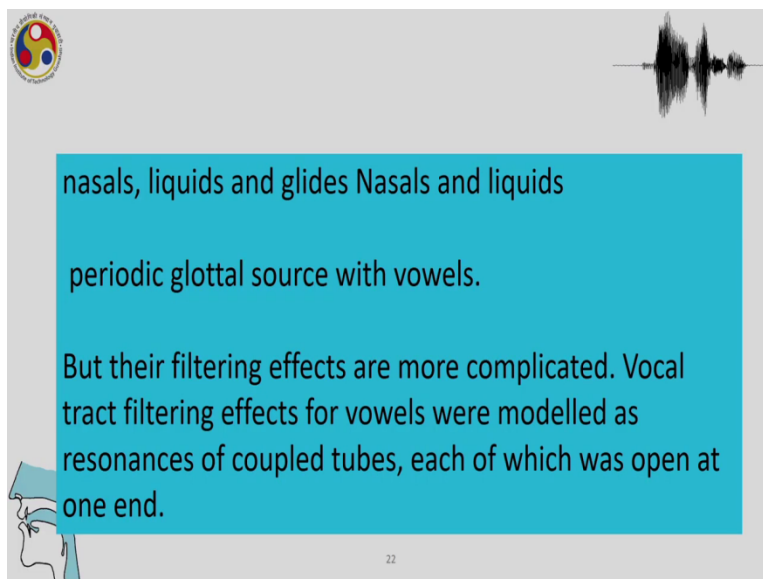
Consonants

Alone or in combination, these sources serve as input for the vocal tract filter which modulates this input, as different frequency components of the source are passed through the filter

20

So, for the production of sounds, all these sources can be combined the glottal source, the vocal tract filter and the frequency components.

(Refer Slide Time: 09:54)



Slide 22 features a logo in the top left corner, a waveform in the top right, and a profile diagram of the human head in the bottom left. The main text is centered in a blue box.

nasals, liquids and glides

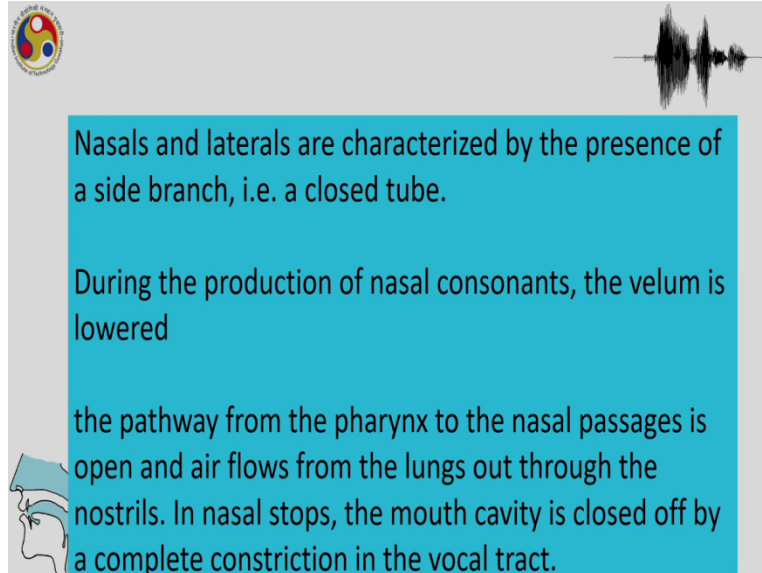
Nasals and liquids
periodic glottal source with vowels.

But their filtering effects are more complicated. Vocal tract filtering effects for vowels were modelled as resonances of coupled tubes, each of which was open at one end.

22

So, and after the combination of all the sources, we have the peculiar properties of nasals, liquids glides and also stops and fricatives. So, nasals, liquids and glides have periodic glottal source with vowels, but they are filtering effects are more complicated vocal tract filtering effects vowels are modeled as resonances of coupled tubes, each of which was open at the end but it is modeled slightly differently for these consonants, it is modeled as closed at one end.

(Refer Slide Time: 10:26)



The slide features a logo in the top left corner, a waveform in the top right, and a sagittal cross-section of the human head in the bottom left. The main text is contained within a light blue rectangular box.

