AUTOMATIC SCREENING OF CHILDREN WITH SPEECH SOUND DISORDERS USING PARALINGUISTIC FEATURES

Mostafa Shahin[1,2], Beena Ahmed[1], Daniel V. Smith[2], Andreas Duenser[2], Julien Epps[1,2]

[1]School of Electrical Engineering and Telecommunications, UNSW, Sydney, Australia
[2]Data61, CSIRO, Tasmania, Australia

Speech Sound Disorder (SSD)

❑ Mouth, jaw, tongue, lips, palates and other articulators all work together to produce human speech.

❑ Failure in moving these parts appropriately affects speech intelligibility and causes speech disorders.

❑ Causes of speech disorder in children, including:

➤ Muscle weaknesses, e.g. dysarthria

➤ Neurological disorder, e.g. apraxia

➤ Structural abnormalities, e.g. cleft lip/palate

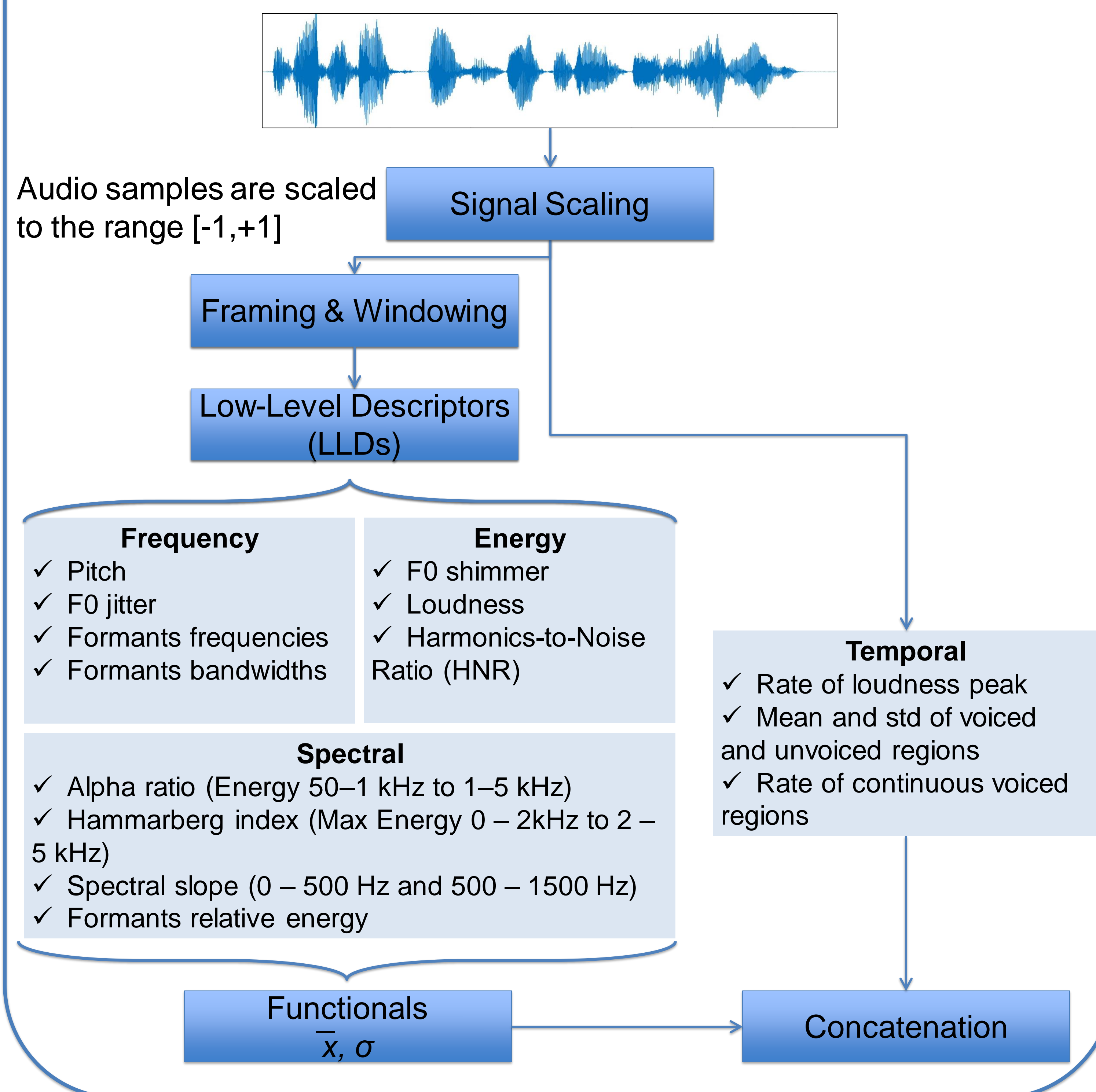➤ Unknown, e.g. articulation/phonological disorder
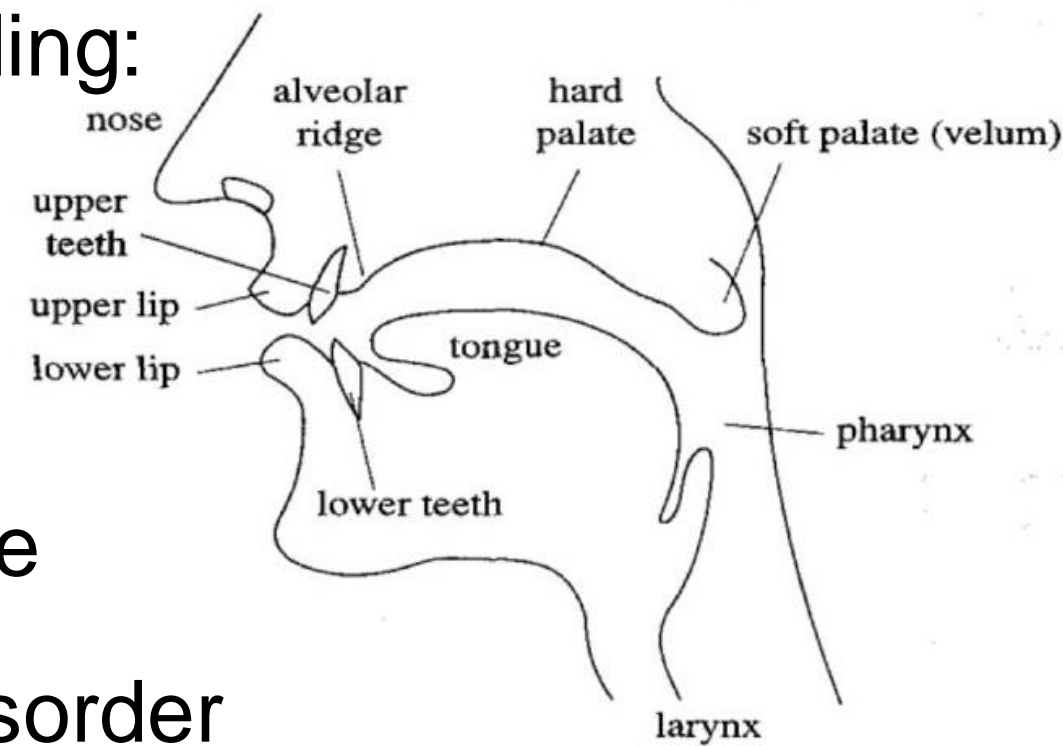
Motivation

