ALFORM: Articulatory Labeling For Optimal Real-time MRI Experiments

18-844 Final Slides

Joseph Konan

Investment Thesis

(The Elevator Pitch)

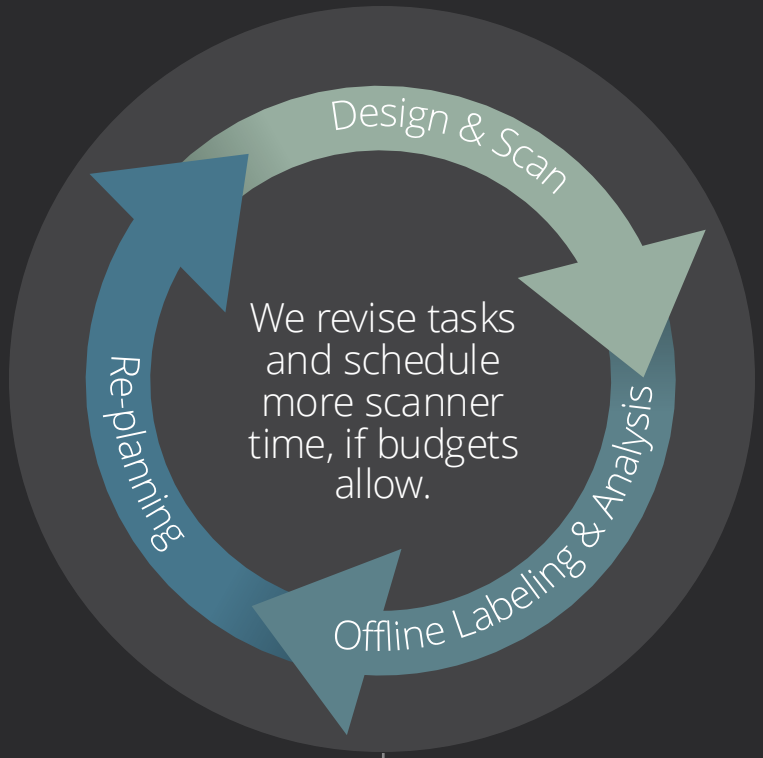
- *FOR* MRI technologists and researchers
- *WHO* need reliable articulatory data without wasting scanner time
- *THE* ALFORM system *IS A* real-time articulatory tracking tool
- *THAT* shows which articulators are being used and how often, while you scan
- *UNLIKE* expensive trial-and-error experiments that only discover missing articulators after the fact
- *OUR PRODUCT* gives instant feedback so you can focus on the right movements and collect better data, faster

PROBLEM
(Before ALFORM)

During the scan:

We guess which tasks will engage target articulators.

We collect RT-MRI data with no real-time feedback on articulator usage.



After the scan:

We manually label articulators offline.

We discover that key articulators were under-sampled or missed.

Wasted scanner time

Expensive RT-MRI sessions that don't capture what we need

Incomplete biased datasets

Important articulators (e.g., velum, tongue root) under-represented.

Slow, non-scalable science

Manual labeling + re-scans limit sample sizes & slow down progress

Here's How We Enhance Productivity

By Improving Data Collection Quality

Initiate

Researchers set target articulators & tasks.

Launch ALFORM with the RT-MRI scan.

Annotate

ALFORM tracks each frame in real time.

Summarizes usage for every articulator.

Calibrate

Identify under-sampled articulators by speaker.

Add tasks to boost underused categories.