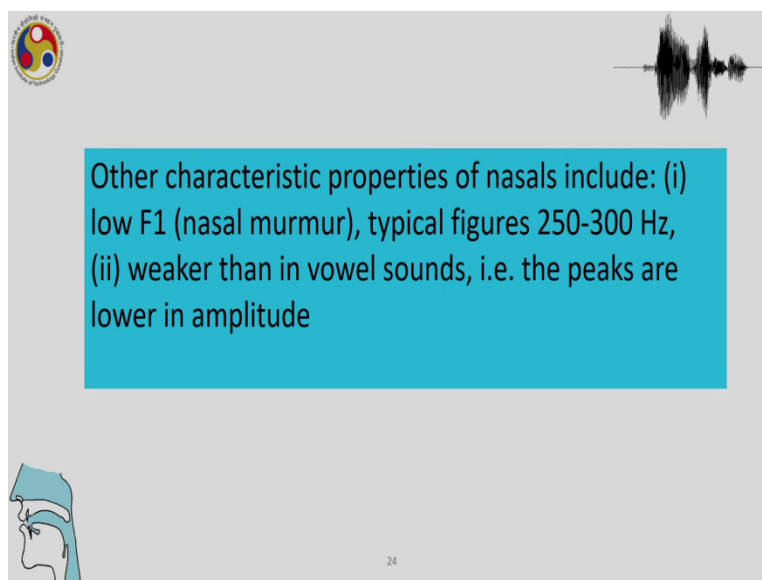
Nasals and laterals are characterized by the presence of a side branch, i.e. a closed tube.

During the production of nasal consonants, the velum is lowered

the pathway from the pharynx to the nasal passages is open and air flows from the lungs out through the nostrils. In nasal stops, the mouth cavity is closed off by a complete constriction in the vocal tract.

And so that is it is a closed tube. During the production of nasal consonants a velum is lowered, and the pathway from pharynx to the nasal passages open and air flows from the lungs through the nostrils. In nasal stops, the mouth cavity is closed off by complete constriction in the vocal tract.

(Refer Slide Time: 10:46)

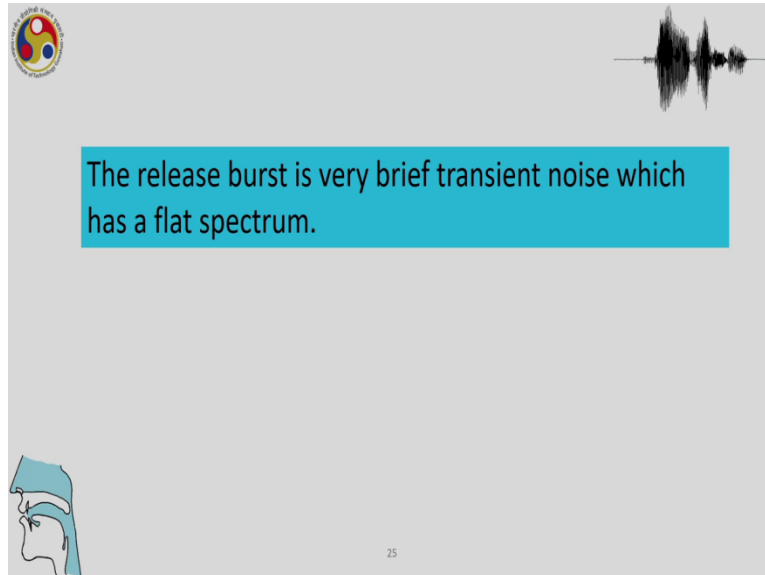


The slide features a logo in the top left corner, a waveform in the top right, and a sagittal cross-section of the human head in the bottom left. The main text is contained within a light blue rectangular box.

Other characteristic properties of nasals include: (i) low F1 (nasal murmur), typical figures 250-300 Hz, (ii) weaker than in vowel sounds, i.e. the peaks are lower in amplitude

And also recall that we talked about low f1 for nasals and around 250 to 200 hertz and weaker than in vowels, the peaks are lower in amplitude.