❑ Generally, children do not get assessed; parents and teachers often do not notice speech problems until too late.

❑ Delay in treatment can cause serious problems such as social impairment, mental health disorder and learning disabilities.

❑ Subjective screening of children with speech disorders is costly, time consuming and can be infeasible.

❑ Automatic speech analysis offer a practical alternative to human screening.

❑ Restricted research in this area due to lack of speech disorder corpora and unreliability of low-level annotation (e.g. phoneme-level)

Objective

Develop an accurate subject-level automatic screening method for detecting speech disorder in children using high-level acoustic measures.

Paralinguistic Features

❑ A set of low-level descriptors representing the prosodic, spectral and voice quality features of the speech.

❑ Speech disorders affect the prosodic characteristics and quality of speech

❑ Two standard sets, GeMAPS (62 parameters), eGeMAPS (88 parameters)

Audio samples are scaled to the range [-1,+1] → **Signal Scaling**

→ **Framing & Windowing**

→ **Low-Level Descriptors (LLDs)**

Frequency
✓ Pitch
✓ F0 jitter
✓ Formants frequencies
✓ Formants bandwidths

Energy
✓ F0 shimmer
✓ Loudness
✓ Harmonics-to-Noise Ratio (HNR)

Temporal
✓ Rate of loudness peak
✓ Mean and std of voiced and unvoiced regions
✓ Rate of continuous voiced regions

Spectral
✓ Alpha ratio (Energy 50–1 kHz to 1–5 kHz)
✓ Hammarberg index (Max Energy 0 – 2kHz to 2 – 5 kHz)
✓ Spectral slope (0 – 500 Hz and 500 – 1500 Hz)
✓ Formants relative energy

Functionals $\bar{x}, \sigma$ → **Concatenation**

Experiment

Speech Corpus

✓ UltraSuite dataset, a speech recordings and ultrasound repository of speech therapy sessions

✓ 58 Typically Developing (TD) children and 28 children with SSD

✓ Age range from 6 - 10 years old

Classification Method

✓ SVM classifier with 4 fold CV matched by age and gender

✓ Recursive Feature Elimination (RFE) feature selection method

✓ Unweighted Average Recall (UAR) classification metric
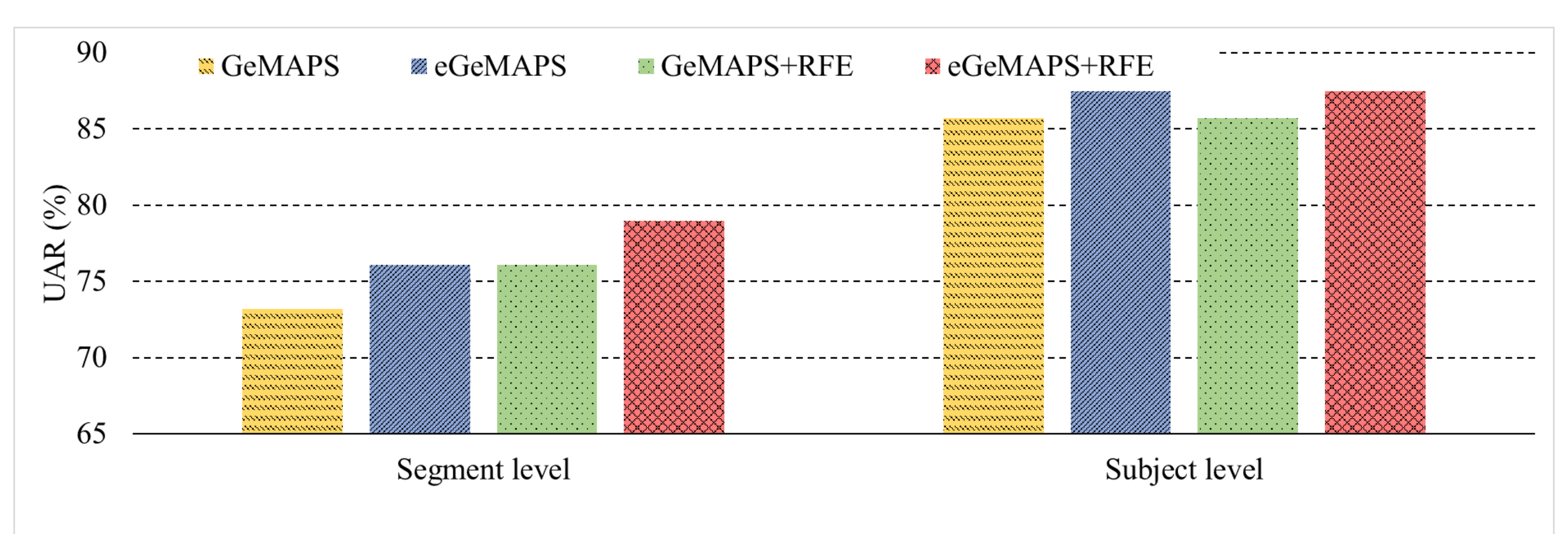
Subset	Type	N# Children Female	N# Children Male	Age
UXTD	TD	31	27	9y 3m ± 1y 10m
UXSSD	SSD	2	6	7y 7m ± 1y 6m
UPX	SSD	4	16	8y 4m ± 2y 2m

