Dual Acoustic Models and Probabilistic Cross-Lingual Speaker Adaptation for Bilingual Speech Synthesis for a Monolingual Speaker

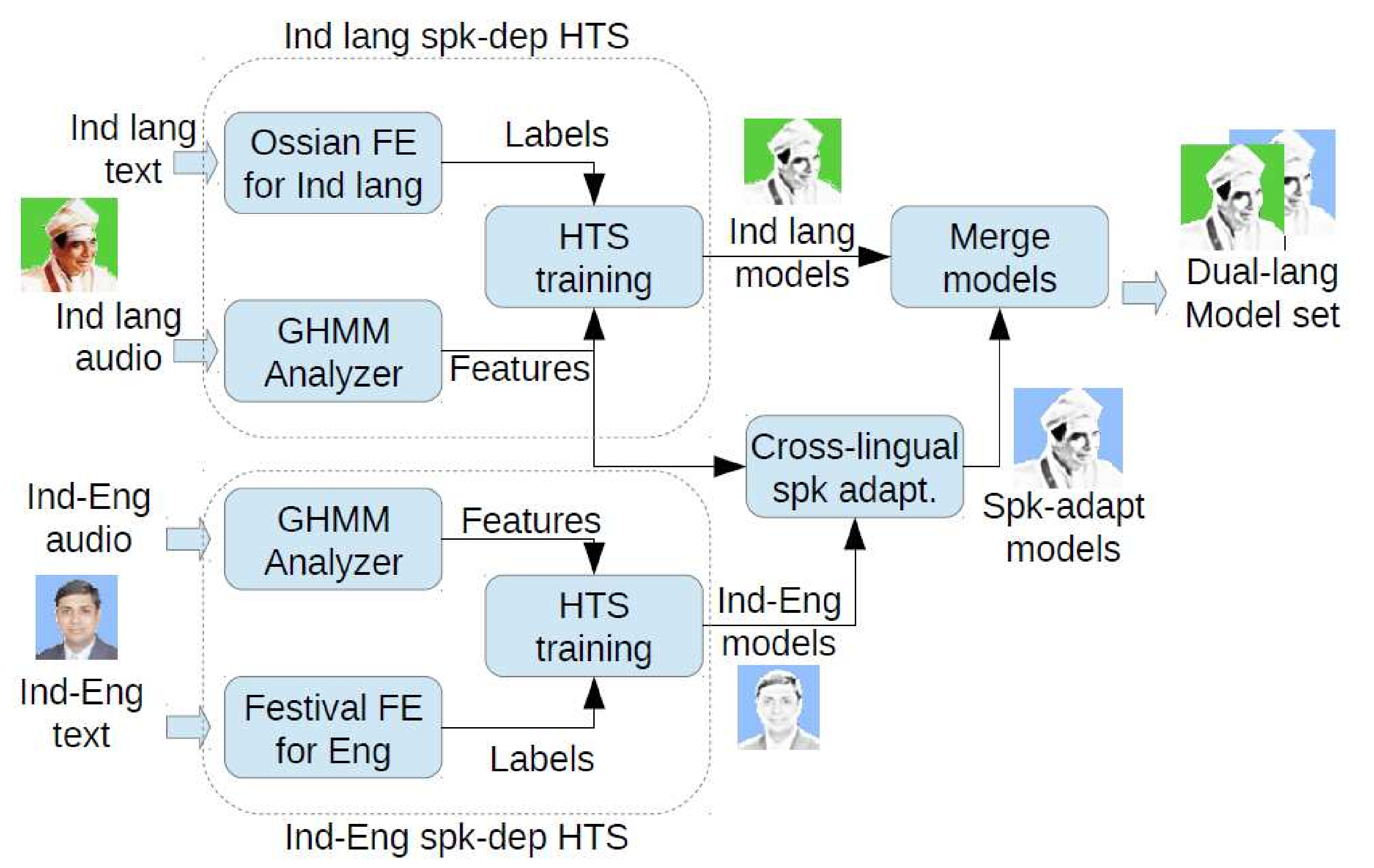
Simple4All Consortium Submission to Blizzard-2014 Spoke Task

I. Objective

- **Task: Bilingual Speech Synthesis for a Monolingual Speaker**
  - **To sythesize dual-language utterances, primarily a native language (Indian) intersperced with words from a non-native language (English)**