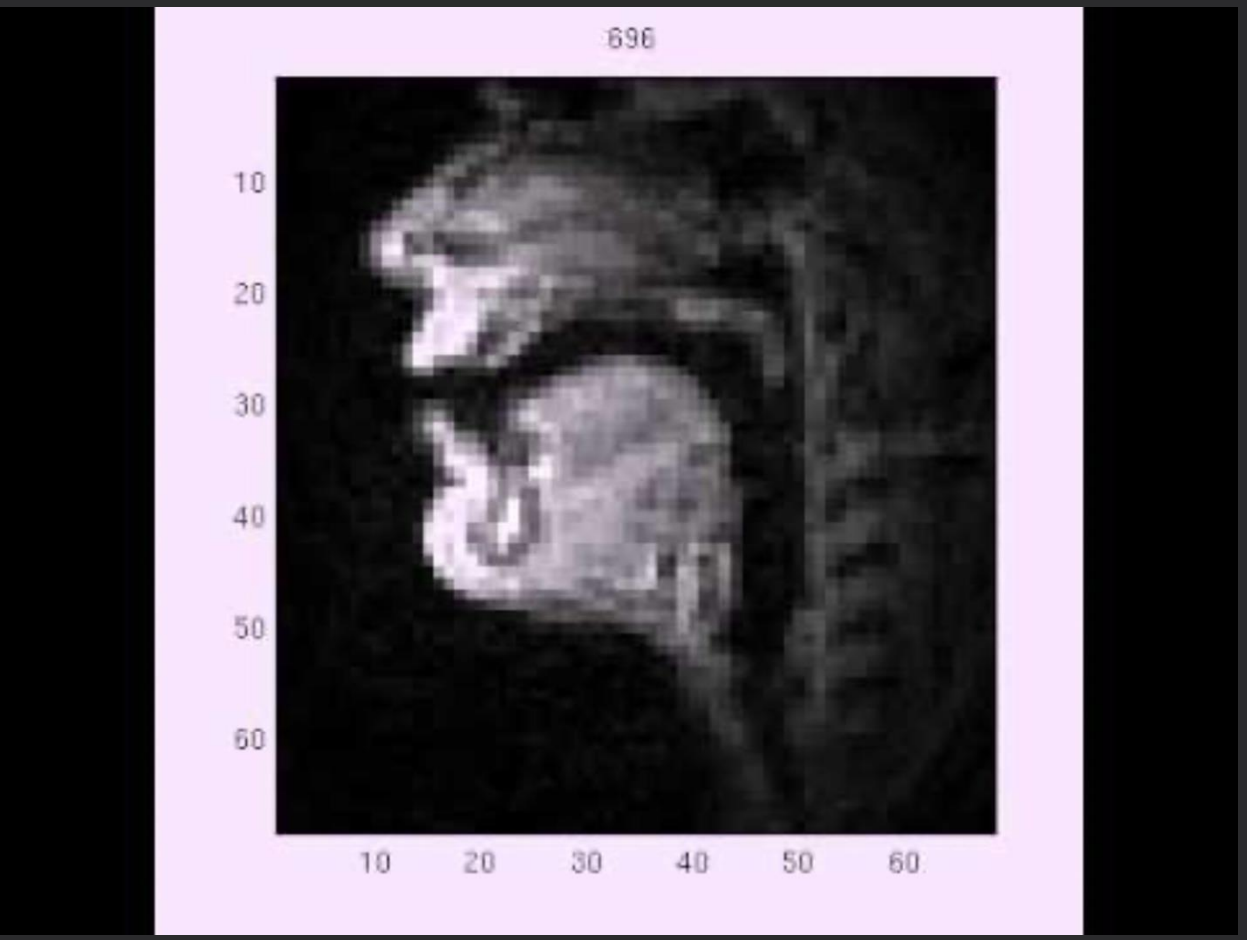
Validate

End with balanced articulatory data.

Fewer re-scans, better articulatory analysis.

