Speech is …

“… the most sophisticated behaviour of the most complex organism in the known universe!”

Speech is … *continuous*

“*Why are you early you owl?*”

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

---

Speech is … *continuous*

“*We were away a year ago.*”

The only silence is the ‘g’ in “ago”!

Speech is … *variable*

There are >7,500,000,000 people in the world … and they all speak differently

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
Top 20 Languages

- Chinese
- English
- Spanish
- Hindi
- Arabic
- Bengali
- Russian
- Portuguese
- Japanese
- German
- French
- Panjabi
- Javanese
- Bihari
- Italian
- Korean
- Telugu
- Tamil
- Marathi
- Vietnamese
- Malay
- Urdu

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
Accent Variation in the UK

“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”

If you’re interested, check out ...
http://www.bbc.co.uk/voices-recordings/index.shtml
Speech is ... *variable*

- 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

- *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)

**Question ...**

What ‘external factors’ might influence the way someone speaks?
External Factors

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

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*.

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
Speech is … ambiguous

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

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

- “to” vs. “two” vs. “too”
- “hear” vs. “here”
- “fork handles” vs. “four candles”
Ambiguity in Speech

“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

- HE USED TO HIT THE KEYBOARD TOO HARD BEFORE WE GOT VOICE-ACTIVATED COMPUTERS.

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
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 ‘McGurk’ Effect**

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 xxwxy"

  … but, it’s much more like this:

  "ax axxxe a ax xeeex xxe xxxxxe axax"

Speech is … **multidisciplinary**
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

Any Questions?
Next time …

Sounds and Symbols