

COM3502-4502-6502 SPEECH PROCESSING

Lecture 5

The Nature of Speech

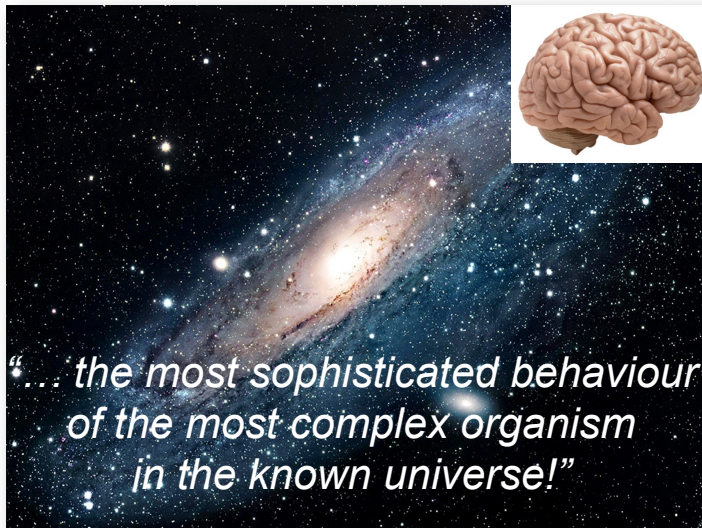


The
University
Of
Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 5, slide 1

Speech is ...

Moore, R. K.
(2007). Spoken
language
processing:
piecing together
the puzzle.
*Speech
Communication*,
49, 418-435.



*"... the most sophisticated behaviour
of the most complex organism
in the known universe!"*

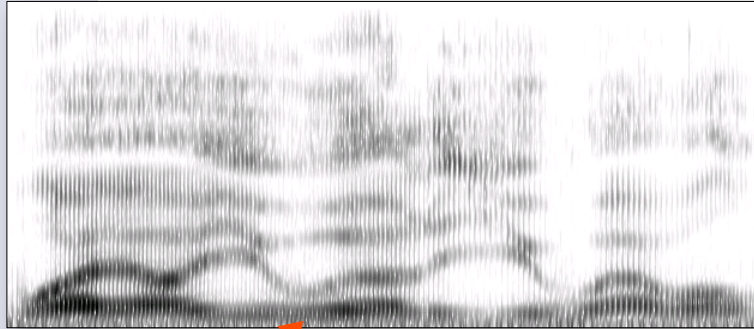


The
University
Of
Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 5, slide 2

Speech is ... *continuous*

"Why are you early you owl?"



Can't see where one word stop and another begins (*there are no gaps*)

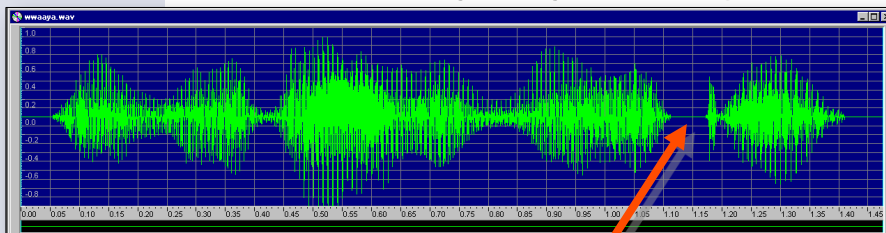


The University
Of
Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 5, slide 3

Speech is ... *continuous*

"We were away a year ago."



The only silence is
the 'g' in "ago" !



The University
Of
Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 5, slide 4

Speech is ... *variable*

العربي български català 中国话 hrvatski česky
english ελληνικά עברית हिंदी italiano 日本語
한국어 românește русский српски ภาษาไทย



There are >7,700,000,000 people
in the world

... and they all speak differently



The
University
Of
Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 5, slide 5

Speech is ... *variable*



There are 3000-8000 languages:

- 6500 'living' languages
- 83% are limited to single countries
- 52% are spoken by less than 10,000 people
- 28% are spoken by less than 1000 people