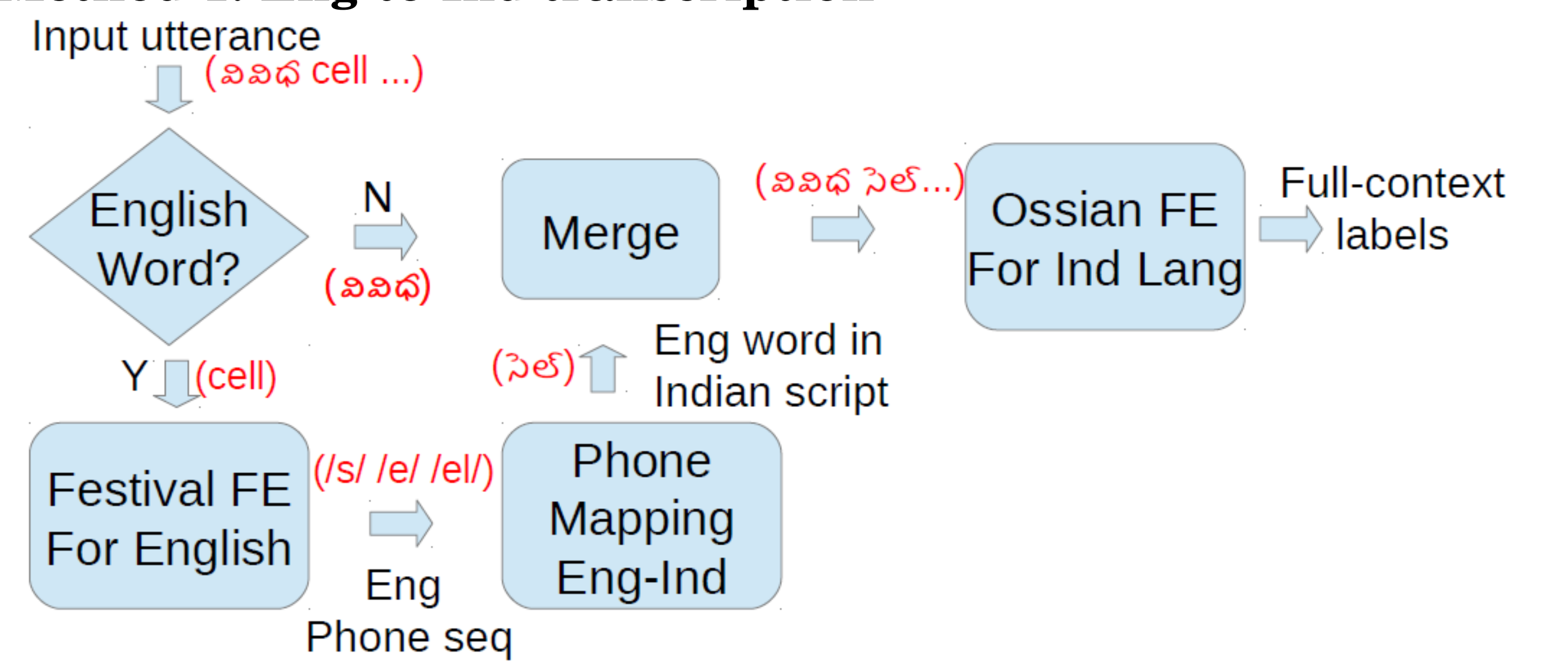
- Training data
  - Single speaker data only in Indian language (a few hundred utterances)
    - Example: "प्रसिद्ध कबीर अध्येता, पुरुषोत्तम अग्रवाल का यह शोध आलेख, उस रामानंद की खोज करता है “
  - Audio data (16kHz, 16 bits) along with text in Indian script (UTF-8)
- Test data
  - Example: “Under 19 cricket world cup में सोमवार को अफ़गानिस्तान ने ऑस्ट्रेलिया को हराकर, बड़ा उलटफेर किया है”