Unique Real-Time Challenge

Very Challenging For Human Annotators



Articulatory Labels

About 70 to 140 per second

Articulators

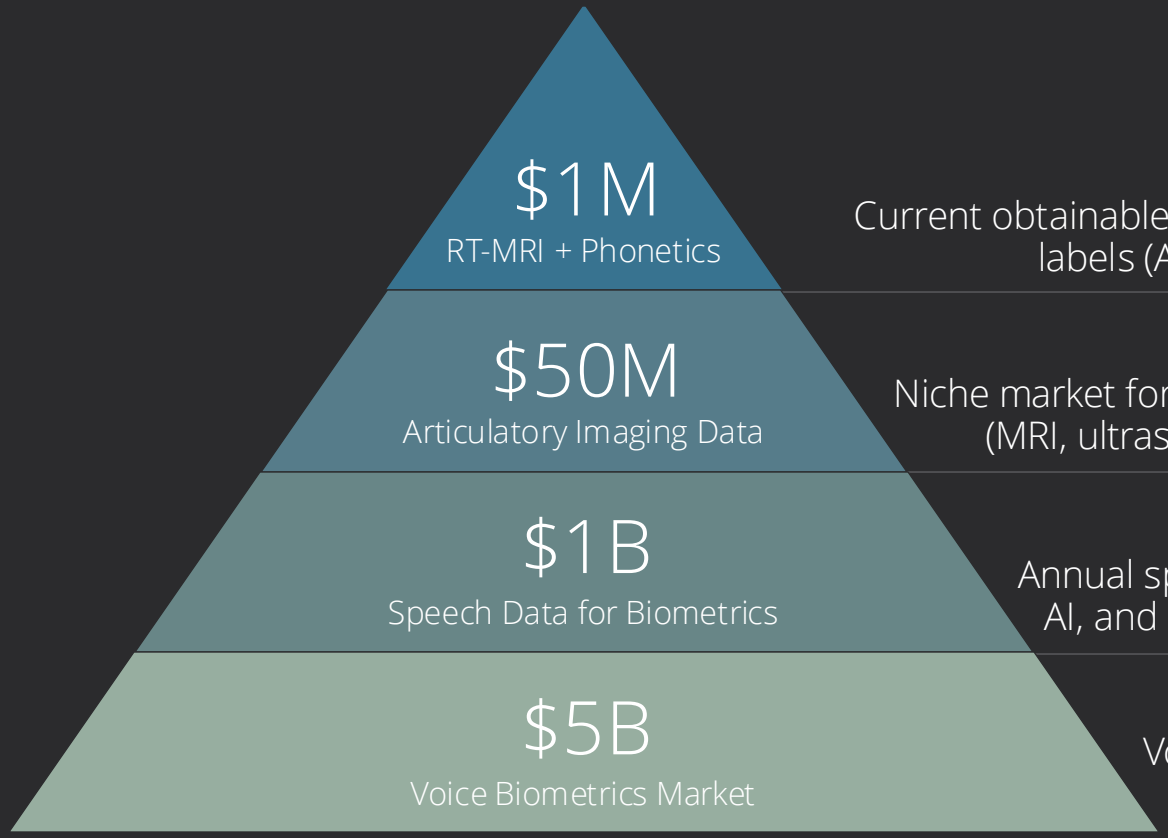
About 7 per speech sound

Speech Sounds

About 10–20 per second

A Blue Ocean Market

With Critical Need For Real-Time MRI Speech Data



Current obtainable niche for RT-MRI speech with phonetic + articulator labels (ALFORM-like data), on the order of ~$1–3M per year.

Niche market for high-fidelity articulatory speech & imaging datasets (MRI, ultrasound, EMA), roughly ~$10–50M per year worldwide.

Annual spend on speech/audio data & labeling for ASR, voice AI, and biometrics; sub-$1B now, plausibly ~$1–2B by 2030.

Voice biometrics market for authentication & speaker ID, ≈$2–5B today & growing toward ~$10B by 2030.

How Customers Use Our Product

In Medical Applications

A Jetson AGX Orin is used for development, with the compute + monitor + interface under 100W.

Stricter medical certification is required later on. A medical cart like this would be the next phase.

How Our Product Works

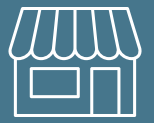
With Deep Learning Technology

Enhanced Phonetic Annotated Datasets With Articulations

- Have phoneticians expertly annotate speech recordings
- Map phonetic labels to physiological articulations
- Final annotations are speech data & articulatory label pairs

Optimized Recurrent And Convolutional Neural Architectures

- Recurrence enables efficient temporal processing
- Convolutions enable hierarchical representation
- Minimize cross entropy loss of predicted alignments

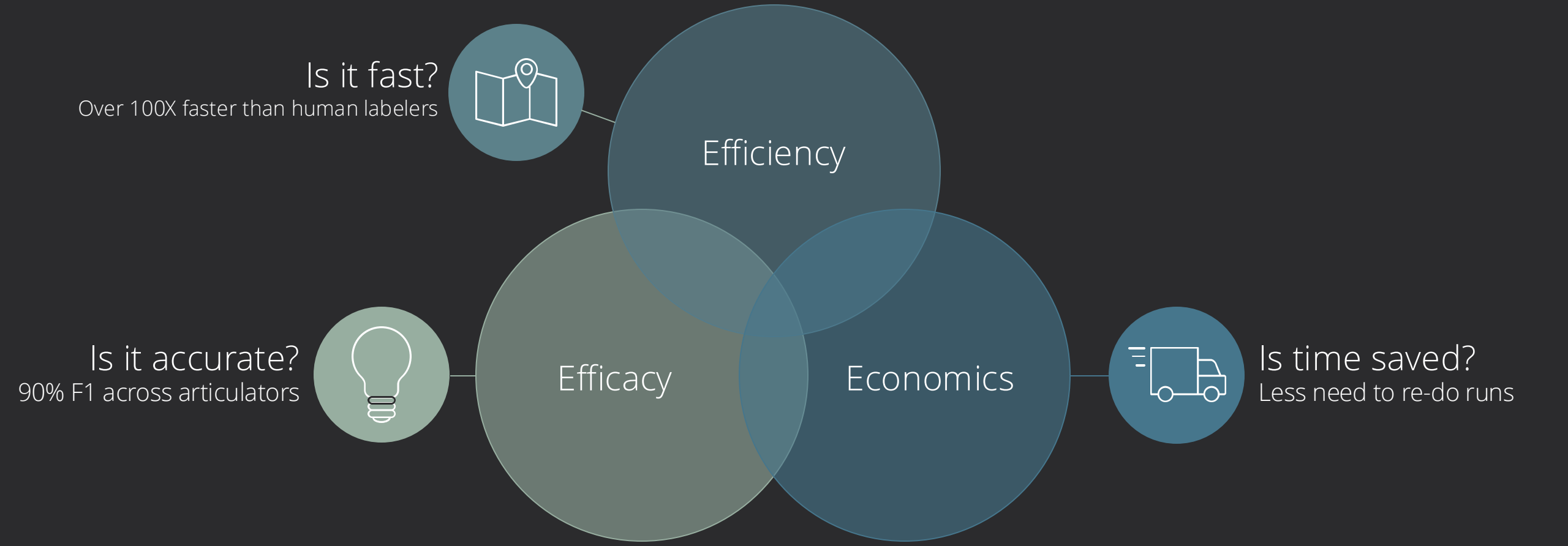


Inference Hardware Optimization Of Performance-Per-Watt

- Models are quantized from 32bit to 8bit
- Different inference hardware are experimented with
- Different wattages are experimented with

