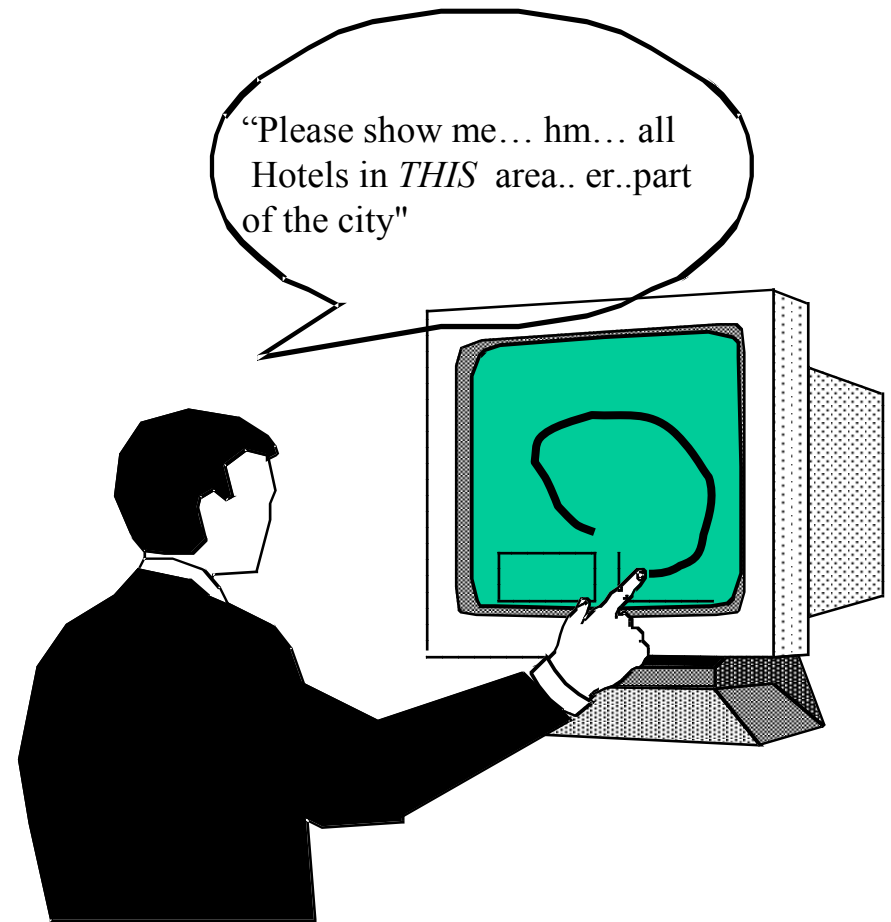


Speech Recognition

25. Aug. 2008

Better Human-Machine Interaction

- Speaking
- Pointing,
- Gesturing
- Hand-Writing
- Drawing
- Presence/Focus of Attention
- Combination
 - Sp+HndWrtg+Gestr.
 - Repair
- Multimodal NLP & Dialog
- Learning from Experience



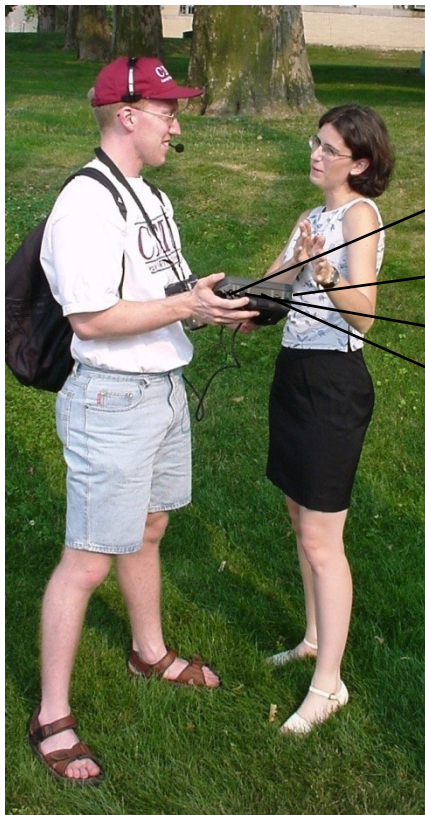
Interaction with Humanoid Robots



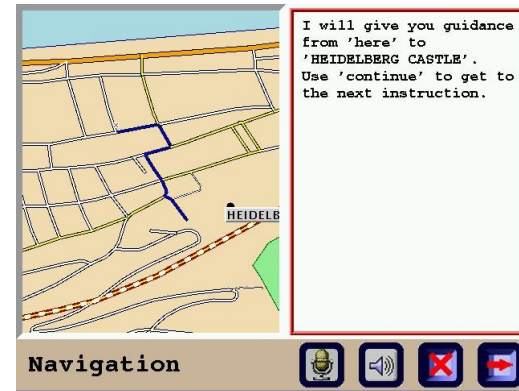
- Some Observations:
 - Multimodal Input
 - Robustness
 - Dialog
 - Learning from Interaction
 - Understanding the Context
 - Direct Interaction vs. Implied

LingWear:

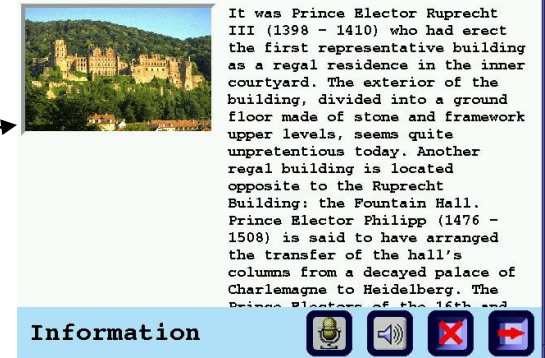
Wearable Language Assistance for the Information Warrior



Navigation



Information Access



Document Translation



水栖、陆栖、两栖动物化石及鱼、蟹、龟、
窝、蚌贝等；有各种草席、木席植物化石，
达数百种。
这是一个中外早已知名的科研现场。二
代初，受到国内外地质考古学家的重视，
行系统的考查研究，为世界地质科研提供
典论著。直到七十年代末我国人类学家在
区作了进一步的考查研究。

Meetings



Alex Waibel – Speech Recognition

Human-Human Communication in a Multilingual Distributed Context



你们的评估准则是什么



Meetings, Lectures

- Participants are Remote *and* Local
- Participants Speak Different Languages
 - Cross-Lingual Dialog *and* Translation of Monolingual Dialog
- Invisible Computer Provides Transparent Services
 - Translation
 - Smmarization

Speech



Transcript: Onune baksana be adam!