II. Dual Language Acoustic Modeling - Overview

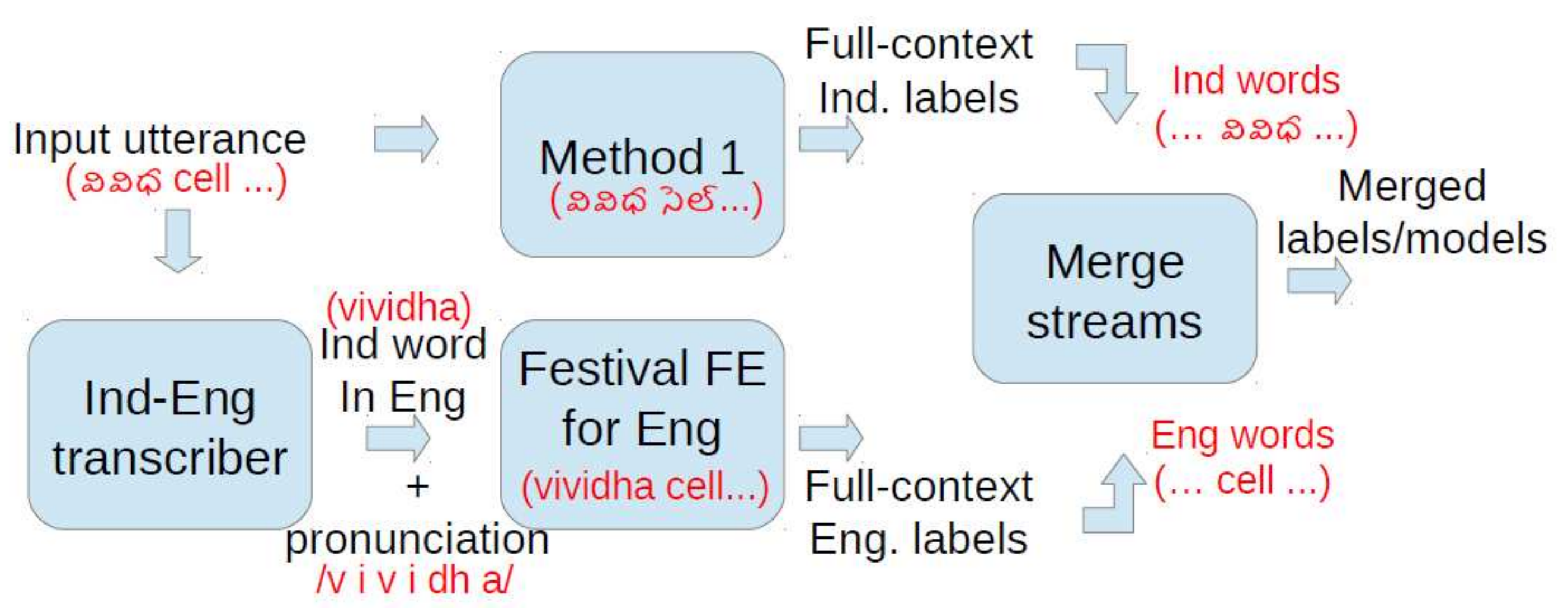


III. Label Generation

Method-1: Eng-to-Ind transcription



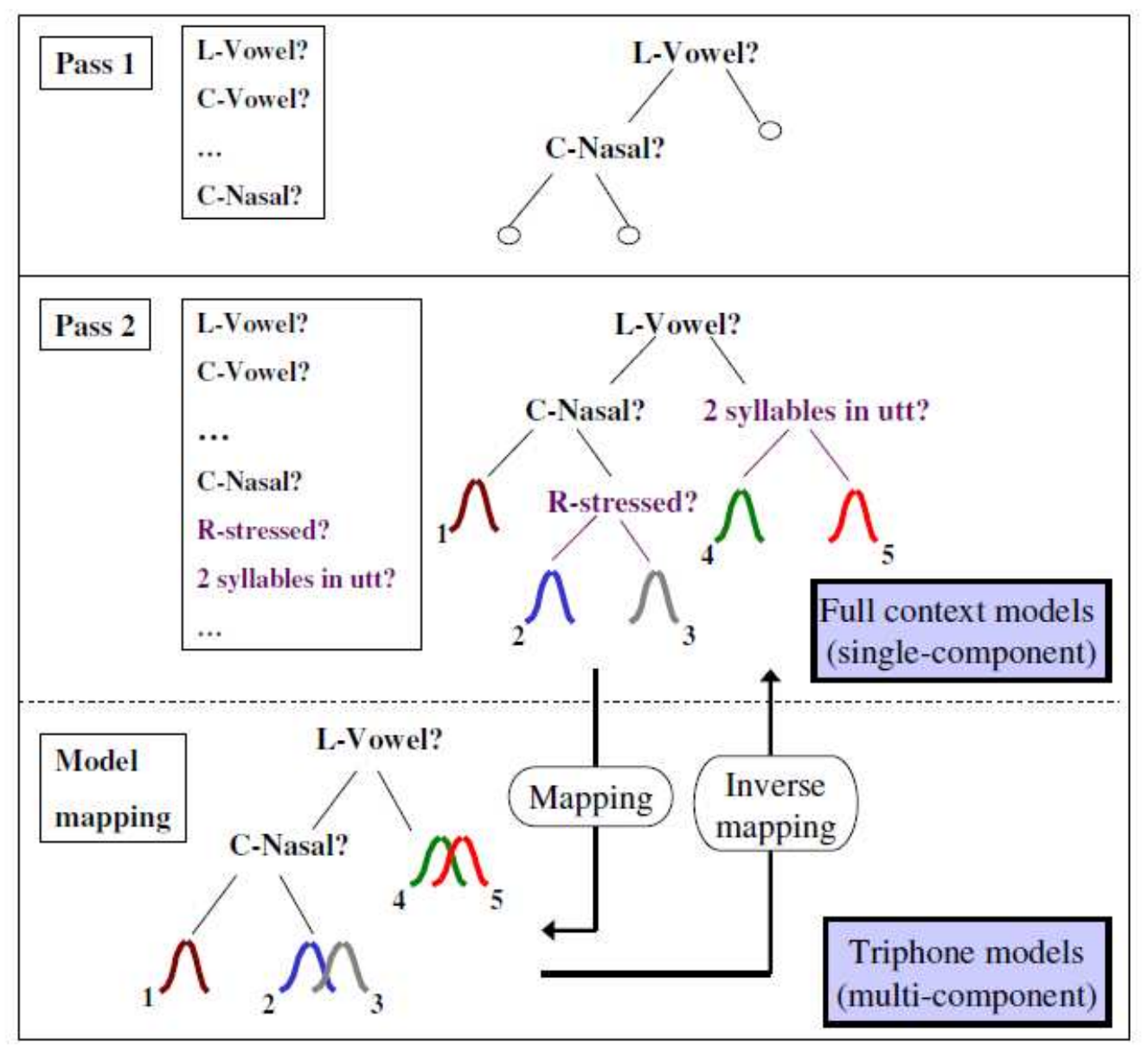
Method-2: Dual front-end with filler words