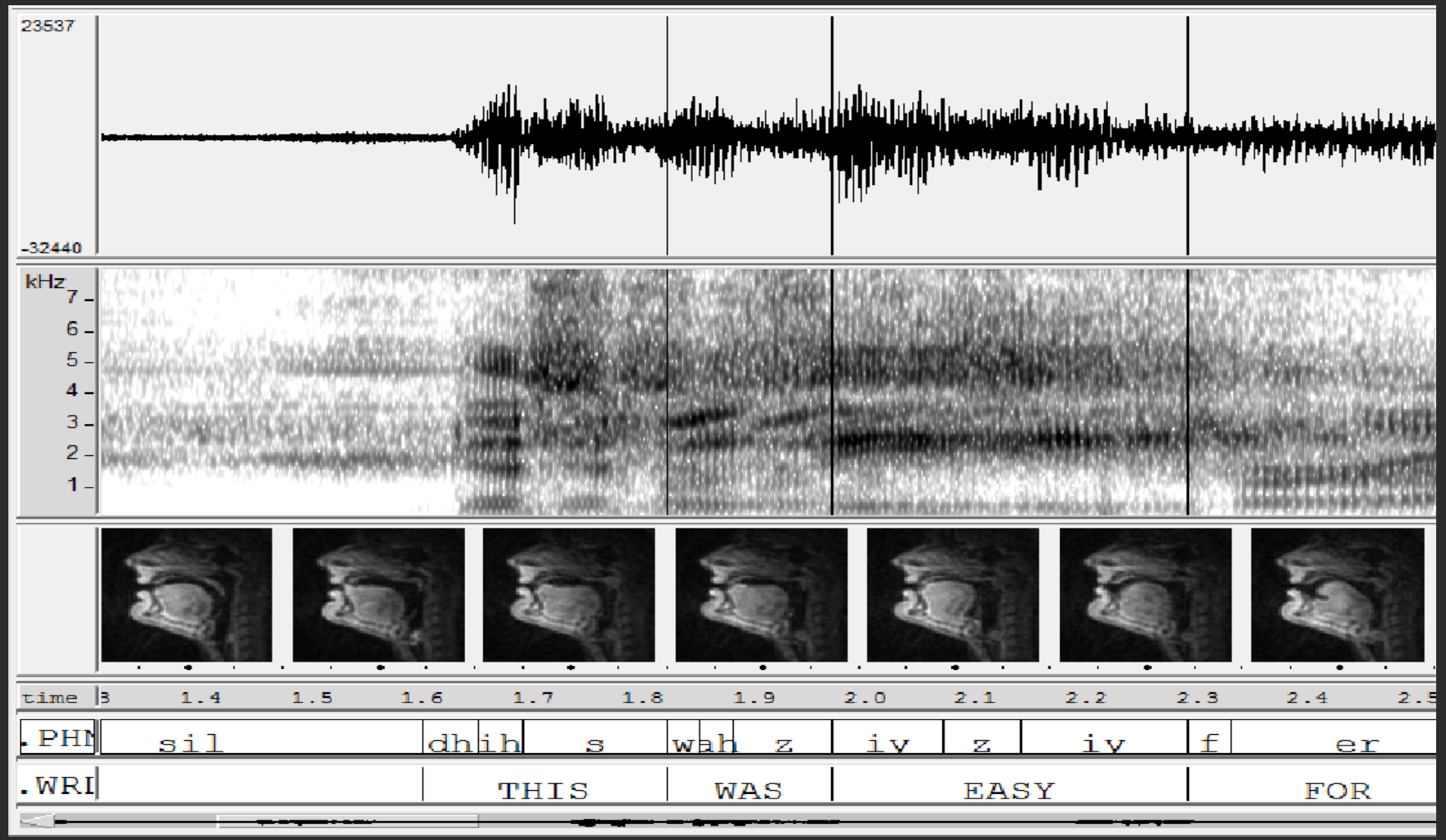
What Makes Our Product Different

Efficiency, Efficacy, and Economics



Our Primary Modality Is Speech

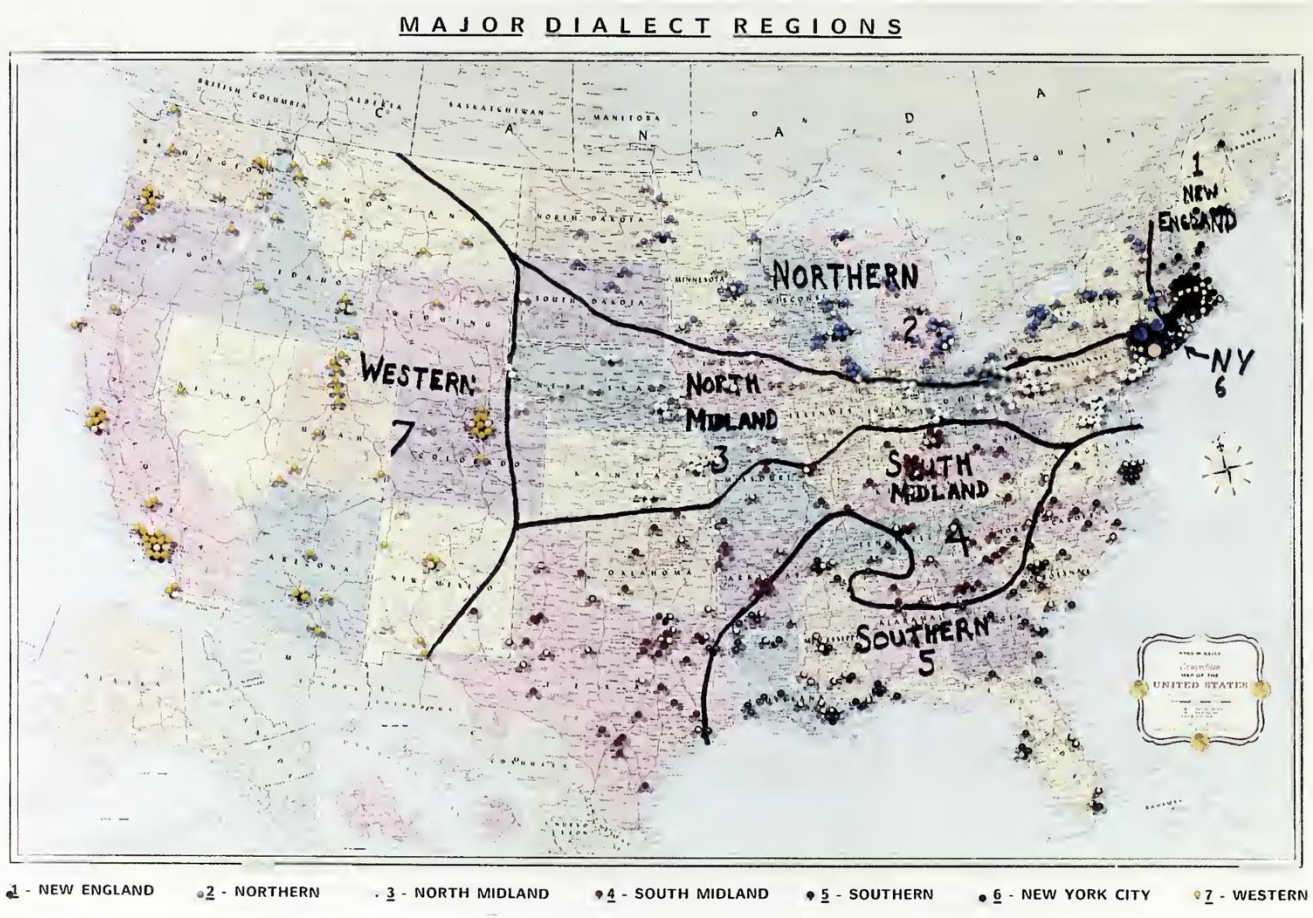
Each frame has 1 phoneme label and 7 articulatory labels