Turkish

Language ID

Bus Station

Acoustic Scene

Angry

Emotion ID

Negotiation

Discourse Analysis

Umut

Speaker ID

Meeting

Topic ID

Istanbul

Entity Tracking

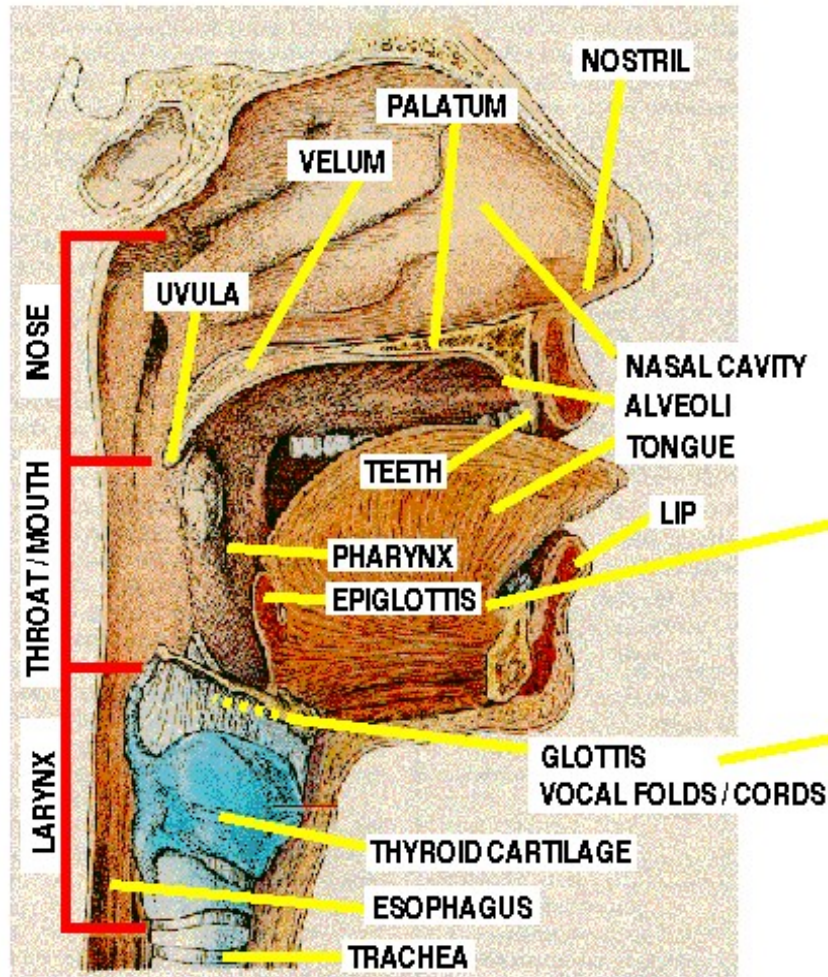
Speech

- Acoustic Phonetics, Speech Production and Perception
- Linguistics and Psychology
- Speech Recognition
 - Isolated and Continuous Word
 - Large Vocabulary Continuous Speech (Read Speech)
 - Conversational Speech
- Speaker Recognition
- Speaker Verification
- Language Identification
- Emotion Recognition
- Speech Synthesis
- Topic Identification
- Spoken Language Understanding
- Dialog Processing
- Machine Translation

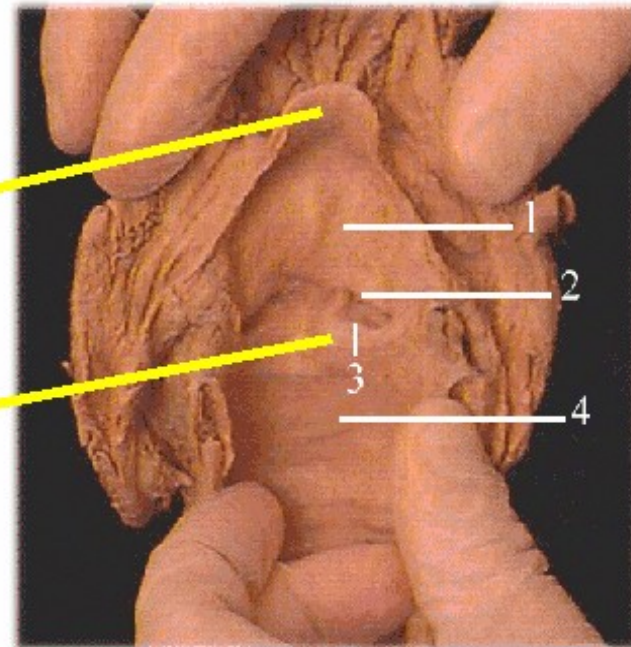
A Few Related Sciences

- Statistics
- Biology
- Linguistics
- Psychology
- Physiology / Anatomy
- Mathematics, Physics
- Electrical Engineering, Communication Theory
- Information Theory, Coding Theory
- Signal Processing
- Pattern Recognition
- Artificial Intelligence
- Language Processing

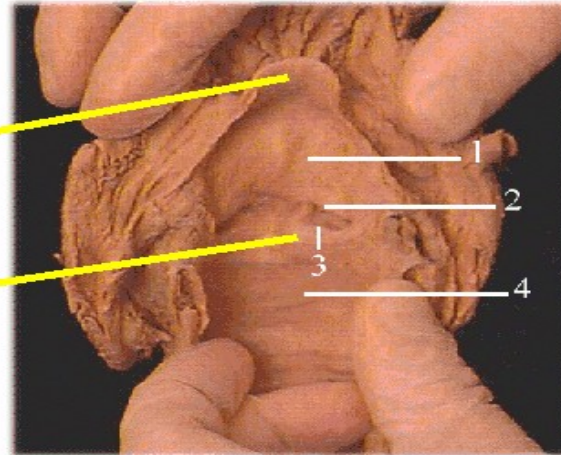
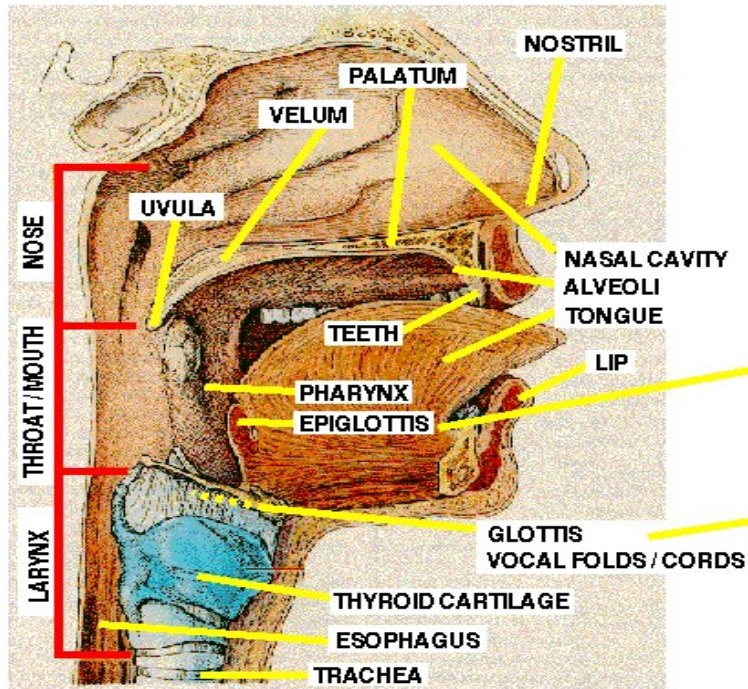
Anatomy of Speech Production



