EEM.ssr: speaker & speech recognition (EEEM034)

Speech communication

by
Dr Philip Jackson

Centre for Vision, Speech & Signal Processing (CVSSP)
Department of Electronic Engineering

Origins of speech communication

- Speech has evolved as a mechanism for communicating information (a message) from one person to another

- In humans, speech takes advantage of:
  - Versatile and agile vocal and articulatory physiology
  - Sophisticated languages that have developed
  - Complexity of the human brain for recognition

- The message undergoes several transformations as it is transmitted from one person to another

Enhanced fitness for survival

The development of speech communication in *homo sapiens* has been driven by evolutionary forces:

- relaying the location of natural resources
- escaping predators
- coordinating hunting
- passing on learnt skills
- monitoring the well being on community
- …and wooing!

Complexity of communication demands intelligence

- Communication is vital in social organisms:
  - Warning and alarm calls of animals
  - Imitation of sounds by birds including speech
  - Small vocabulary of sounds: great apes, whales

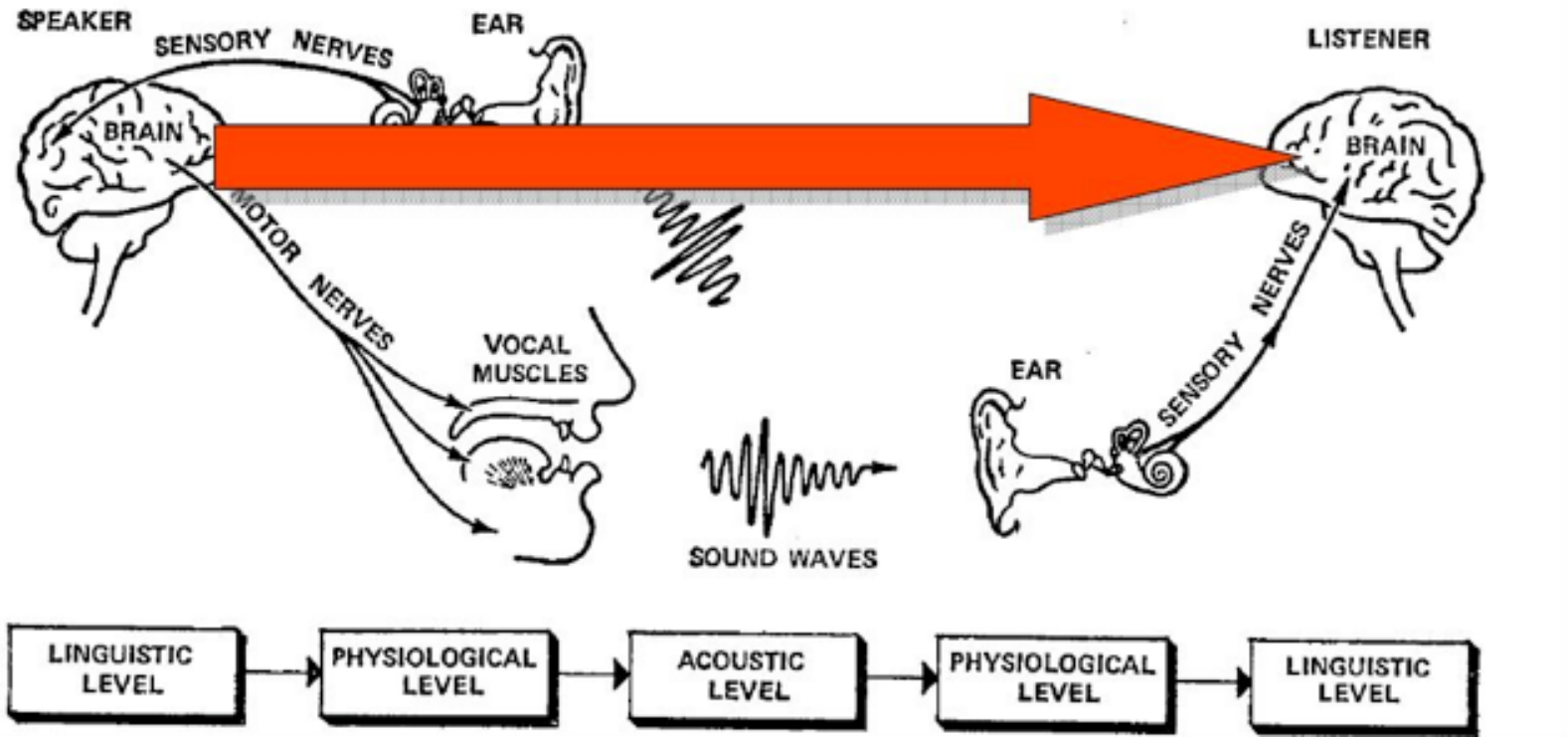
- Yet spoken language is very complex
  - requires fusion of many sources of knowledge