Here Are Our Demographics

Focused On 8 Dialect Regions For American English

Figure 3.1: Map of TIMIT Dialect Regions

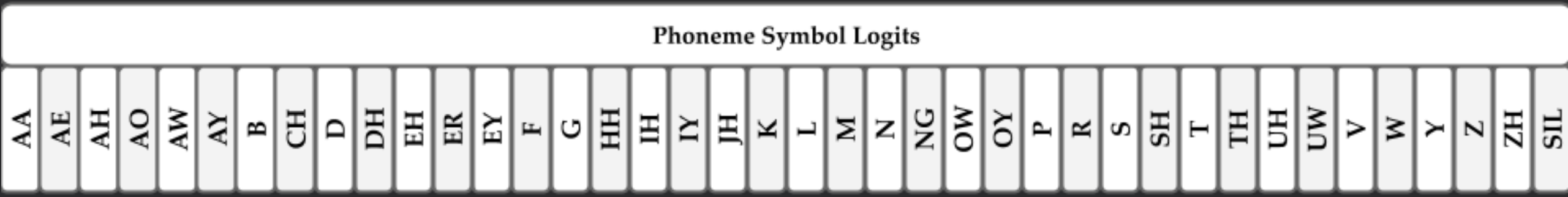


Courtesy of Texas Instruments, Inc.