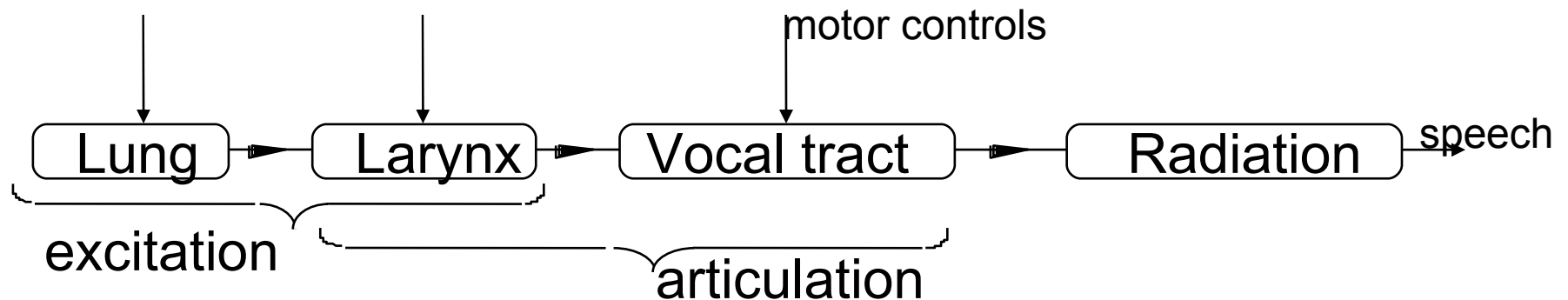
University
Erlangen
Department of
Phoniatrics and
Pedaudiology
Waldstr.1
D-91054 Erlangen