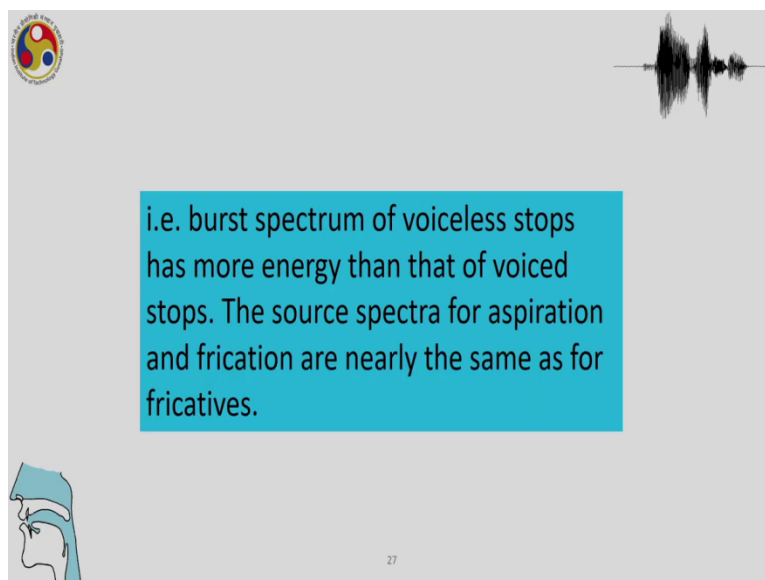
(Refer Slide Time: 11:02)



The release burst is very brief transient noise which has a flat spectrum.

The release burst is a very brief transient noise, which has a flat spectrum.

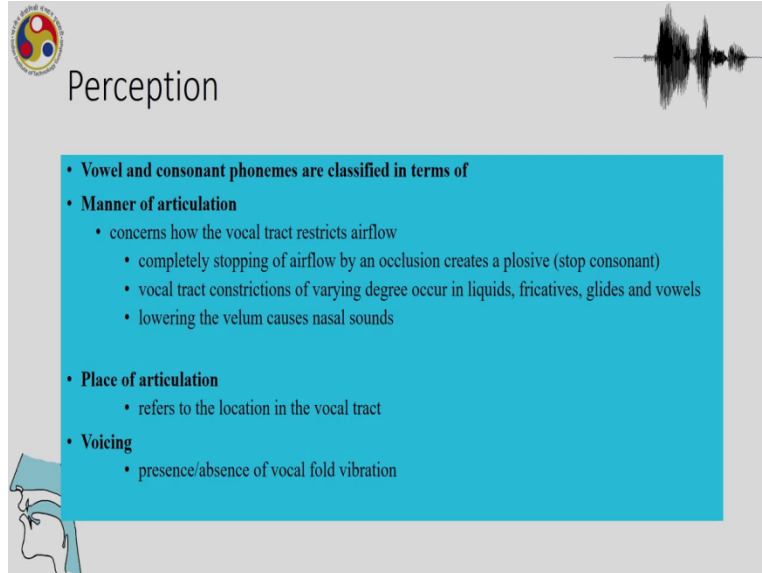
(Refer Slide Time: 11:08)



i.e. burst spectrum of voiceless stops has more energy than that of voiced stops. The source spectra for aspiration and frication are nearly the same as for fricatives.

And also voiceless stops has more energy than that of voice stops and burst spectrum. The source spectra for aspiration and frication are nearly the same, that is, the noise.

(Refer Slide Time: 11:18)