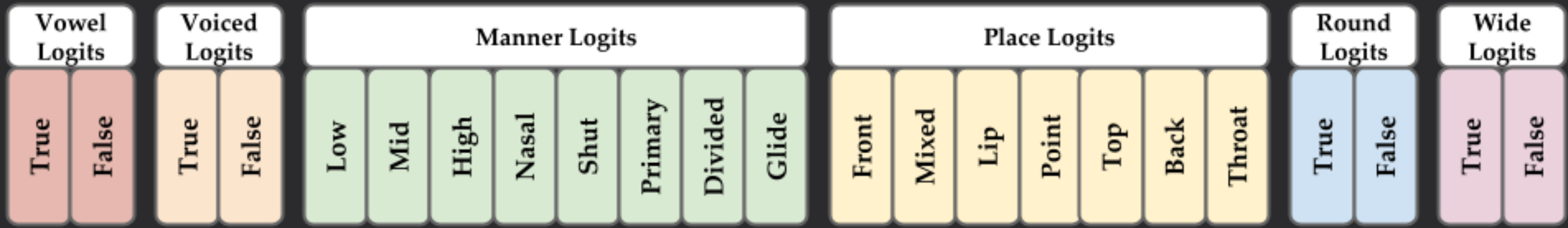
Table 3.1: Dialect distribution of speakers

Dialect Region		# Male Speakers	# Female Speakers	Total # Speakers
Name	Code (dr)			
New England	1	31 (63%)	18 (27%)	49 (8%)
Northern	2	71 (70%)	31 (30%)	102 (16%)
North Midland	3	79 (67%)	23 (23%)	102 (16%)
South Midland	4	69 (69%)	31 (31%)	100 (16%)
Southern	5	62 (63%)	36 (37%)	98 (16%)
New York City	6	30 (65%)	16 (35%)	46 (7%)
Western	7	74 (74%)	26 (26%)	100 (16%)
Army Brat (moved around)	8	22 (67%)	11 (33%)	33 (5%)
Total # Speakers:		438 (70%)	192 (30%)	630 (100%)