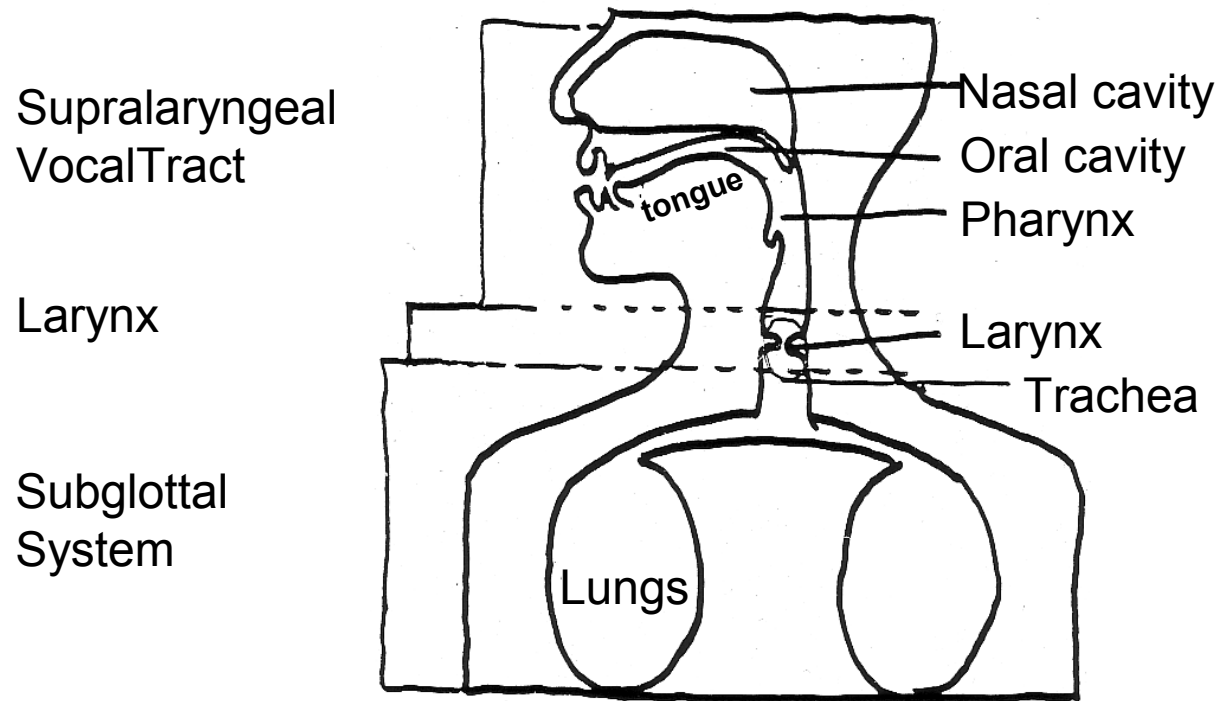
Speech Production



The three physiologic components of human speech production

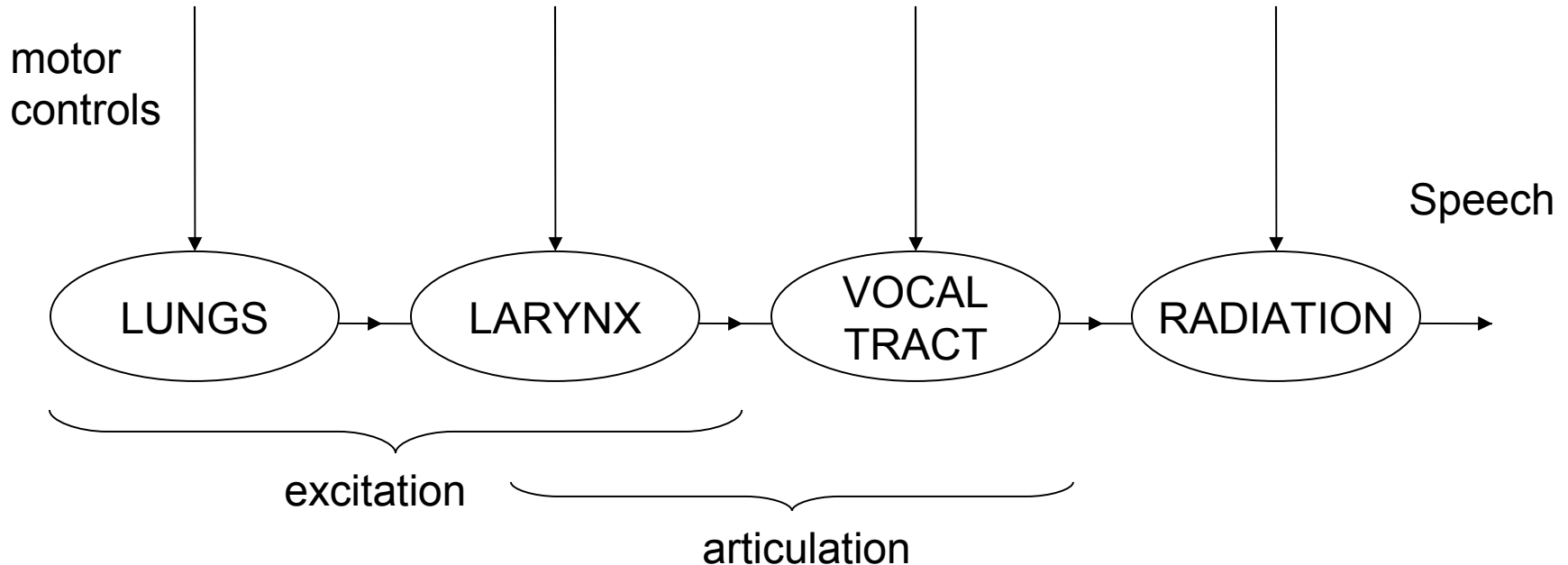


Speech Production



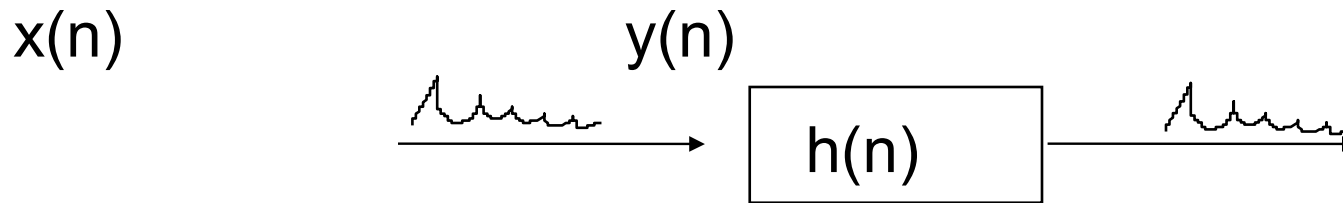
- The three physiologic components of human speech production -

Speech Production (cont.)



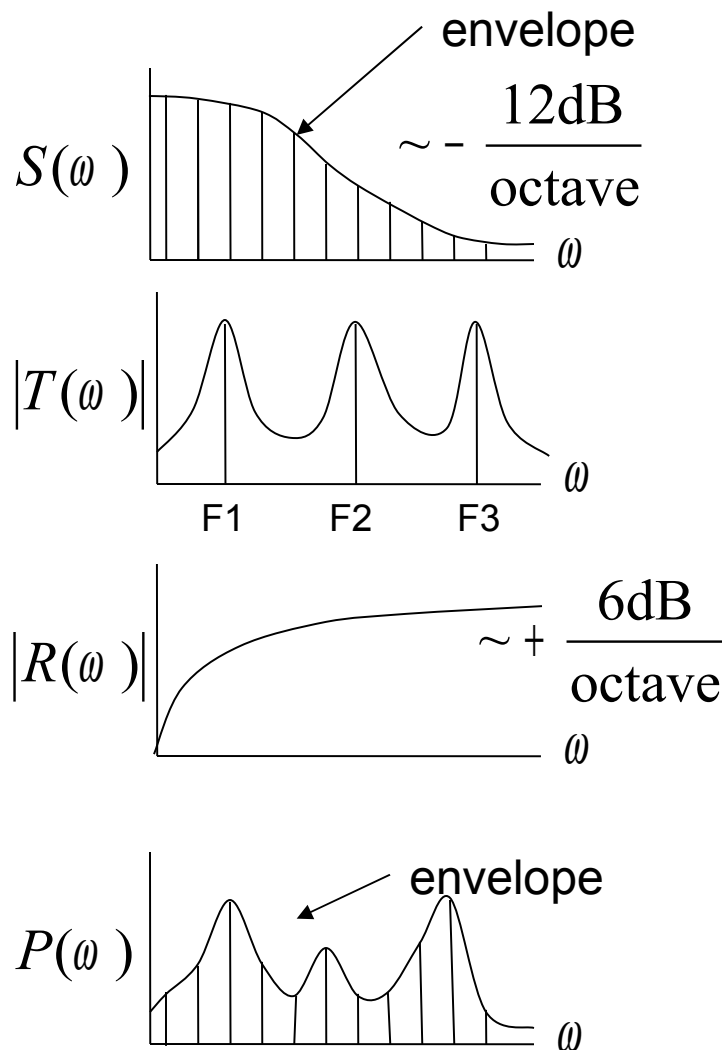
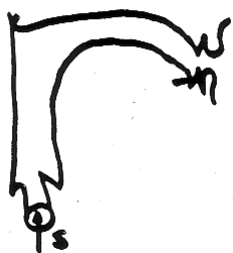
- Functional Block Diagram of Speech Production -

Convolution

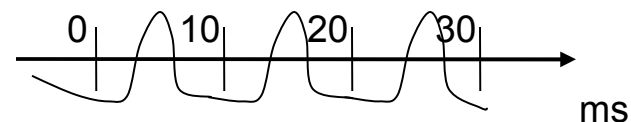


$$y(n) = x(n) * h(n) = \sum_{k=-\infty}^{\infty} x(k) h(n - k)$$

Transfer Functions of the Different Components of Speech Production



Periodic excitation (Vowel)

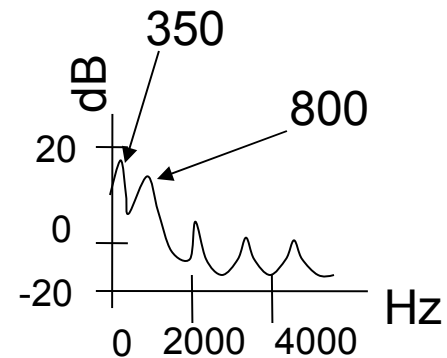
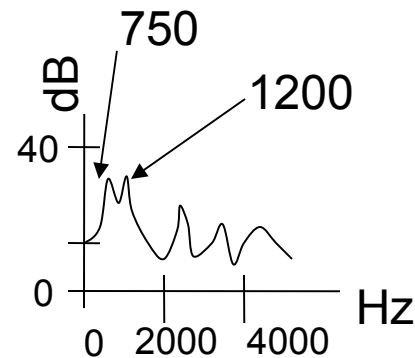
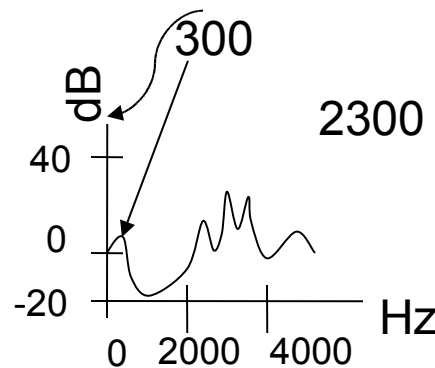


