Ultrax2020: Ultrasound-based Technologies for Speech Therapy

Aciel Eshky, Manuel Sam Ribeiro, Korin Richmond, and Steve Renals
Centre for Speech Technology Research, School of Informatics, University of Edinburgh

THE UNIVERSITY *of* EDINBURGH **informatics**

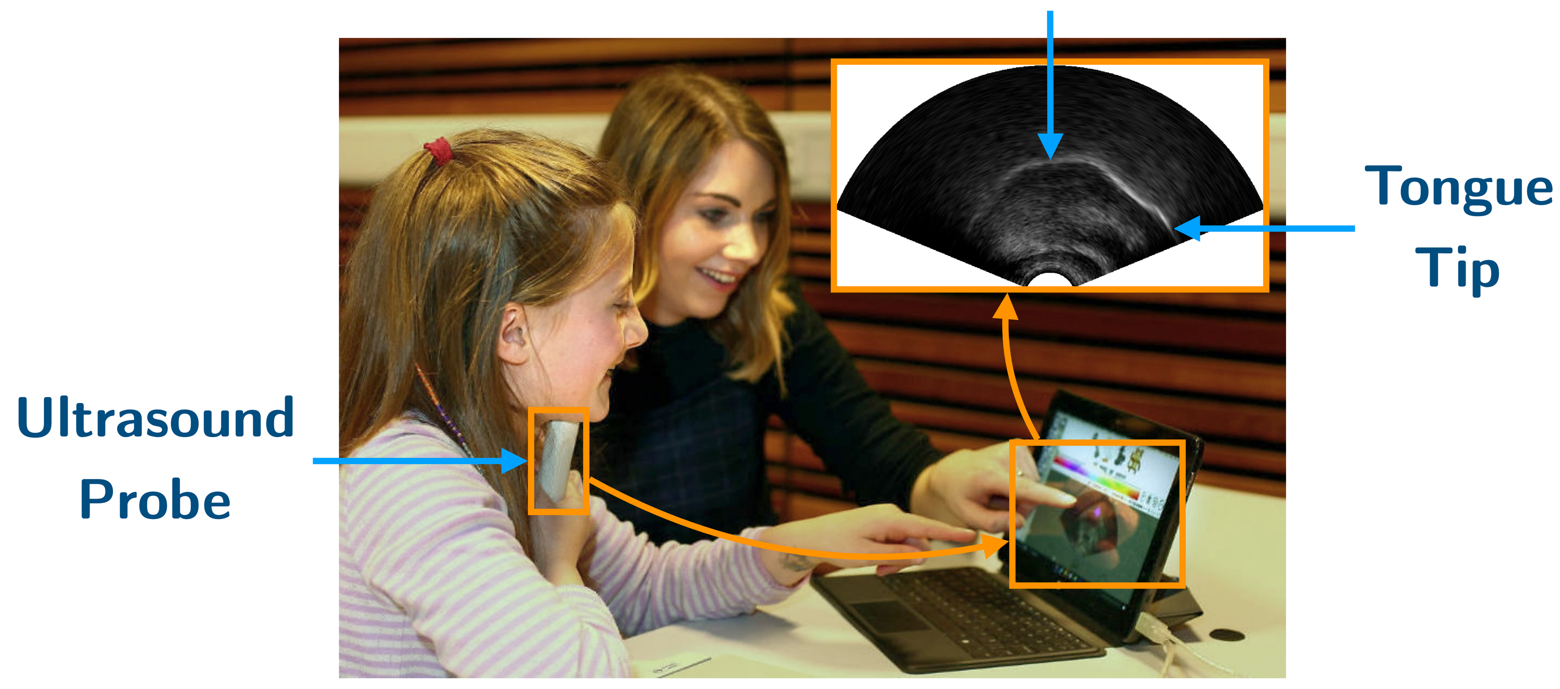
Speech Sound Disorders

- Native English speakers master vowel & consonant production by age 8
- **11.4%** of UK **8 year-olds** suffer from speech sound disorders ranging from **mild** (common distortions) to **severe** (unintelligible)
- Patients suffer long-term challenges: social, psychological, educational

Speech therapy is amenable to automation

- Current clinical practice is **perception-based**:
  Speech and language therapists (**SLTs**) elicit prompts from patients and phonetically transcribe what they hear
  Pros: fast
  Cons: subjective and can be unreliable (low SLT agreement)