Tradition Phoneme Recognition Has One Classification Task



Articulation Involves Seven Separate Classifications



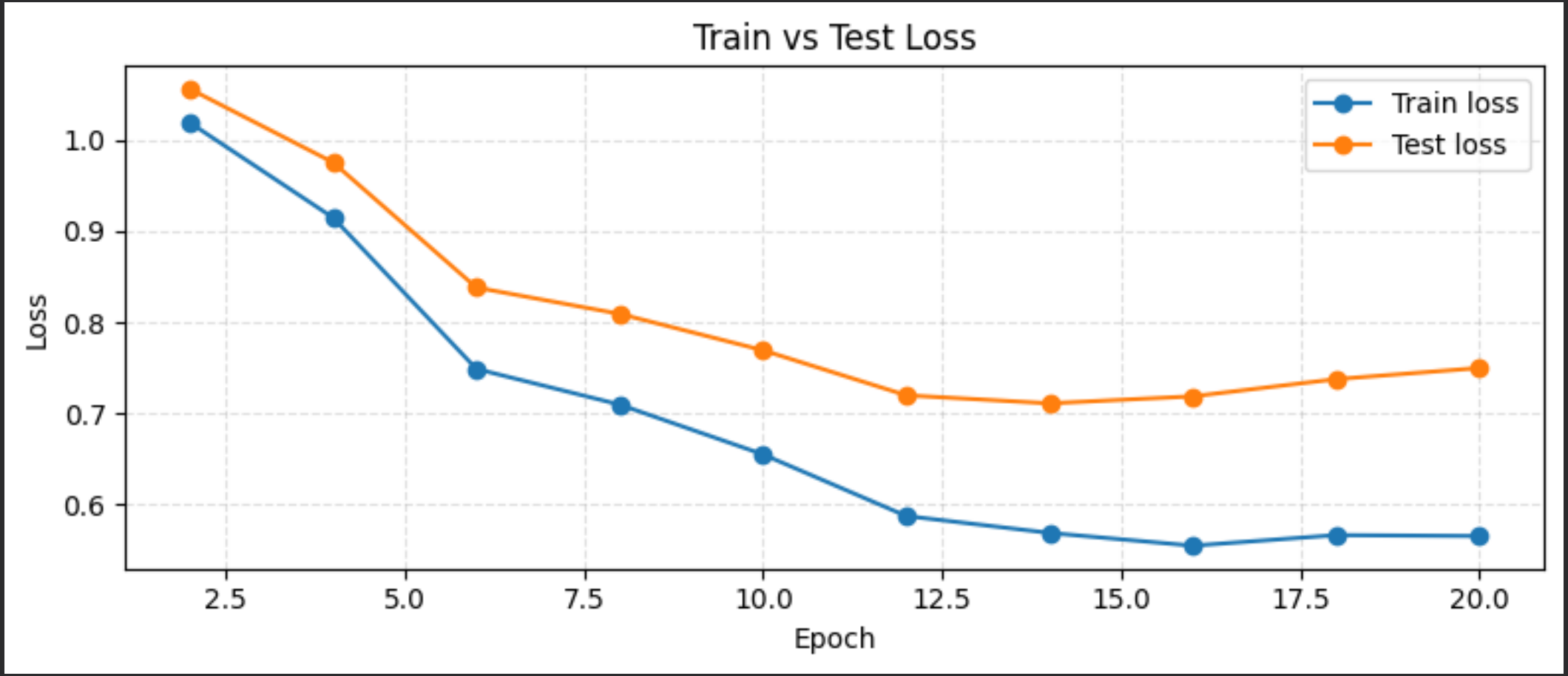
Here Are Our Metrics

On Articulator Recognition By Gender & Dialect

	Vowel	Voiced	Manner	Place	Modified	Round	Wide
FEMALE	93%	97%	87%	85%	96%	93%	87%
MALE	92%	97%	86%	85%	96%	94%	89%
DR1	92%	97%	85%	85%	95%	92%	87%
DR2	93%	97%	87%	86%	97%	95%	89%
DR3	92%	97%	87%	85%	97%	95%	89%
DR4	92%	96%	86%	84%	96%	94%	88%
DR5	92%	97%	85%	84%	96%	93%	87%
DR6	92%	97%	87%	85%	96%	94%	87%
DR7	92%	97%	86%	85%	97%	93%	89%
DR8	92%	96%	86%	85%	97%	94%	88%
ALL	92%	97%	86%	85%	96%	94%	88%