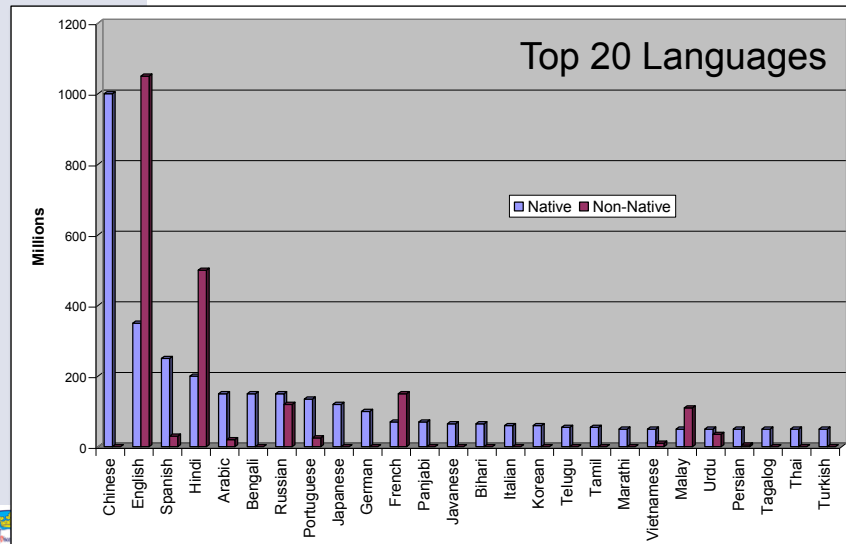
- According to the National Geographic Society, one language dies out every two weeks
- Half of the world's languages will be out of use by 2100



The
University
Of
Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 5, slide 6

Speech is ... *variable*



The University of Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 5, slide 7

Speech is ... *variable*



- Within any given language, groups of individuals use local dialects and accents
- A '**dialect**' is when different words are used
 - “*snotter clot*” is Geordie (*Newcastle*) for “*handkerchief*”
 - “*china*” is Cockney rhyming slang for “*mate*” (short for “*china plate*”)
 - “*beck*” is Yorkshire (*Old Norse*) for “*stream*”
- An '**accent**' is when different sounds are used
 - “*bath*” and “*path*” (rhymes with “*cat*” in the north of England)
 - “*bath*” and “*path*” (rhymes with “*car*” in the south of England)
- Accents and dialects reflect regional and/or social differences



The University of Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 5, slide 8

Speech is ... *variable*

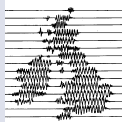
The British Isles



"When a sailor in a small craft faces the might of the vast Atlantic Ocean today, he takes the same risks that generations took before him"



ABI



Accents of the British Isles



The University Of Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 5, slide 9

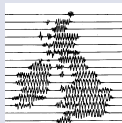
Speech is ... *variable*

If you're interested, check out ...

<http://www.bbc.co.uk/voices/recordings/index.shtml>



ABI



Accents of the British Isles



The University Of Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 5, slide 10

Speech is ... *variable*



'Between'

- Even within a given language/dialect/accent group, different individuals sound different
- 'Inter-speaker variation' is caused by:
 - age (e.g. *child, teenager, adult, senior citizen*)
 - gender
 - physical characteristics (e.g. *vocal tract size*)
 - social habits

'Within'

- 'Intra-speaker variation' can arise from:
 - physiological factors (e.g. *bad health, fatigue*)
 - psychological factors (e.g. *mood, emotion*)
 - external factors (e.g. *the environment*)



University
Of
Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 5, slide 11

Question ...

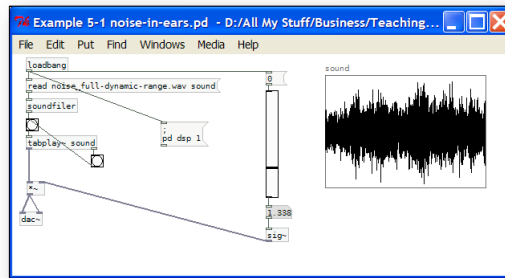
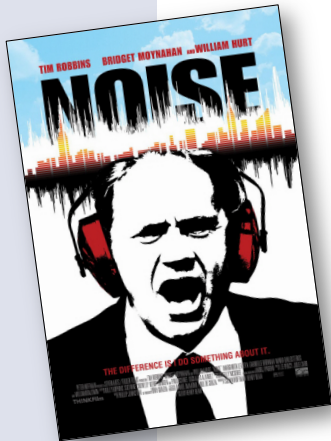
What 'external factors' might influence the way someone speaks?