IV. Acoustic Models training

- **GlottHMM based TTS models**
- **Speaker dependent models for Indian language**
- **Cross-lingual adaptation of an Indian accented English voice using Indian language data**
  - **Two pass decision tree generation**
  - **Decode Indian language data using triphone models**
  - **CMLLR adaptation of the English model set**
  - **Merge Indian speaker-dependent and adapted English acoustic model sets**

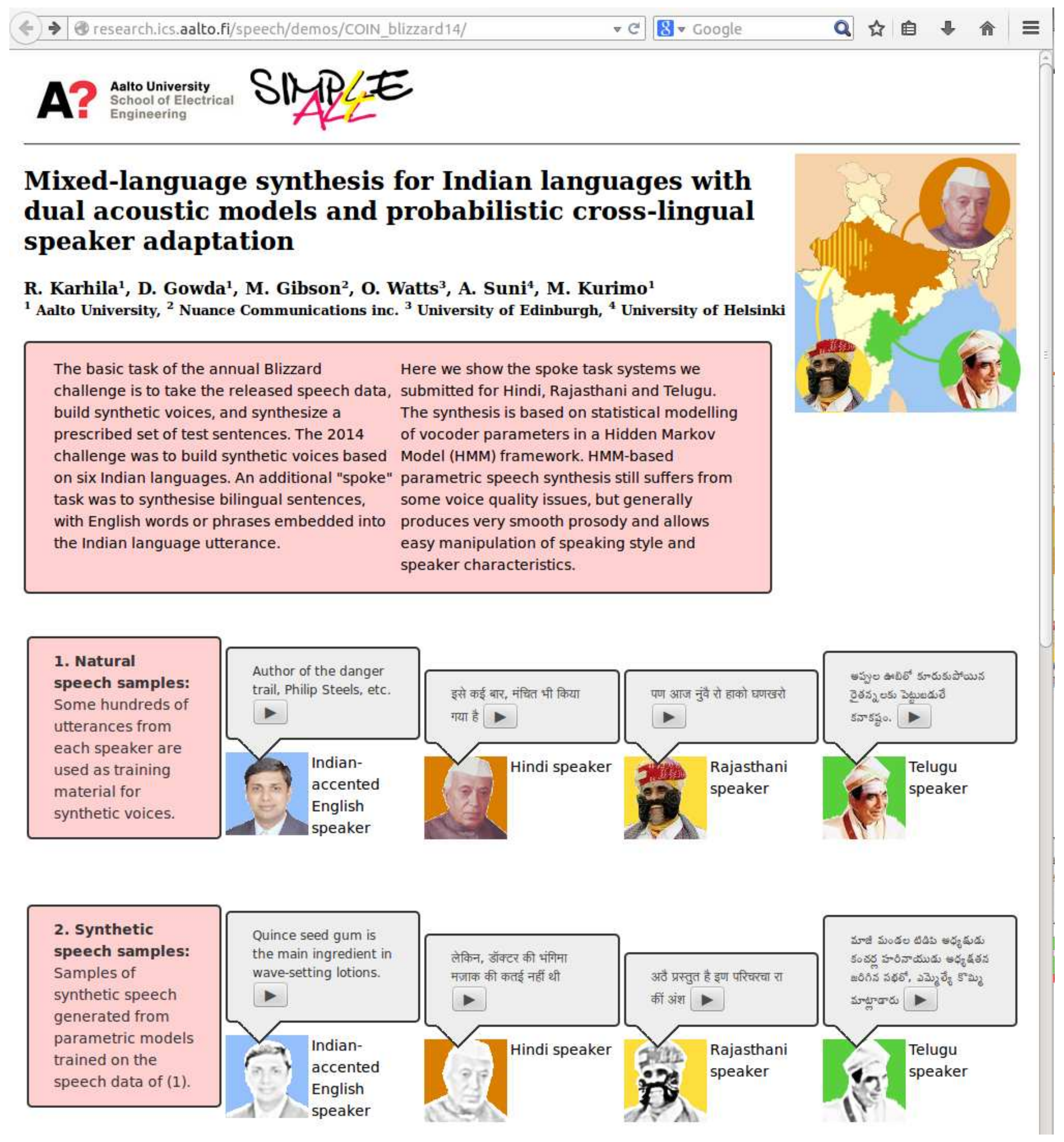
V. Two-pass decision tree generation



- **1st pass: Generate tree for ASR type triphones**
- **2nd pass: Extend ASR tree to TTS type full-context tree**
- **Train the TTS tree leaf Gaussians**
- **Finally pool all TTS Gaussians under each ASR tree leaf to form Gaussian mixture models**

VI. Demo Page

http://research.ics.aalto.fi/speech/demos/COIN_blizzard14/



research.ics.aalto.fi/speech/demos/COIN_blizzard14/
Google
Aalto University School of Electrical Engineering

Mixed-language synthesis for Indian languages with dual acoustic models and probabilistic cross-lingual speaker adaptation

R. Karhila[1], D. Gowda[1], M. Gibson[2], O. Watts[3], A. Suni[4], M. Kurimo[1]
[1] Aalto University, [2] Nuance Communications inc. [3] University of Edinburgh, [4] University of Helsinki

The basic task of the annual Blizzard challenge is to take the released speech data, build synthetic voices, and synthesize a prescribed set of test sentences. The 2014 challenge was to build synthetic voices based on six Indian languages. An additional "spoke" task was to synthesise bilingual sentences, with English words or phrases embedded into the Indian language utterance.

