Next Generation Tools for Spoken Language Science & Technology

Johns Hopkins University

GitHub network

- 📍 Contributors
- 📍 Dependents

WFST-BASED MODELS | **END-TO-END MODELS**

Timeline:
- Siri
- amazon echo
- Google Home
- Assistant
- Whisper
- 2010 ... 2023
- KALDI
- Seq2Seq
- ESPnet
- SpeechBrain
- Transformer
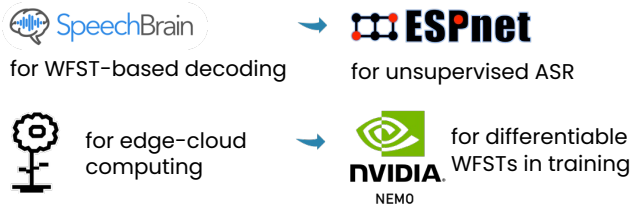- **Next-gen Kaldi**
- DNN + WFST
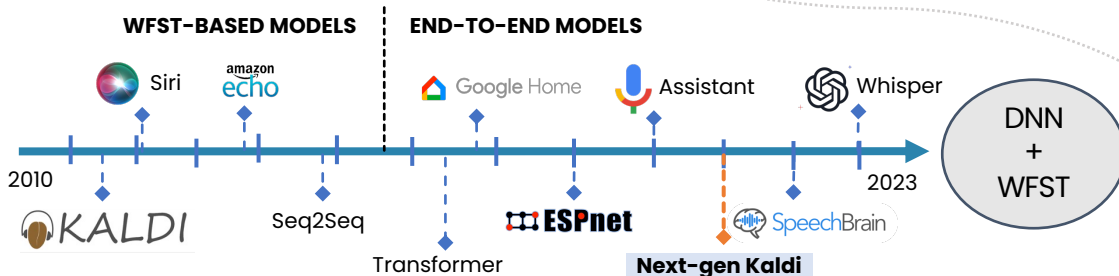
Goals of the project: (1) to make speech recognition accessible to both beginners and advanced researchers and create new training materials; (2) to leverage existing deep learning frameworks (such as PyTorch); (3) to enhance the toolkit for spoken language technology.

Contributions: We develop a suite of speech tools geared towards: (1) efficient data pipelines for audio/speech data, (2) combining neural networks with WFST-based decoding to enable modularity, and (3) training and deployment recipes. To ease adoption, we organized community tutorials and hackathons, and supported integration of our tools into other popular speech frameworks.

Community tutorials/hackathons:

- INTERSPEECH 2021 BRNO | CZECHIA — 30 August – 3 September 2021 — Speech everywhere!
- INTERSPEECH 2023 — 20th–24th August 2023 · Dublin, Ireland
- SLT 2022

Integration with other frameworks:

- → SpeechBrain — for WFST-based decoding
- → ESPnet — for unsupervised ASR
- → (edge-cloud) for edge-cloud computing
- → NVIDIA NEMO — for differentiable WFSTs in training

Enabled speech tasks:

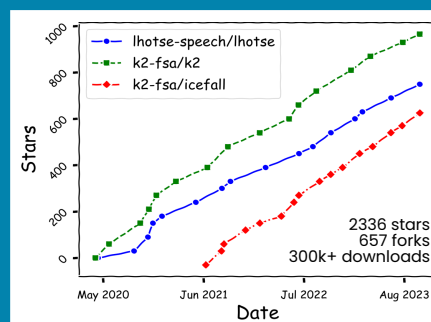| Task	Details
Speech recognition	25 recipes (with check-points) in 3 languages
Weakly supervised ASR	ASR training with non-verbatim transcripts
Multi-speaker ASR*	Including data simulation workflows
Contextual ASR*	WFST-based and deep biasing
Keyword search	Integration with Kaldi and ESPNet
Speech translation	Arabic and Mandarin dialectal ST

Only existing open-source implementation — *PhD dissertation project*

Overview of the tools

- **k2** — **k2** implements loss functions through GPU-based differentiable WFSTs.
- **LHOTSE** — **Lhotse** provides data processing pipelines and manifests for public data.
- **ICEFALL** — **Icefall** recipes combine Lhotse and k2 with PyTorch-based deep neural networks.

- lhotse-speech/lhotse
- k2-fsa/k2
- k2-fsa/icefall

2336 stars
657 forks
300k+ downloads

- → Kaldi-style reproducible recipes in PyTorch environment.
- → Model checkpoints and training logs available through HuggingFace.
- → On-device deployment support through Sherpa and ONNX.

🔥 PyTorch   🤗 Hugging Face