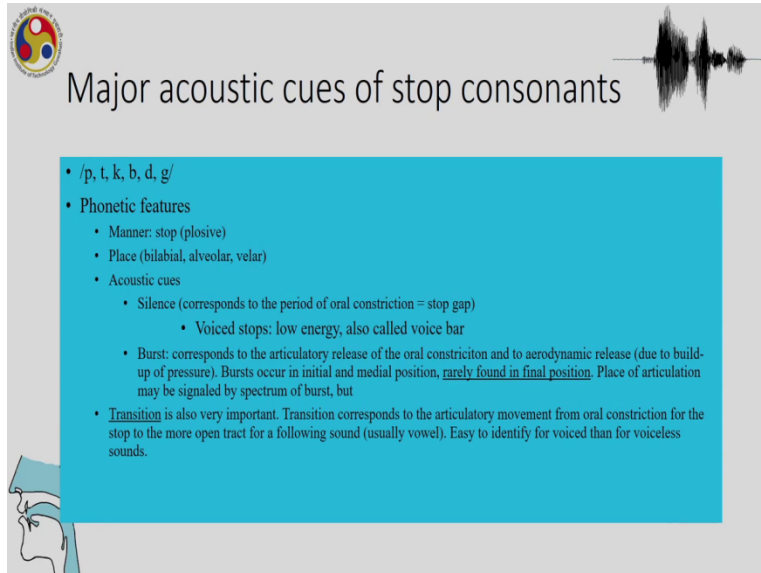
Perception

- **Vowel and consonant phonemes are classified in terms of**
- **Manner of articulation**
 - concerns how the vocal tract restricts airflow
 - completely stopping of airflow by an occlusion creates a plosive (stop consonant)
 - vocal tract constrictions of varying degree occur in liquids, fricatives, glides and vowels
 - lowering the velum causes nasal sounds
- **Place of articulation**
 - refers to the location in the vocal tract
- **Voicing**
 - presence/absence of vocal fold vibration

Now, when it comes to perception, so, all these things that we had studied as a result of the source filter theory of consonant production, now, we have to be aware that all of this contributes to the perception of speech. So, when we talk about consonants, when you talk about manner of articulations, so, the way stop completely stops the airflow, and other sounds like liquids, fricatives glides and vowels have varying degree of occlusions.

And lowering the velum causes nasal sounds, place of articulation, which refers to the location of the occlusion in the vocal tract and voicing his presence, absence of the vocal fold vibration, and this is the most basic information about articulatory phonetics that you already know from the previous lectures.

(Refer Slide Time: 12:22)



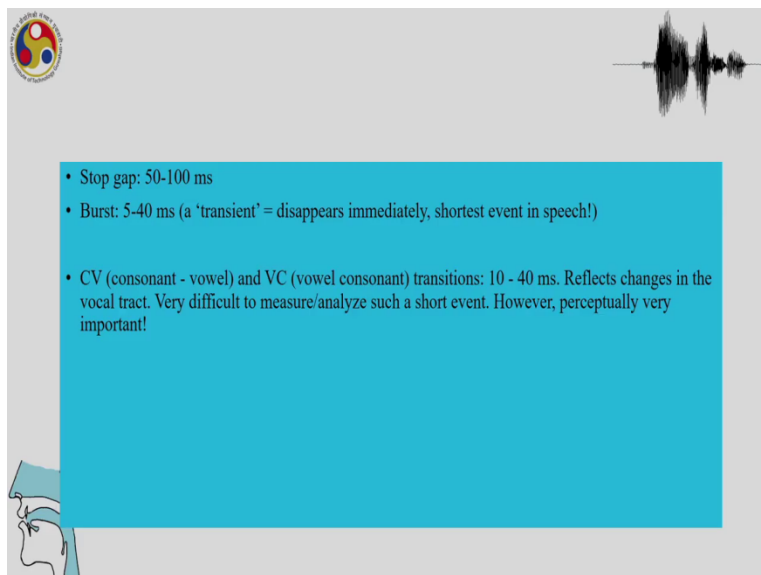
The slide features a logo in the top left corner, a waveform in the top right, and a profile of a human head in the bottom left. The main content is a blue box with the following text:

Major acoustic cues of stop consonants

- /p, t, k, b, d, g/
- Phonetic features
 - Manner: stop (plosive)
 - Place (bilabial, alveolar, velar)
 - Acoustic cues
 - Silence (corresponds to the period of oral constriction = stop gap)
 - Voiced stops: low energy, also called voice bar
 - Burst: corresponds to the articulatory release of the oral constriction and to aerodynamic release (due to build-up of pressure). Bursts occur in initial and medial position, rarely found in final position. Place of articulation may be signaled by spectrum of burst, but
 - Transition is also very important. Transition corresponds to the articulatory movement from oral constriction for the stop to the more open tract for a following sound (usually vowel). Easy to identify for voiced than for voiceless sounds.

And coupled with the major acoustic cues, stop, which has a low energy and the voice bar that you saw for voice stops, and the burst that you saw for stops because of the release and buildup of pressure and the release and place of articulation signaled by the spectrum of burst.

(Refer Slide Time: 12:50)



The slide features a logo in the top left corner, a waveform in the top right, and a profile of a human head in the bottom left. The main content is a blue box with the following text:

- Stop gap: 50-100 ms
- Burst: 5-40 ms (a 'transient' = disappears immediately, shortest event in speech!)
- CV (consonant - vowel) and VC (vowel consonant) transitions: 10 - 40 ms. Reflects changes in the vocal tract. Very difficult to measure/analyze such a short event. However, perceptually very important!