- Humans have developed large brains and supreme intelligence in the animal kingdom to deal with it:
  - very large number of neurons, in parallel

Speech signal chain

- **Speech chain** is the sequence of signal transformations as message travels from speaker's mind to listener's:

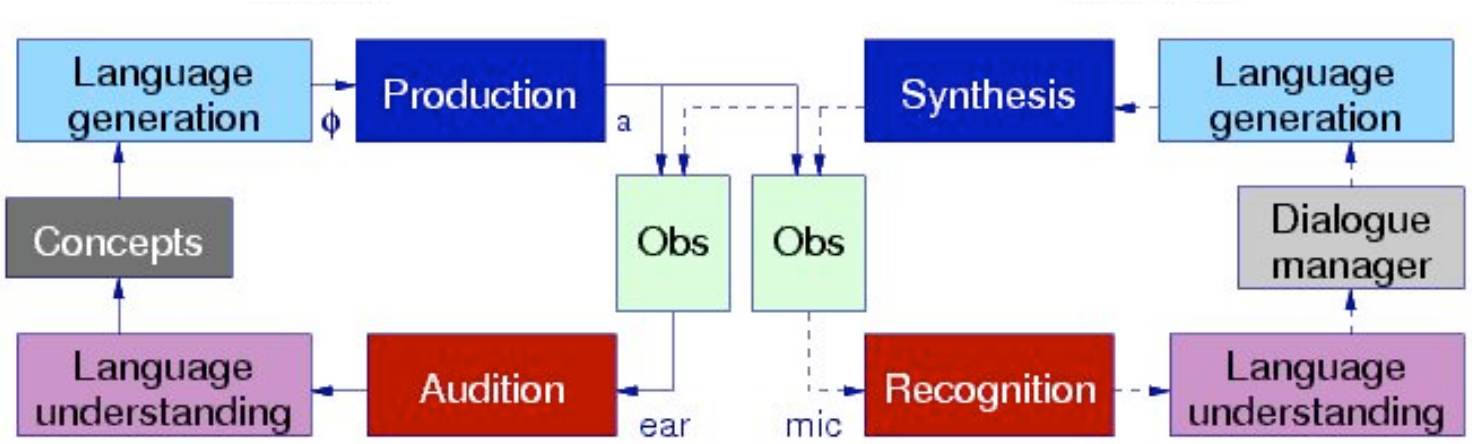
  concept > words > motor commands > sound pressure (mouth) > sound pressure (ears) > auditory nerve impulses > words > received meaning



Speech control loops

- In normal dialogue, there are 3 feedback loops controlling speech:
  - Sensory feedback of articulator motion
  - Auditory feedback of sound produced by the speaker
  - Conversational responses from the listener

- Feedback control provides resilience and robustness:
  - For language acquisition (i.e., for learning to speak)
  - For directing dialogue to achieve common understanding
  - For speaking with damaged or obstructed articulators
  - For speaking in an adverse noise environment