Vocal Tract Transfer Functions for Different Vowels

Vocal Tract Shapes

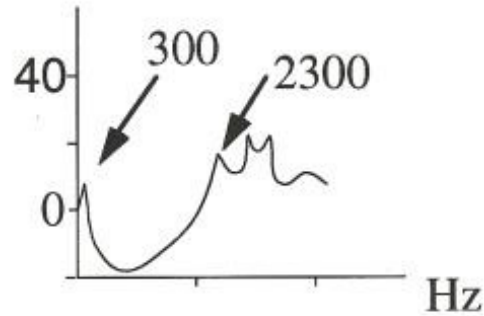


Resulting Transfer Functions (Spectra)

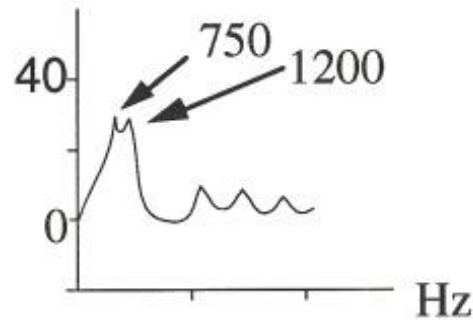


Vocal tract & Transferfunctions

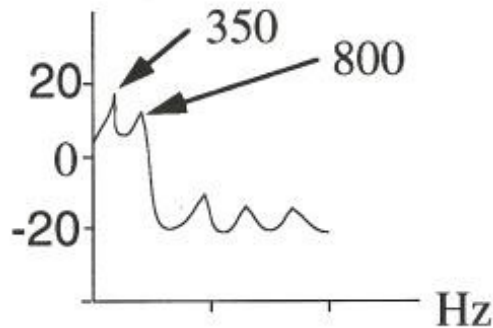
/i/



/a/



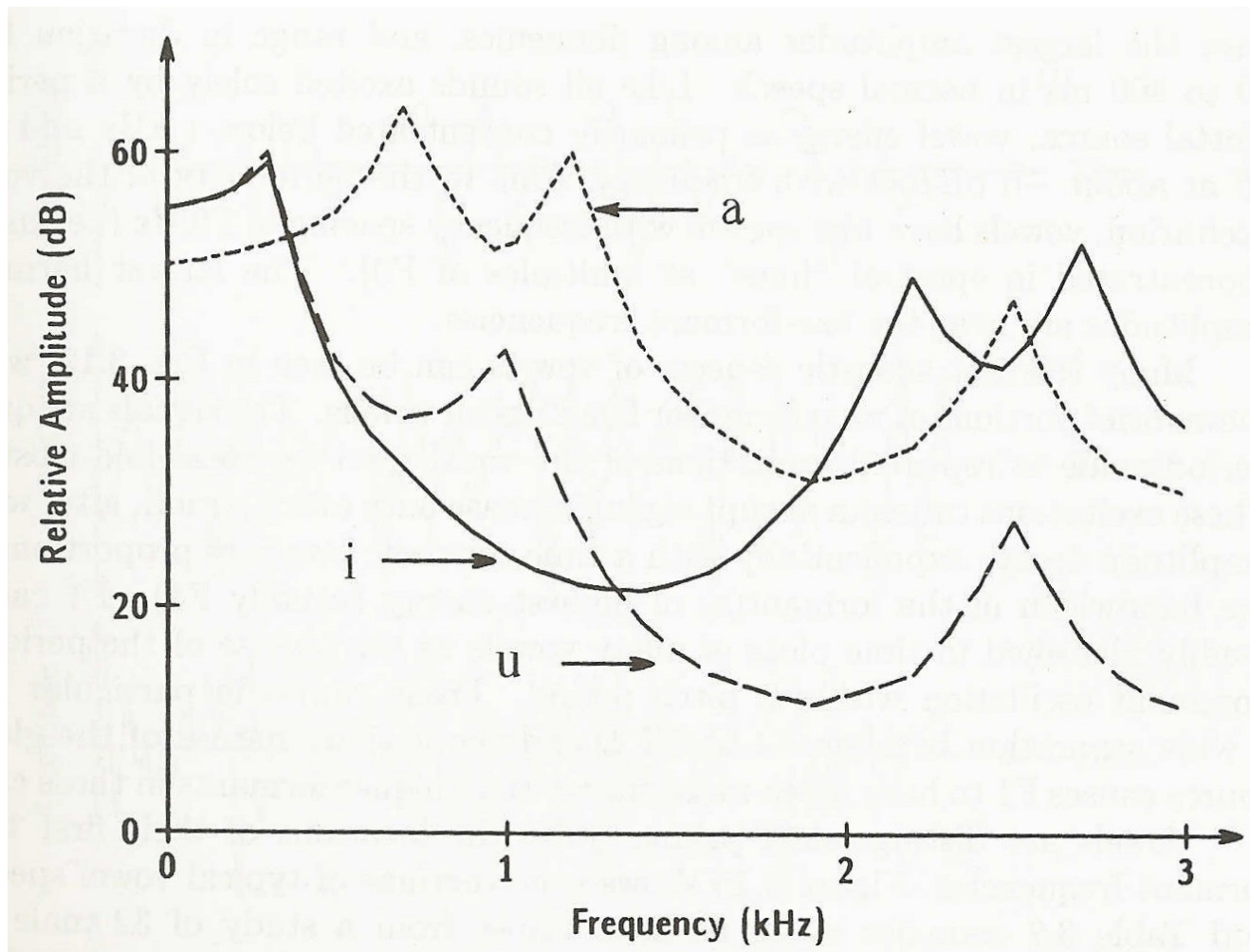
/u/



Vokaltraktformen

Resultierende
Transferfunktionen

Vowels /a/, /i/ and /u/