And also the transition, which is very important form and transitions, which we talked about. So remember the period of silence that we talked about, there is a stop gap of 5200 milliseconds. And there is a silence which we also talked about when we talked about in the last class, we

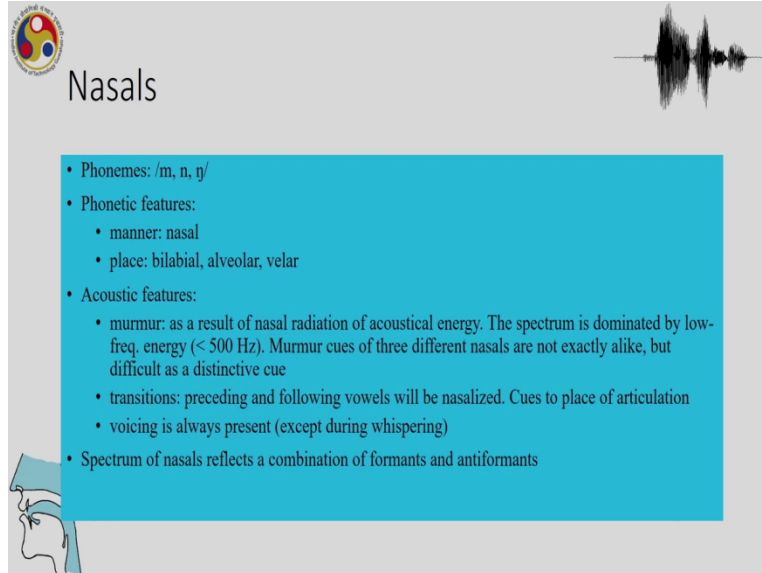
talked about how that is compensated for that silence is not almost is not heard, and those things are actually cues for the perception of different consonants.

(Refer Slide Time: 13:25)

- Phonemes:
- Phonetic features:
- [ʃ] [ʒ] [s] [z] [o] [ɒ] [v] [ʋ]
 - manner: frication
 - place: labiodental, linguadental, alveolar, palatal, glottal
- Acoustic cues:
 - voicing
 - frication noise: noise generated as air is forced through a narrow constriction. Then filtered by the vocal tract.
 - transitions to and from the vowels due to changes in the vocal tract
- sibilants/ stridents have intense noise energy
- non sibilants: weak noise energy

So, phonemes phonetic features like frication and acoustic cues like voicing and sibilants etcetera, what we showed in acoustics, and those things contribute a lot in the perception of the sounds. So, noise generated as air is forced through a narrow constriction for fricatives and that is filtered by the vocal tract and transitions to and from the vowels due to changes in the vocal tract and sibilants have intense noise energy and non sibilants have weak non energy. So, weak noise. So, the noise aspect which is stressed in the acoustic part is also very important in perception.

(Refer Slide Time: 14:08)

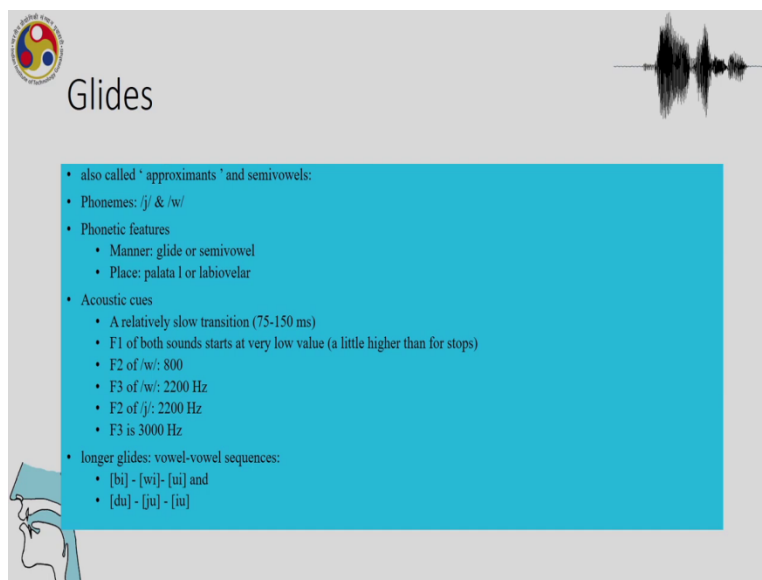


Nasals

- Phonemes: /m, n, ŋ/
- Phonetic features:
 - manner: nasal
 - place: bilabial, alveolar, velar
- Acoustic features:
 - murmur: as a result of nasal radiation of acoustical energy. The spectrum is dominated by low-freq. energy (< 500 Hz). Murmur cues of three different nasals are not exactly alike, but difficult as a distinctive cue
 - transitions: preceding and following vowels will be nasalized. Cues to place of articulation
 - voicing is always present (except during whispering)
- Spectrum of nasals reflects a combination of formants and antiformants