A message: an idea in words

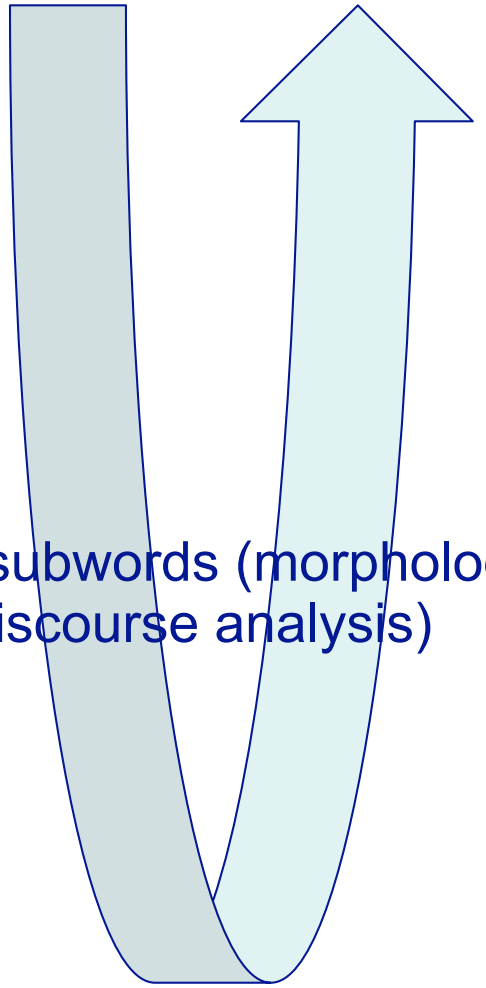
- We develop our understanding of the world at the same time as we learn to talk:
  - Many of our concepts and basic ideas are formed in words;
  - So, language strongly influences the way we think!

- Ideas are concepts or **abstract** notions
- Words reveal the **surface** form of a language, which is structured by grammar and syntax

- A spoken utterance represents our ideas and intentions in the form of words and sounds

Levels of encoding and interpretation

- Society
  - social function or aim
- Pragmatics
  - communicative intent or objective
- Semantics
  - idea or meaning
- Syntax
  - structure of sounds (phonology), of subwords (morphology), of words (grammar), or dialogues (discourse analysis)
- Empirics
  - speech patterns (phonetics)
- Physics
  - acoustic signal



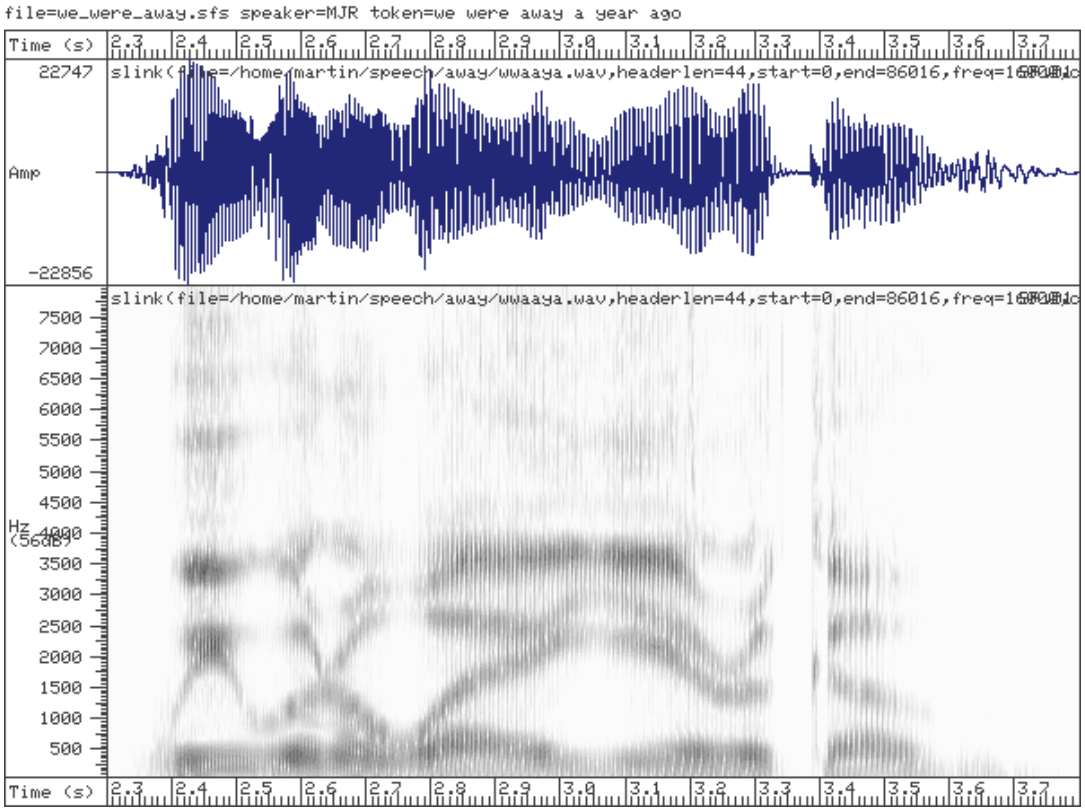
Relating words, articulation and sounds of a language