Different Vocal Tract Shapes

BEAT



BIT



BAIT



BET



BAT



BART



BALL



BOY



BUTCH



BOOT



BUT

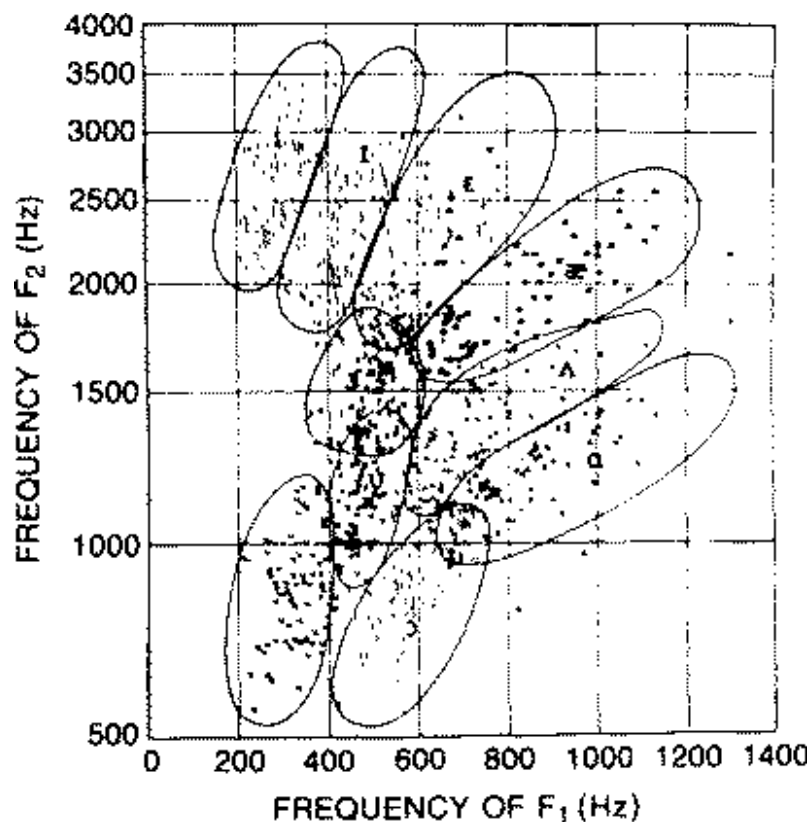


BIRD

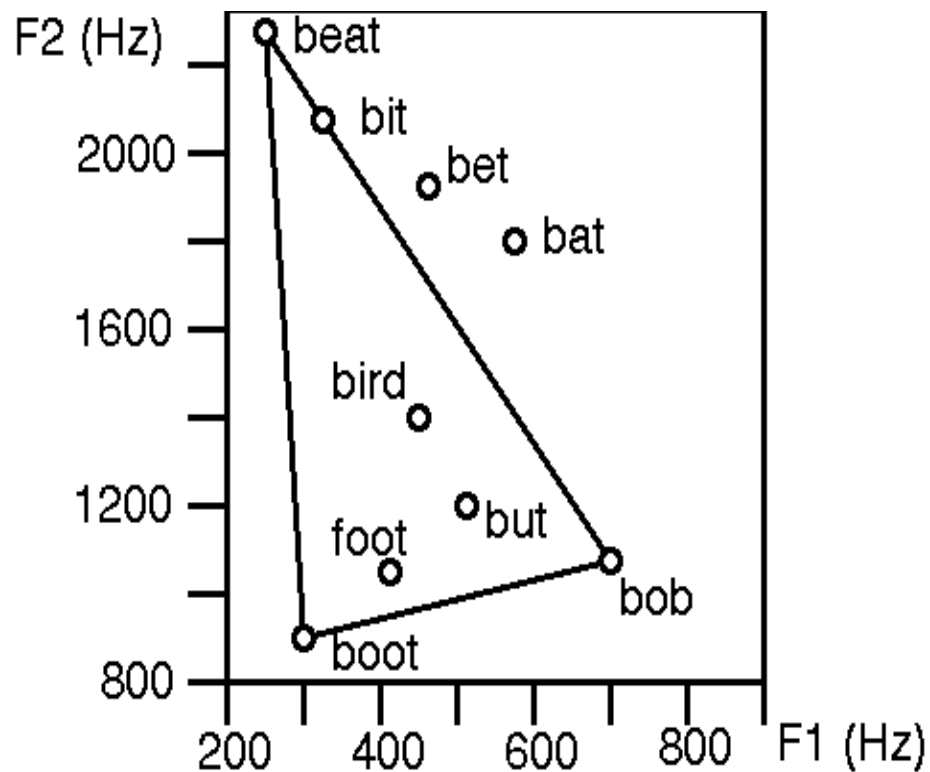


Formants

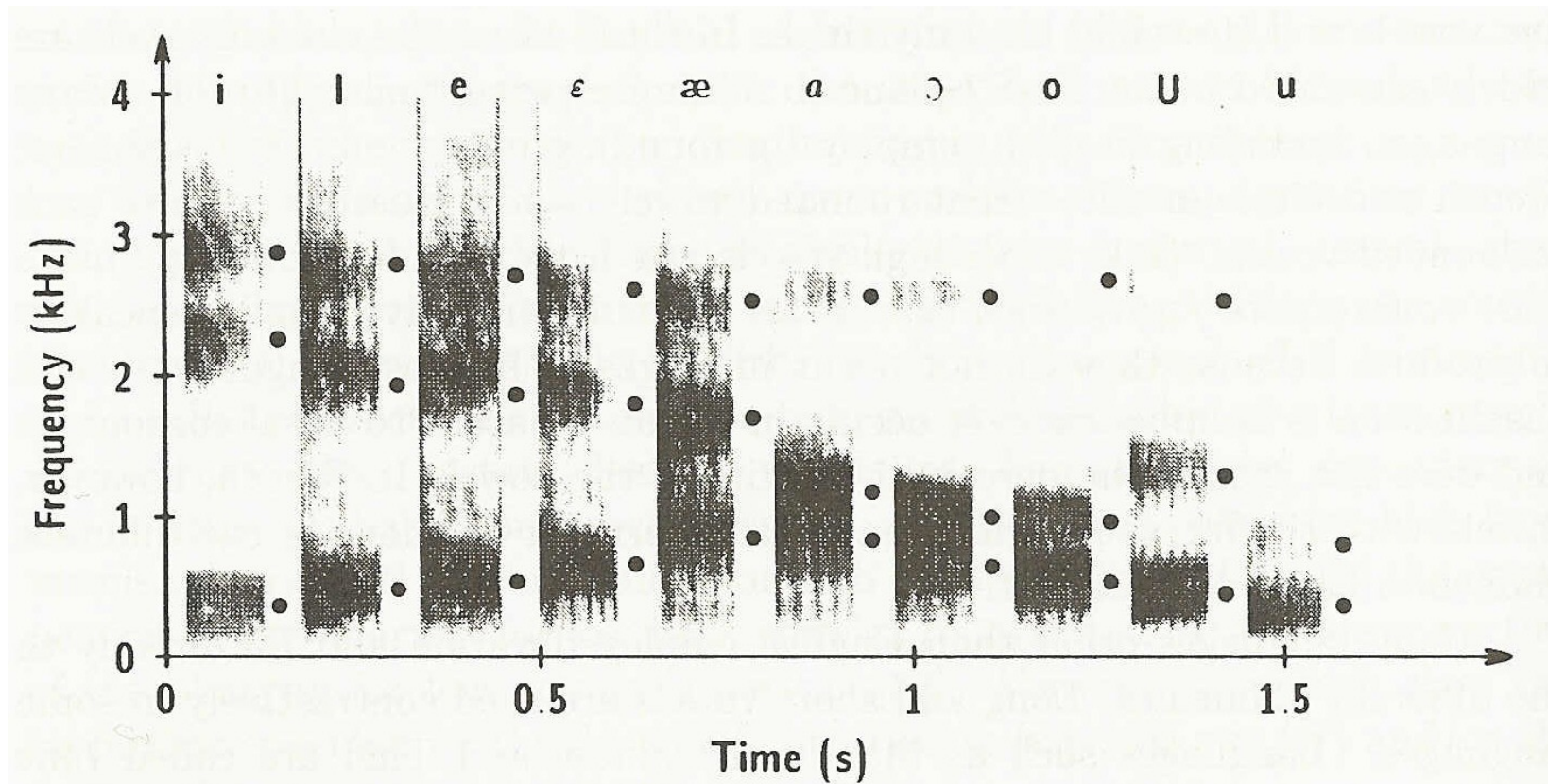
The resonance frequencies of the vocal tract transfer function are called formants. In practice, only the first few formants are of interest.