And similarly for nasals. The spectrum is dominated by low frequency energy, etc.

(Refer Slide Time: 14:20)

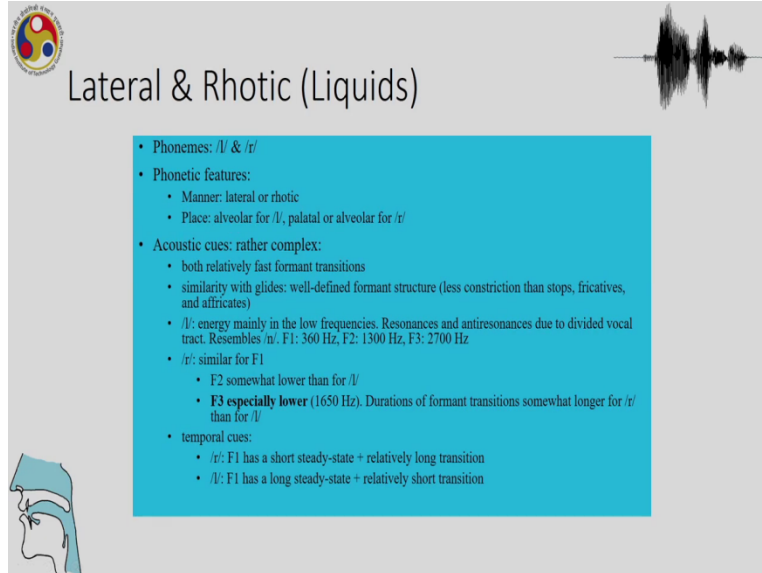


Glides

- also called 'approximants' and semivowels:
- Phonemes: /j/ & /w/
- Phonetic features
 - Manner: glide or semivowel
 - Place: palata l or labiovelar
- Acoustic cues
 - A relatively slow transition (75-150 ms)
 - F1 of both sounds starts at very low value (a little higher than for stops)
 - F2 of /w/: 800
 - F3 of /w/: 2200 Hz
 - F2 of /j/: 2200 Hz
 - F3 is 3000 Hz
- longer glides: vowel-vowel sequences:
 - [bi] - [wi] - [ui] and
 - [du] - [ju] - [ɰ]

And also nasal pole and zero, which we had talked about yesterday are also cues for nasals. And similarly for glides are relatively slow transition from 75 to 100 milliseconds and f1 of both the sound sounds starts very low. The important thing about glides is that there is this movement which is quite apparent in the acoustics and that is also a cue for its perception.

(Refer Slide Time: 14:53)

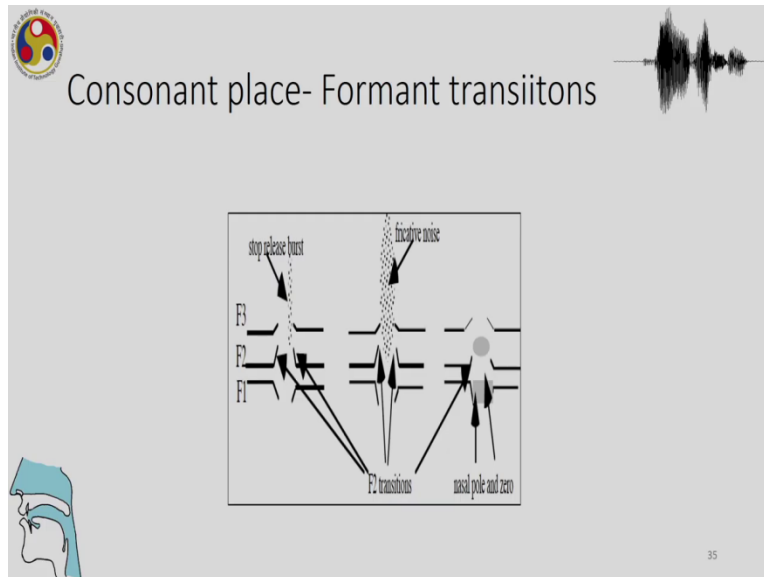


Lateral & Rhotic (Liquids)