CV	N# Training Segments TD	N# Training Segments SSD	N# Test Segments TD	N# Test Segments SSD
Fold 1	4287	9758	482	3349
Fold 2	3995	9895	774	3212
Fold 3	4365	10298	404	2809
Fold 4	4066	9370	703	3737
Average	4178	9830	590	3276
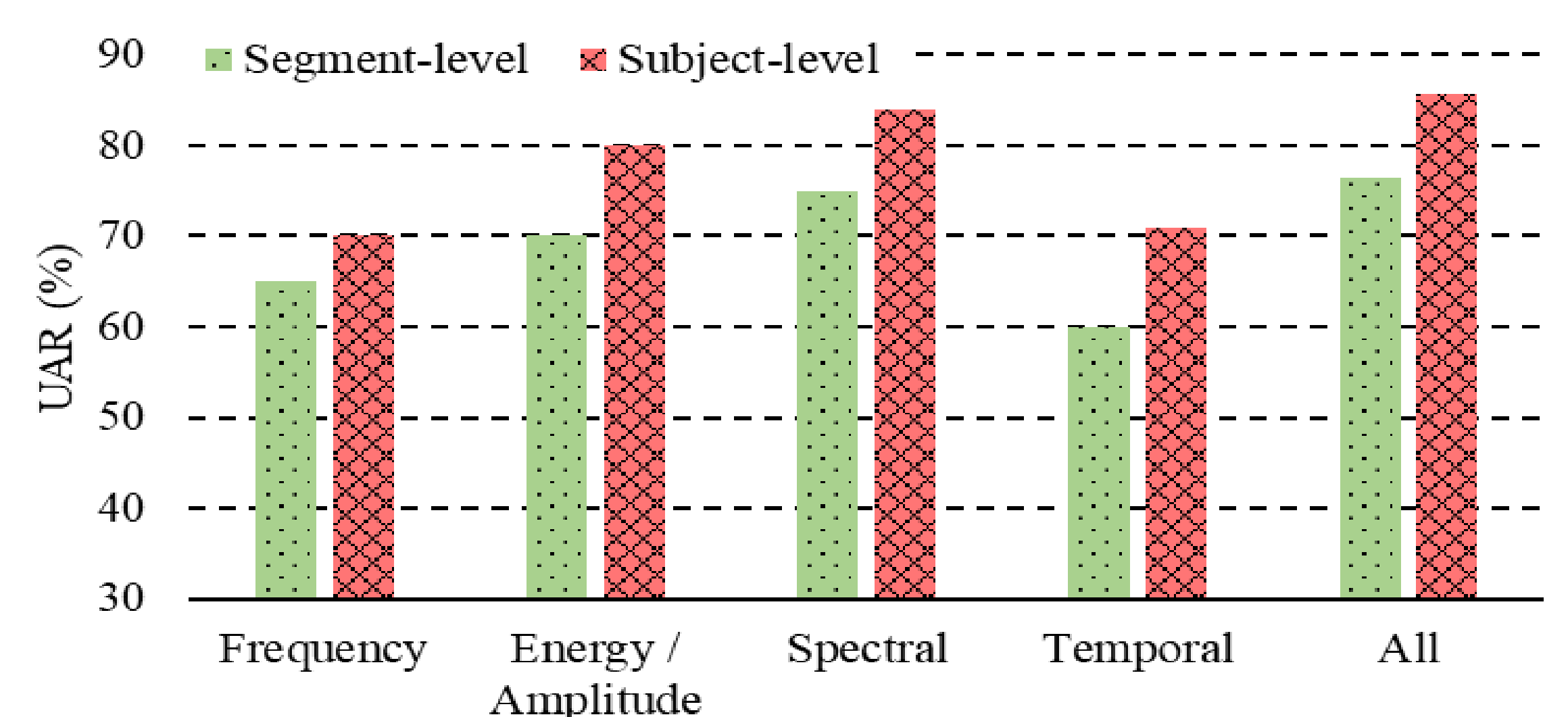
Conclusion

❑ An automatic screening method for children with speech sound disorder based on paralinguistic features.

❑ Comparing between two standard paralinguistic acoustic parameter sets, the GeMAPS and the extended eGeMAPS.

❑ Achieving segment-level and subject-level UAR of ~79% and ~87% respectively.

❑ The extended eGeMAPS feature set, which contains cepstral parameters and more dynamic information, was shown to outperform the abstracted GeMAPS in all experiments.

❑ ~84% subject-level accuracy could be obtained using only the spectral features.

❑ The energy and amplitude related features achieved high subject-level classification score of ~80%.