The
University
Of
Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 5, slide 12

External Factors



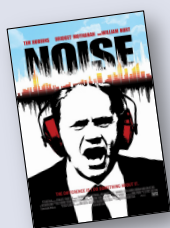
Lombard, E. (1911). Le sign de l'élévation de la voix. *Ann. Maladies Oreille, Larynx, Nez, Pharynx*, 37, 101–119.



The University
Of
Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 5, slide 13

External Factors



- Noise
 - hyper-articulation
 - the 'Lombard Effect'
- Vibration
 - chest
 - vocal tract
- The task
 - casual conversation
 - reading out loud
 - lecturing
- The listener
 - a child ('parentese')
 - a non-native person
 - a hearing-impaired individual
 - an animal
 - a machine (!)
- The cognitive load
 - interaction with other tasks
 - stressful/emotional situations
- Alcohol/drugs
 - impaired motor control



The University
Of
Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 5, slide 14

Speech is ... *variable*



“fish and chips” → “fish ’n’ chips”

“temporary” → “tem’pry”

“can be” → “cam be”

“bread and butter” → “bre’m butter”

This is not ‘sloppy’ speaking!

This is a normal part of communicative behaviour that balances the demands of *speaking* against the demands of *listening*



The University
Of
Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 5, slide 15

Speech is ... *variable*



- People can (*choose to*) speak:

- loudly *or* softly
- quickly *or* slowly
- clearly *or* mumbled
- formally *or* casually

... and this can change in the course of a single utterance

- There is also inherent variation

- even a trained phonetician cannot produce the same word in *exactly* the same way on different occasions



The University
Of
Sheffield.

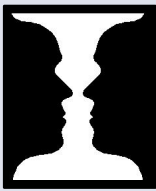
COM3502-4502-6502 Speech Processing: Lecture 5, slide 16

Speech is ... *ambiguous*



- Variability means that signals we'd like to be the same are actually quite different ...

speech speech speech speech speech



- Ambiguity arises when signals we'd like to be different end up being the same ...

“to” vs. “two” vs. “too”
“hear” vs. “here”

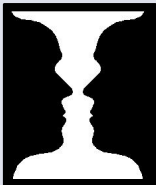
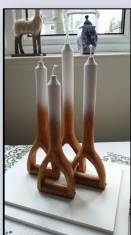
← ‘homophones’



The University Of Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 5, slide 17

Speech is ... *ambiguous*



“fork handles” vs. “four candles”



The University Of Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 5, slide 18

Speech is ... *ambiguous*

“four candles” vs. *“fork handles”*

“great ape” vs. *“grey tape”*

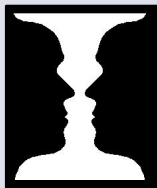
“law and order” vs. *“Laura Norder”*

“Joe is office head” vs. *“Joe is off his head”*

“This nudist play will wreck a nice beach”

vs.

“This new display will recognise speech”



Speech is ... *language+*

What's
being
said

- Rich in *‘linguistic’* information
 - *“that isn’t it, is it?”*
 - *“what’s in the road ahead?”* vs. *“what’s in the road, a head?”*

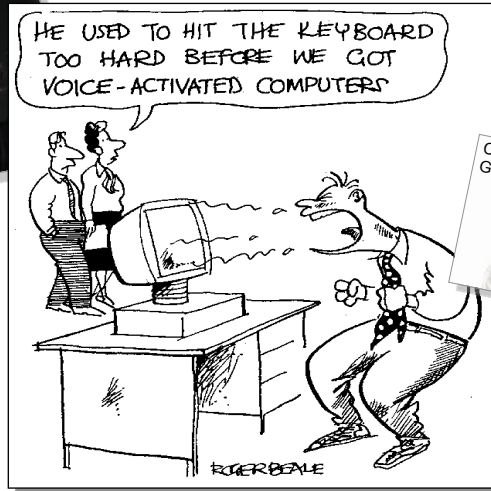
How it's
being
said

- Rich in *‘para-linguistic’* information
 - individuality
 - attitude
 - emotion