- New clinical approach utilises **instrumental methods**:
  SLTs record the patient's **speech audio** augmented with **ultrasound tongue imaging**:



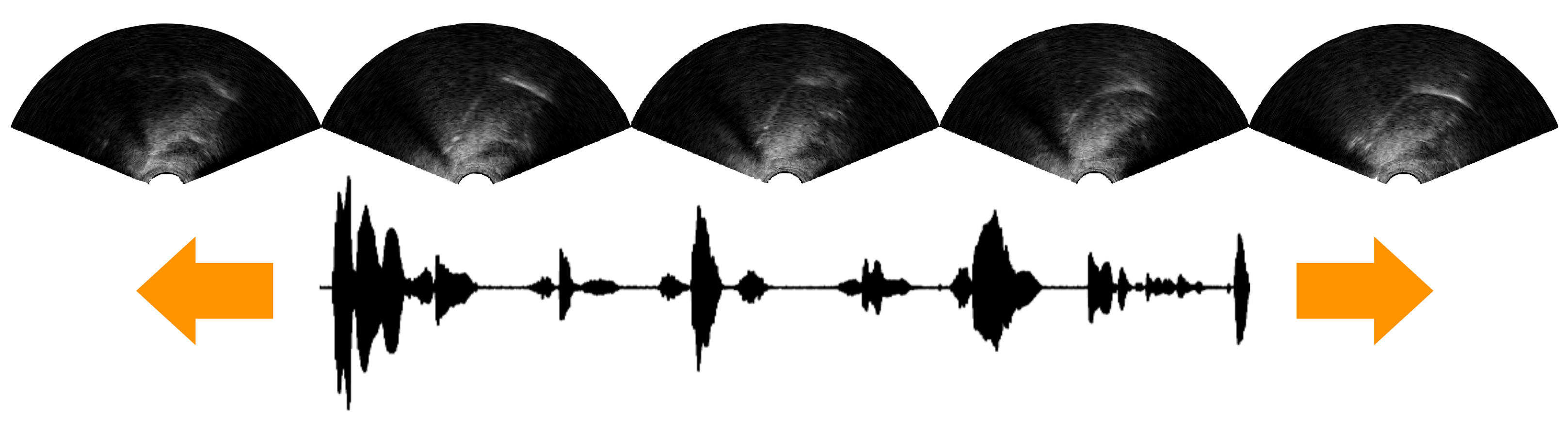
  Pros: visual biofeedback reduces subjectivity and increases reliability (higher SLT agreement)
  Cons: recorded data is time-consuming to manually process by SLTs
  → instrumental methods not yet clinically viable

- **Ultrax2020: automate** data processing using **machine learning**
  - Minimise manual processing → make instrumental methods viable
  - Example tasks shown below (1-6)

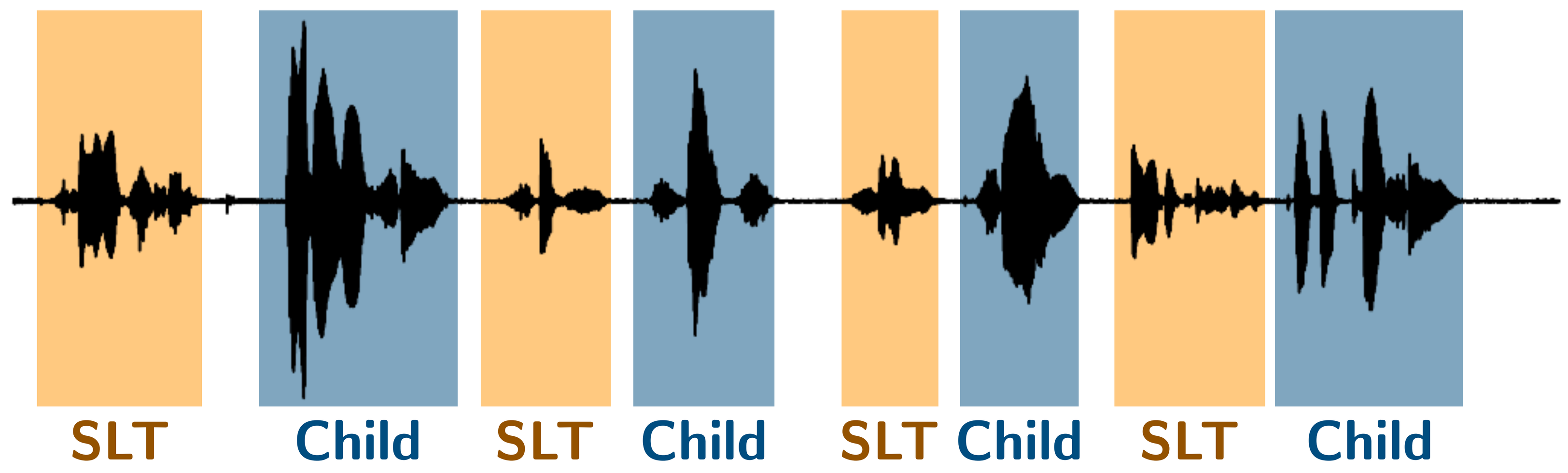
1. Identify ultrasound probe direction using image classification