- Phonology
  - function of sound units within a language
  - how sound units are used to make words
  - **phoneme**
- Phonetics
  - acoustic result of speech articulation
  - how speech sounds are produced
  - **phone**

Speech is not acoustic text

- Written language
  - discrete words separated by spaces
  - usually complete, correct spelling
  - opportunity to skip, skim or re-read

- Spoken language
  - continuous sequence of sounds, usually without spaces
  - often damaged, interrupted, parts mumbled



Written and spoken words

- **Orthographic** (written) and **phonetic** (spoken) forms
- Grapheme-to-phoneme (letter to sound) mapping is not 1-to-1:
  - Some sounds require several letters
    - e.g., “sh”, “ph”
  - Some letters have several pronunciations
    - e.g., “g”, “c”
  - Some sounds have several transcriptions
    - e.g., /f/: “f” and “ph”
  - Some letters produce several sounds
    - e.g., “x” /ks/
  - Some combinations have complex relations
    - e.g., “-ough-”
  - Different accents alter some phonemes
    - e.g., “bath”, “food”

Speech units: vowels, consonants and syllables

- Vowels
  - Vibrating vocal cords in larynx with clear vocal tract
  - Produced using slower extrinsic muscles
- Consonants
  - Usually some occlusion of the vocal tract
  - Sound source can be from larynx, click or hiss
  - Produced using faster intrinsic muscles
- Syllables
  - All languages have CV syllables
  - Basic unit of articulation
  - Vowel glides and consonant clusters

Beyond sound units

- **homophones** or homonyms
  - “to”, “too”, “two”
  - “hear”, “here”
  - “glasses”, for seeing or drinking
- ambiguity of **segmentation**
  - “grey tape” or “great ape”
  - “how do you wreck a nice beach?”
- **intonation** changes meaning
  - “He’s gone” or “He’s gone?”
- **emphasis** or stress
  - “the cat sat on the mat”

Sources of acoustic variability

- Population-specific differences
  - language, dialect, accent, class, style,…
- Individual differences
  - vocal-tract length, sex, age, health, mood,…
- Prosody
  - timing, pitch and intonation
  - intensity, stress and emphasis
- Noise
  - reverberation
  - background noise
  - transmission or coding artefacts

Speech communication summary

- Speech is the natural human modality of interaction
  - An utterance is an acoustical encoding of a message
  - Feedback in the speech loop optimises communication
- Spoken language's characteristics
  - Written vs. spoken language (phonology)
  - Continuous acoustic signal (phonetics)
  - Vowels and consonants in syllables, words and dialogue
- Variability of the speech signal
  - Accent, speaker, performance, noise conditions