Here we show the spoke task systems we submitted for Hindi, Rajasthani and Telugu. The synthesis is based on statistical modelling of vocoder parameters in a Hidden Markov Model (HMM) framework. HMM-based parametric speech synthesis still suffers from some voice quality issues, but generally produces very smooth prosody and allows easy manipulation of speaking style and speaker characteristics.

1. Natural speech samples: Some hundreds of utterances from each speaker are used as training material for synthetic voices.

Author of the danger trail, Philip Steels, etc. ▶ Indian-accented English speaker

इसे कई बार, मंचित भी किया गया है ▶ Hindi speaker

पण आज नुंवै रो हाको घणखरो ▶ Rajasthani speaker

అప్పుల ఊబిలో కూరుకుపోయిన రైతన్నలకు పెట్టుబడులే కనాకష్టం. ▶ Telugu speaker

2. Synthetic speech samples: Samples of synthetic speech generated from parametric models trained on the speech data of (1).

Quince seed gum is the main ingredient in wave-setting lotions. ▶ Indian-accented English speaker

लेकिन, डॉक्टर की भंगिमा मज़ाक की कतई नहीं थी ▶ Hindi speaker

अठै प्रस्तुत है इण परिचरचा रा कीं अंश ▶ Rajasthani speaker

మాజీ మండల టిడిపి అధ్యక్షుడు కంచర్ల హరినాయుడు అధ్యక్షతన జరిగిన సభలో, ఎమ్మెల్యే కొమ్ము మాట్లాడారు ▶ Telugu speaker

Aalto University

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

THE UNIVERSITY OF EDINBURGH

UNIVERSIDAD POLITÉCNICA DE MADRID

TECHNICAL UNIVERSITY OF CLUJ-NAPOCA ROMANIA

EMIME
Effective Multilingual Interaction in Mobile Environments