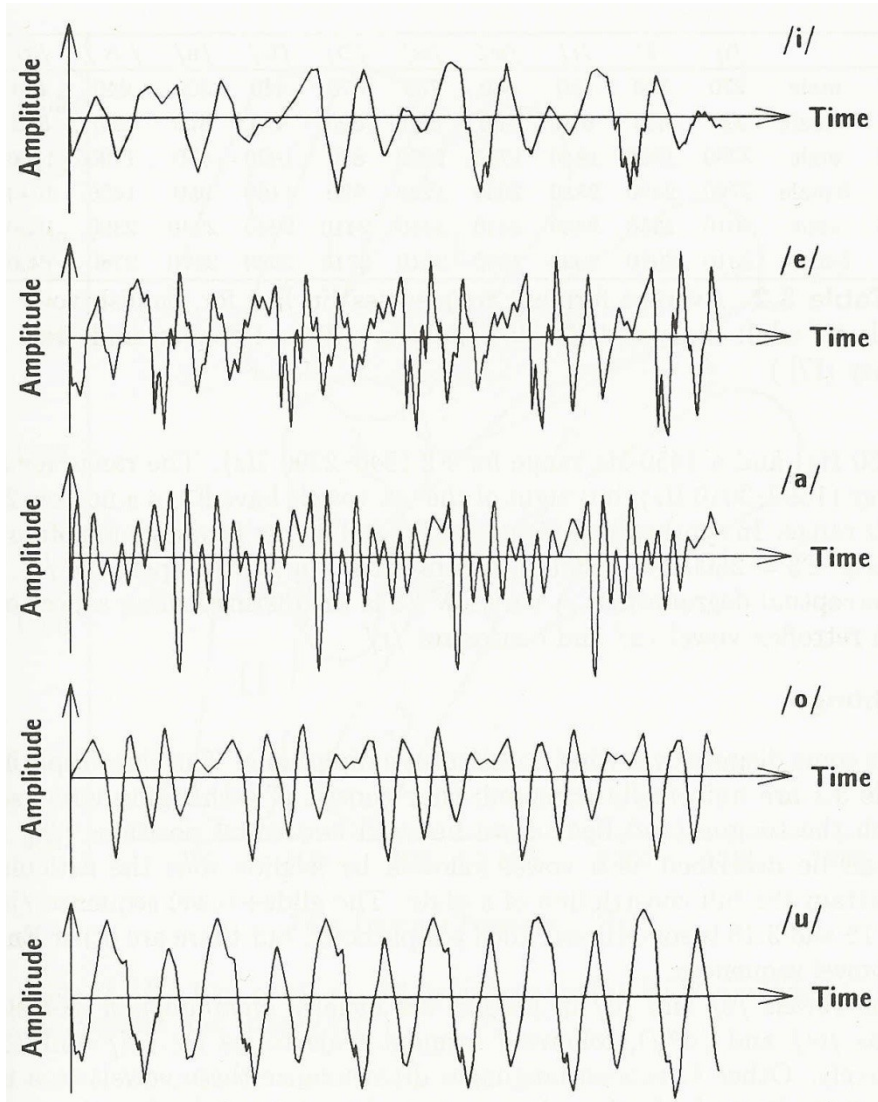
The Vowel-Triangle



Spectrograms



Vowels in the Time Domain

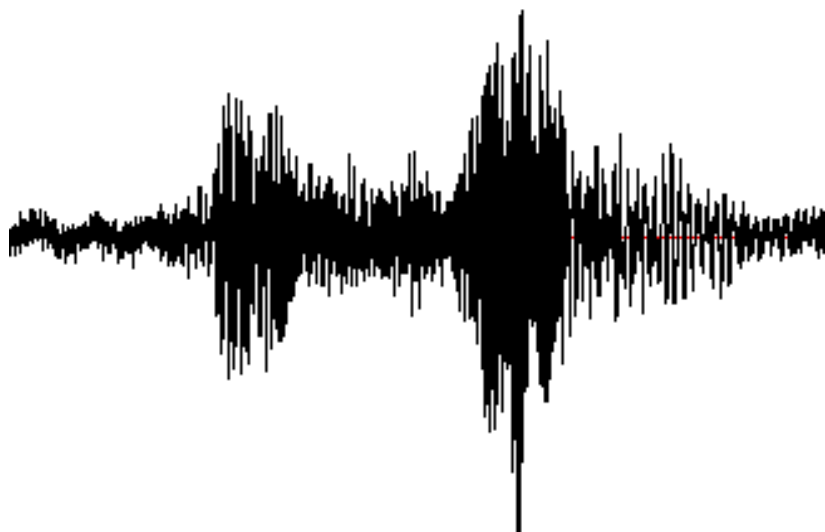


Consonants

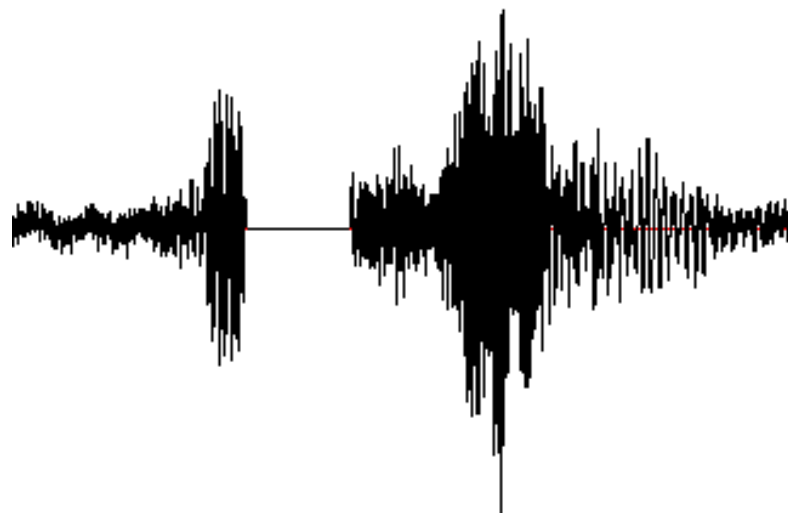
Consonants are sounds which are articulated by temporarily constricting the airflow or stopping the airflow completely.