Other
speaker-
generated
behaviour

- Rich in *‘extra-linguistic’* information
 - breathing noises
 - lip-smacks

Speech is ... *emotional*



The University
Of
Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 5, slide 21

Speech is ... *d'd'dis-disfluent*

- Normal spontaneous speech contains:
 - false-starts
 - repeats
 - filled pauses (*"uhms"* and *"errs"*)
 - overlaps
- Spoken utterances are not as 'well-formed' as written sentences
- So-called 'disfluencies' actually make speech easier (*for a human*) to produce and understand



The University
Of
Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 5, slide 22

Speech is ... *multimodal*



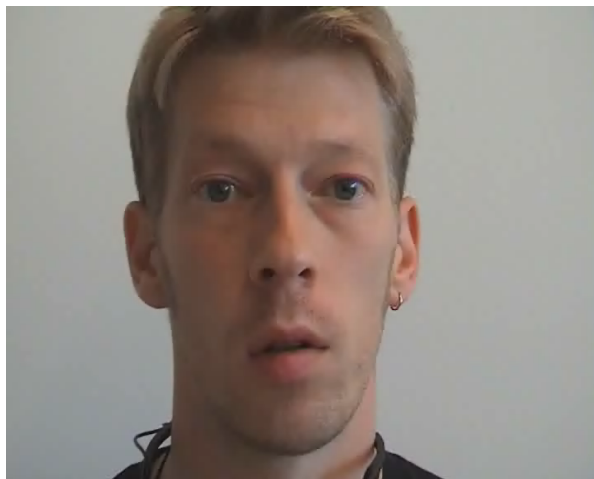
- Speech is not just an acoustic signal, it is also visual
- Listeners can derive a large benefit from seeing a speaker's lips
 - ~12 dB signal-to-noise gain
- The visual information can even *override* the acoustic information
 - the 'McGurk Effect'



The University Of Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 5, slide 23

The 'McGurk' Effect



The University Of Sheffield.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.

COM3502-4502-6502 Speech Processing: Lecture 5, slide 24

Speech is ... *contaminated*

- Everyday speech occurs in normal environments (*not recording studios*) which include:
 - other voices
 - competing sounds
 - noise & distortion
 - reverberation



- It's a bit like this ...

“xn xpplx x dxy kxxps thx dxctxr xwxy”

... but, it's much more like this:

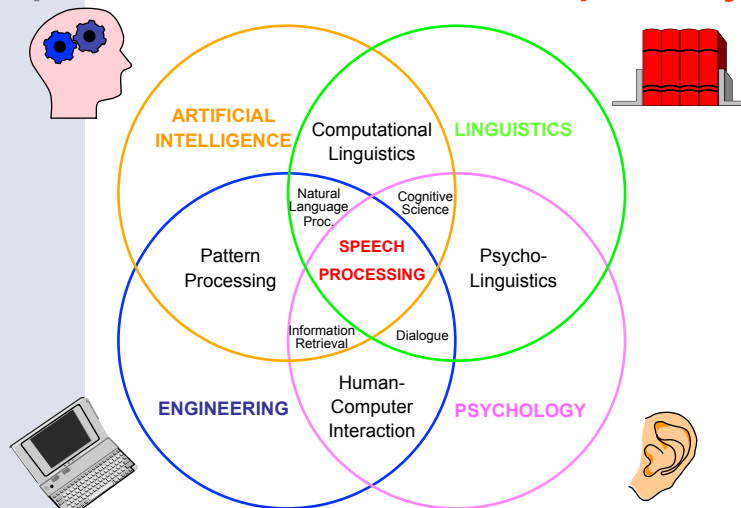
“ax axxxe a xax xeexx xxe xoxxex axax”



The University
Of
Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 5, slide 25

Speech is ... *multidisciplinary*



The University
Of
Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 5, slide 26

This lecture has covered ...

- Why speech is an *interesting* signal to process
- Why speech is a *challenging* signal to process ...
 - continuity
 - variability & ambiguity
 - linguistic richness
 - emotion & disfluency
 - multimodality & contamination
 - multi-disciplinarity



The
University
Of
Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 5, slide 27

Any Questions ?



The
University
Of
Sheffield.

COM3502-4502-6502 Speech Processing: Lecture 5, slide 28

Next time ...

Sounds and Symbols



The
University
Of
Sheffield.