Flip all recordings to the same direction for consistency

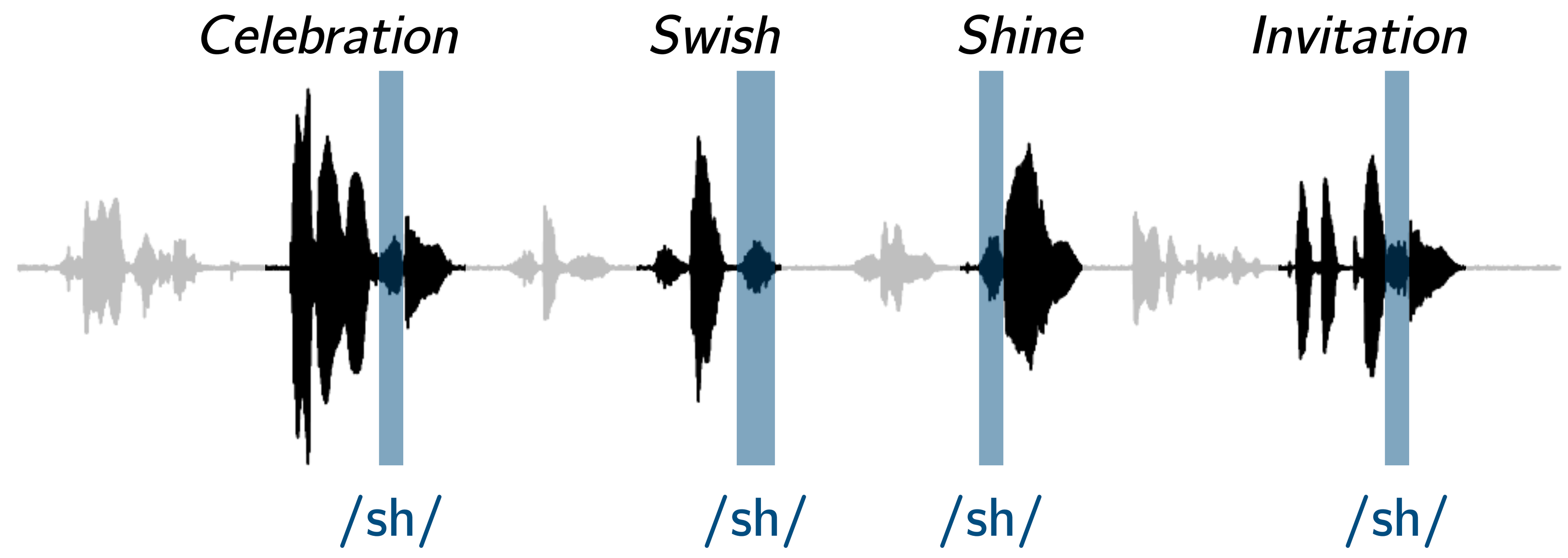
2. Synchronise audio and ultrasound by learning cross-modal embeddings



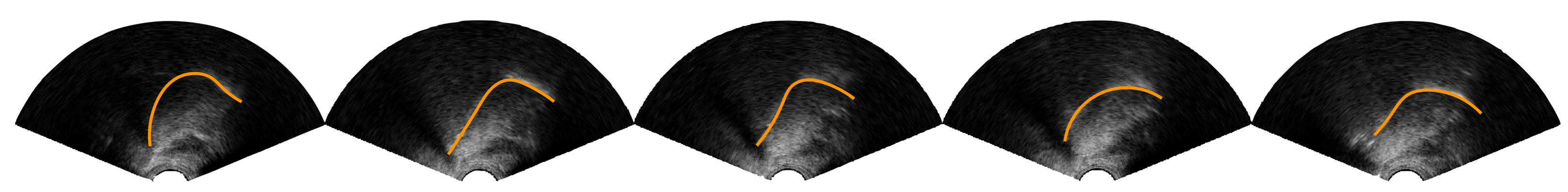
3. Extract child speech using speaker diarisation