- Phonemes: /l/ & /r/
- Phonetic features:
 - Manner: lateral or rhotic
 - Place: alveolar for /l/, palatal or alveolar for /r/
- Acoustic cues: rather complex:
 - both relatively fast formant transitions
 - similarity with glides: well-defined formant structure (less constriction than stops, fricatives, and affricates)
 - /l/: energy mainly in the low frequencies. Resonances and antiresonances due to divided vocal tract. Resembles /n/. F1: 360 Hz, F2: 1300 Hz, F3: 2700 Hz
 - /r/: similar for F1
 - F2 somewhat lower than for /l/
 - **F3 especially lower** (1650 Hz). Durations of formant transitions somewhat longer for /r/ than for /l/
- temporal cues:
 - /r/: F1 has a short steady-state + relatively long transition
 - /l/: F1 has a long steady-state + relatively short transition

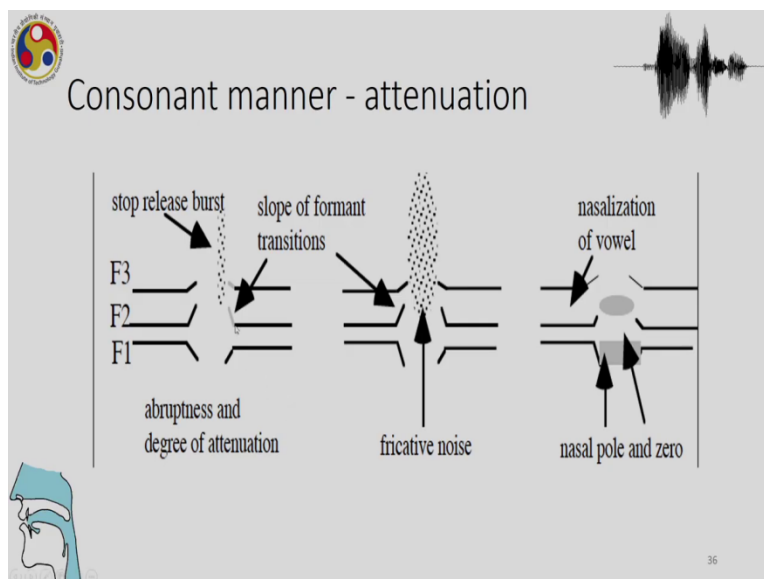
When it comes to the proximate la and ra their, acoustic cues are quite complex. So, they are both very fast moving form and transitions and they have a well defined form and structure and However, one should remember the F three for l is especially lower F1 is also low. And then there are some other cues like F1 is short steady state and long transition, F1 is long steady state for low and short transition. So, even though the sounds are pretty similar, these are the timing of the F1 and transition is different for la and ra.

(Refer Slide Time: 15:40)



So, summarizing like yesterday, so, these are the things which are important. So, the stop release burst, the fricative noise, nasal pole 0 zero and F2 transitions for all of these is important for the place of articulation.

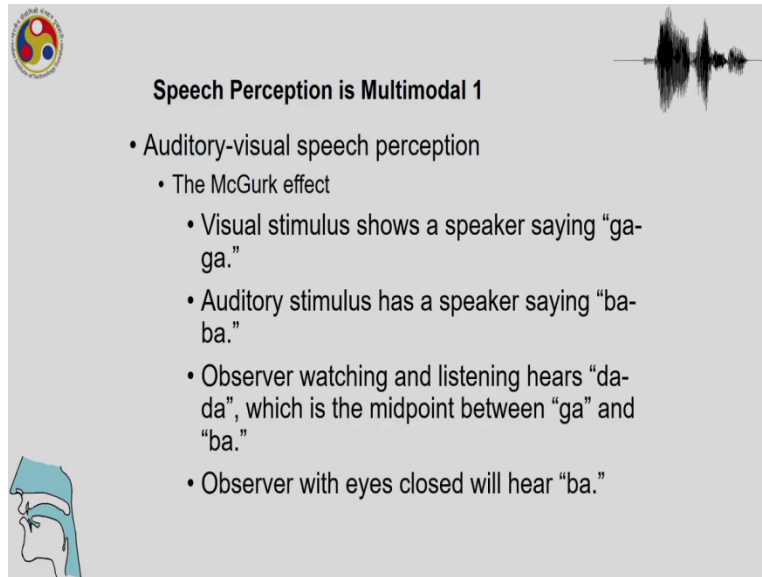
(Refer Slide Time: 16:00)