Listen to these examples:

original recording



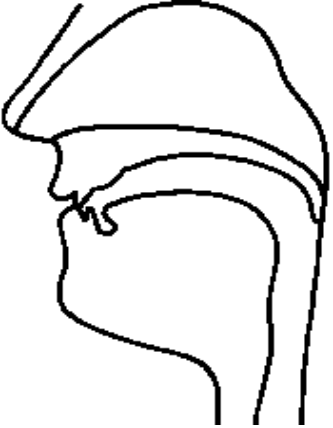


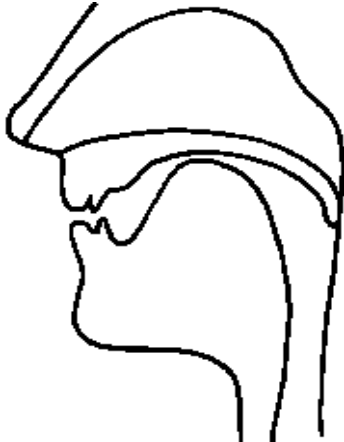
one part blanked out



The blanked-out part sounds like a (plosive) consonant.

Vocal Tract Shapes of Consonants

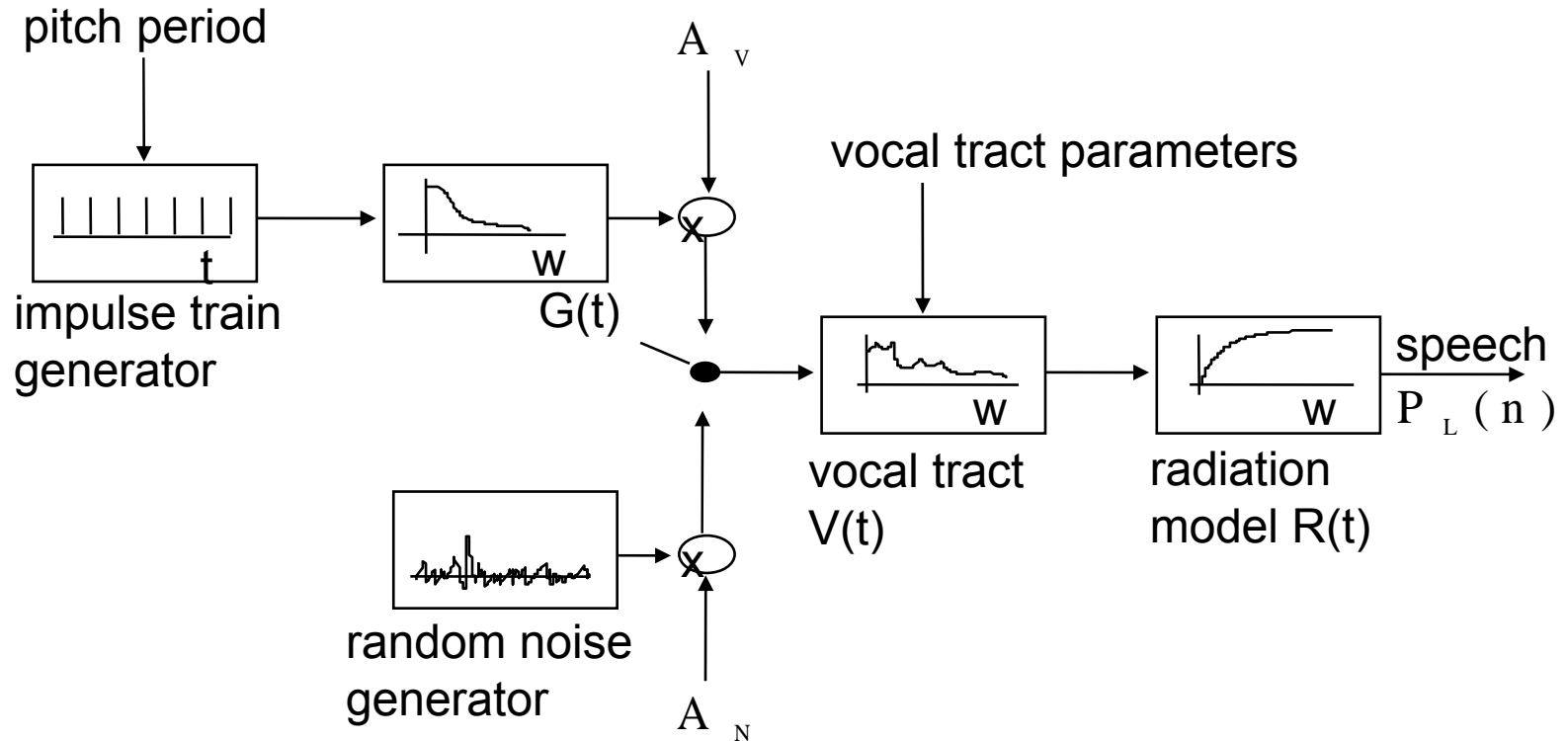
Fricatives

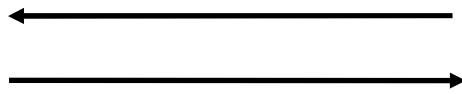
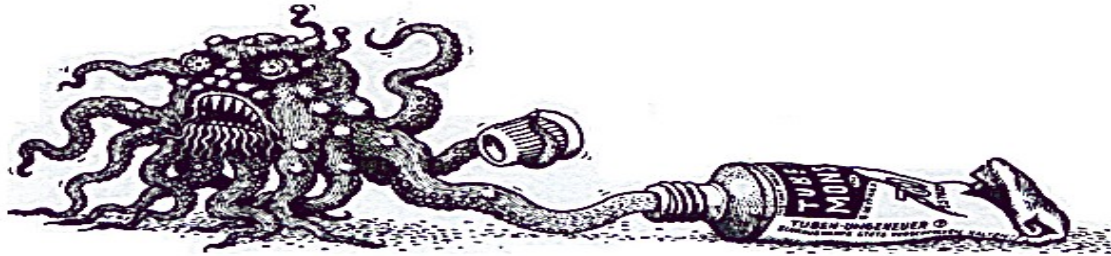
Lip-Teeth Friction	Tongue-Teeth Tongue-Alveoli	Palatal Friction Alveolar Friction	Palatal Friction
			
FAN, VAN	SUE, ZOO	VISION, VICIOUS	YOU

Additionally there is a glottal fricative /**h**/ as in **H**OUSE.

Other languages often also have aspirated velar and palatal fricatives.

Vocal Tract Model of Speech





speech synthesis
speech recognition

Dimensions of Difficulty

- Noise – Environmental, Channel, Reverberation
- Speaker – Male, Female, Children, Elderly
- Acoustic Similarity – Letters, Digits,...
- Vocabulary Size – 10 → 100,000 words
- Speaking Style – Isolated, Continuous Read Speech, Spontaneous, Conversational Speech