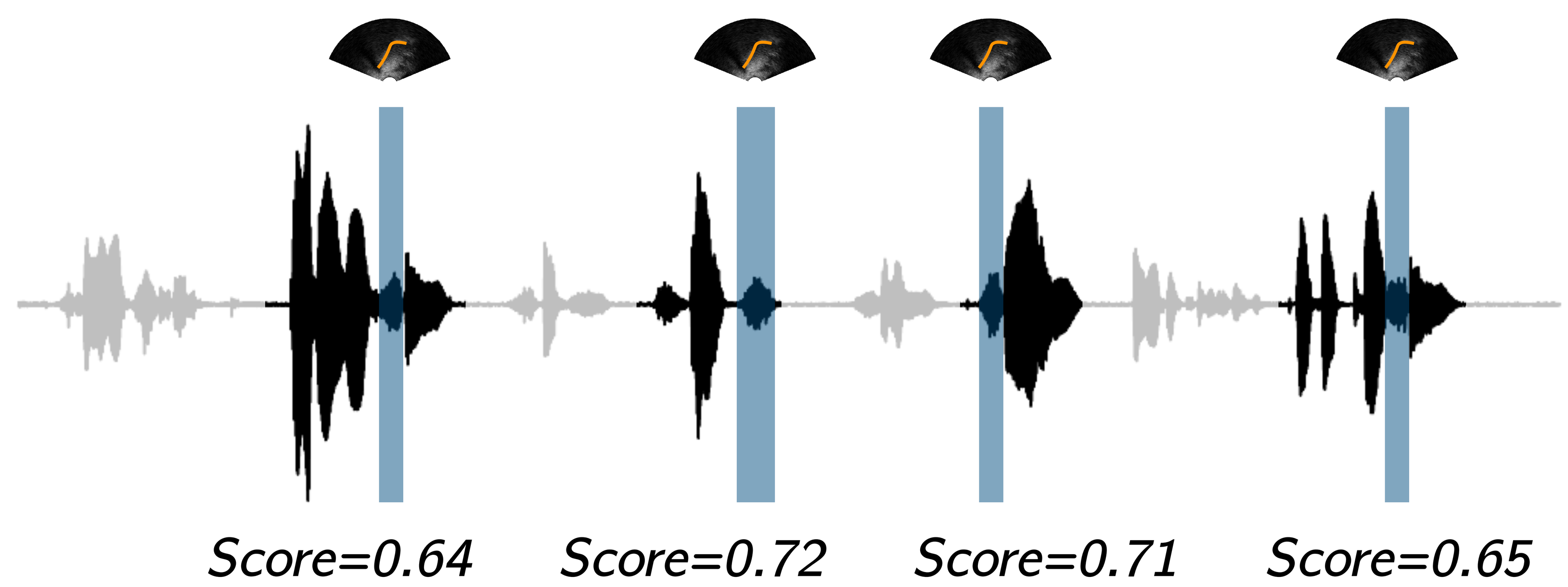
4. Locate target phones using phone alignment



5. Extract tongue surface using recursive Bayesian estimation



6. Determine the goodness of pronunciation using neural networks and adversarial training



Project Outputs: UltraSuite

A collection of data and software available to the research community

Data:
- Typically developing children
- Children with speech sound disorders
- Children with cleft palate (extend techniques to different subgroup)
- Adult Data (improve model performance by augmenting child data with adult data during training)

Software:
- To process and visualise data
- To perform machine learning tasks above

Collaborators and Clinical Partners

- Joanne Cleland and Ellie Sugden from the **University of Strathclyde**
- Alan Wrench from **Articulate Instruments Ltd.**
- **NHS Lothian**, **NHS Greater Glasgow and Clyde**, **NHS Grampian**, **The Nuffield Hearing and Speech Centre** (NHS UCL Hospitals)

Supported by EPSRC Healthcare Partnerships Programme Grant Number EP/P02338X/1 (Ultrax2020)

AI for Social Good 2020, Edinburgh, UK

Contact: {a.eshky, sam.ribeiro, korin.richmond, s.renals}@ed.ac.uk