For manner of articulation, the attenuation the chain, the energy, which is a burst in energy for stops and the noise in fricatives and the low amplitude in nasals and that is important and the abruptness in in the stops, then the fricative noise, the low hissing noise and in nasals the

amplitude difference which makes a difference between the stop and a nasal is an important cue. And also for glides difference between F2 and F3 which is not shown here is also important.

(Refer Slide Time: 16:40)



Speech Perception is Multimodal 1

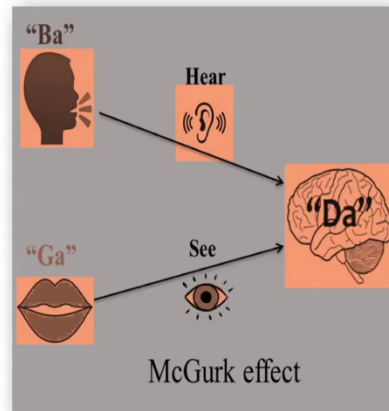
- Auditory-visual speech perception
 - The McGurk effect
 - Visual stimulus shows a speaker saying “ga-ga.”
 - Auditory stimulus has a speaker saying “ba-ba.”
 - Observer watching and listening hears “da-da”, which is the midpoint between “ga” and “ba.”
 - Observer with eyes closed will hear “ba.”

So, let us also talk about end this lecture on speech perception by discussing how speech is multimodal, which means it not only uses the articulatory medium, it also uses of other mediums to for efficient communication, this is with regard to what is called the McGurk effect. So, this is something which is first shown by McGurk.

This is a very interesting experiment which shows the McGurk effect, there is a visual stimulus which shows the speaker saying gaga auditory stimulus, she says, a speaker says, ba ba, and then the listener, listeners hear da da, even though neither of the visual stimulus nor the auditory stimulus says da. So, basically, the listener hears something between gah and ba, because there is a clash between the auditory stimulus and the visual stimulus, there is a sort of a trade off and you hear something in between.

However, if the eyes are closed, then the speaker hears what is said. And then when the eyes are closed, there is no clash with the visual medium.

(Refer Slide Time: 18:04)



The McGurk effect. The lips are moving as if he/she is saying /ga-ga/, but the actual sound being presented is /ba-ba/. The listener, however, reports hearing the sound /da-da/. If the listener closes his eyes, so that he no longer sees the lips, he hears /ba-ba/. Thus, seeing the lips moving influences what the listener hears.

So, as shown in this diagram, the speaker says ba ba and the lip says ga ga and what is played is ba ba, but the perception is da da.

(Refer Slide Time: 18:18)



- Prepared with help from
- Edward Flemming, *24.915 Linguistic Phonetics*. Fall 2015. Massachusetts Institute of Technology: MIT OpenCourseWare, <https://ocw.mit.edu>. License: [Creative Commons BY-NC-SA](#).

39

So, these are, we have used some of our information from the MIT courseware on linguistic phonetics.

(Refer Slide Time: 18:30)



- Steriade, Donca (1997). Phonetics in phonology: the case of laryngeal neutralization. Ms, UCLA.
- Stevens, Kenneth N., and Sheila E. Blumstein (1981). The search for invariant acoustic correlates of phonetic features. Peter D. Eimas and Joanne L. Miller (eds.) Perspectives on the study of speech. Lawrence Erlbaum, Hillsdale.
- Wright, R., Frisch, S., & Pisoni, D. B. (1999). Speech Perception. In J. G. Webster (Ed.), Wiley Encyclopedia of Electrical and Electronics Engineering, Vol. 20 (pp. 175-195). New York: John Wiley and Sons.

And these are some of the papers that we have used to prepare our slides. Thank you very much for your attention. And we will continue with the course phonetics and phonology a broad overview in the following lectures. Thank you.