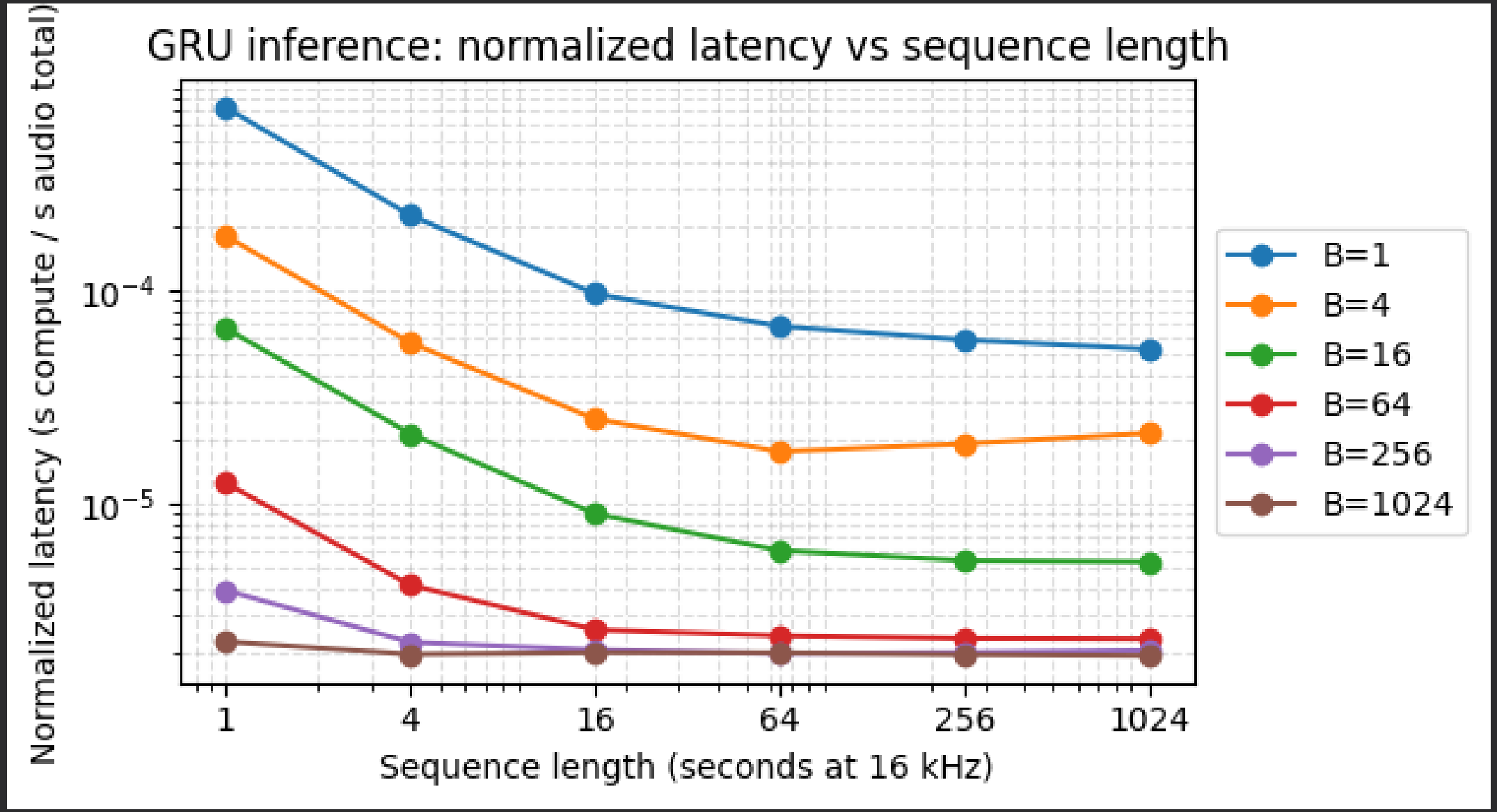
Here Are Our Metrics

On Model Training Optimization

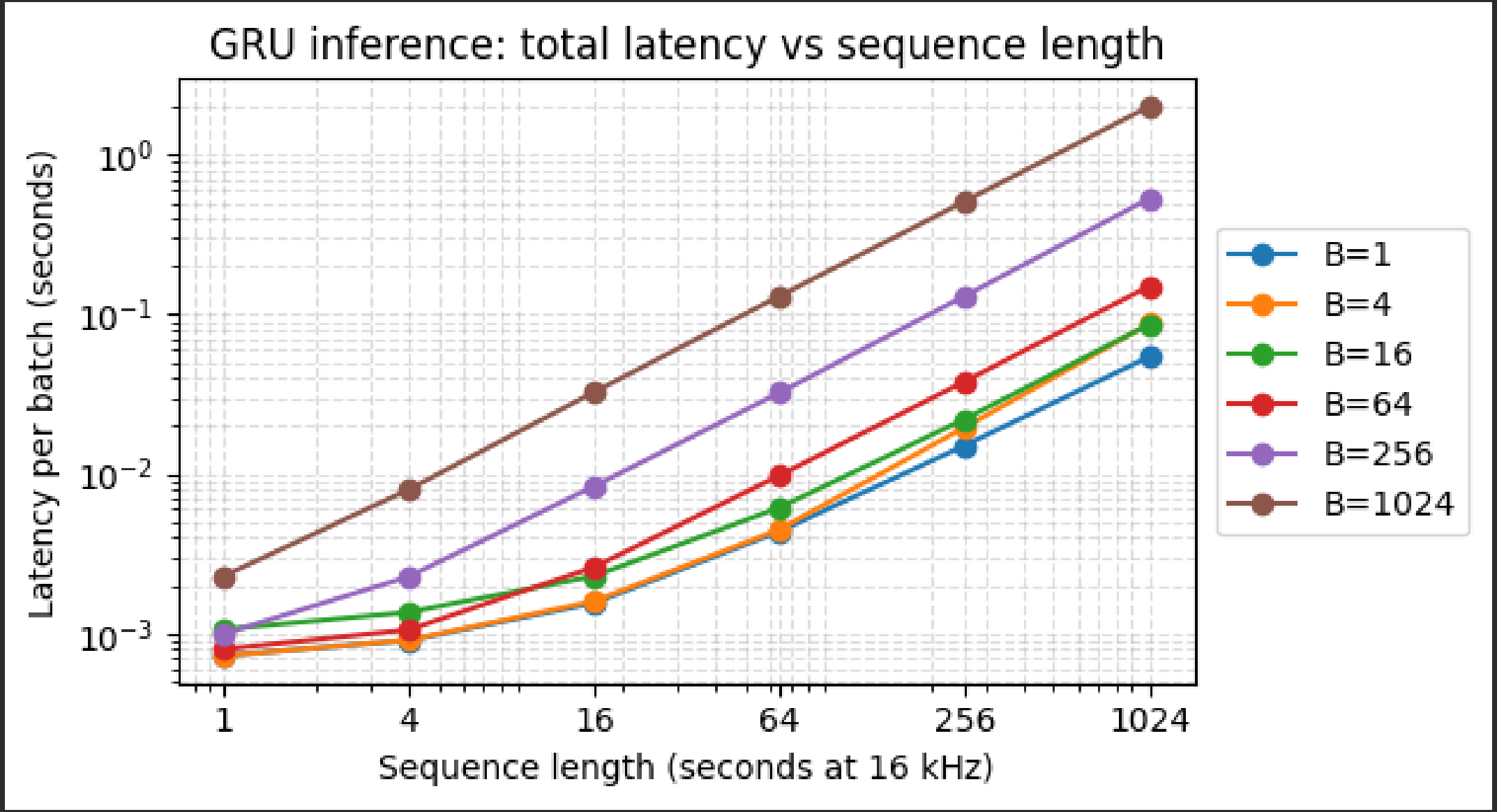


Model overfits, but shows capability of learning more-complex representations.
Future work will look into closing the gap using different loss functions as regularizers.

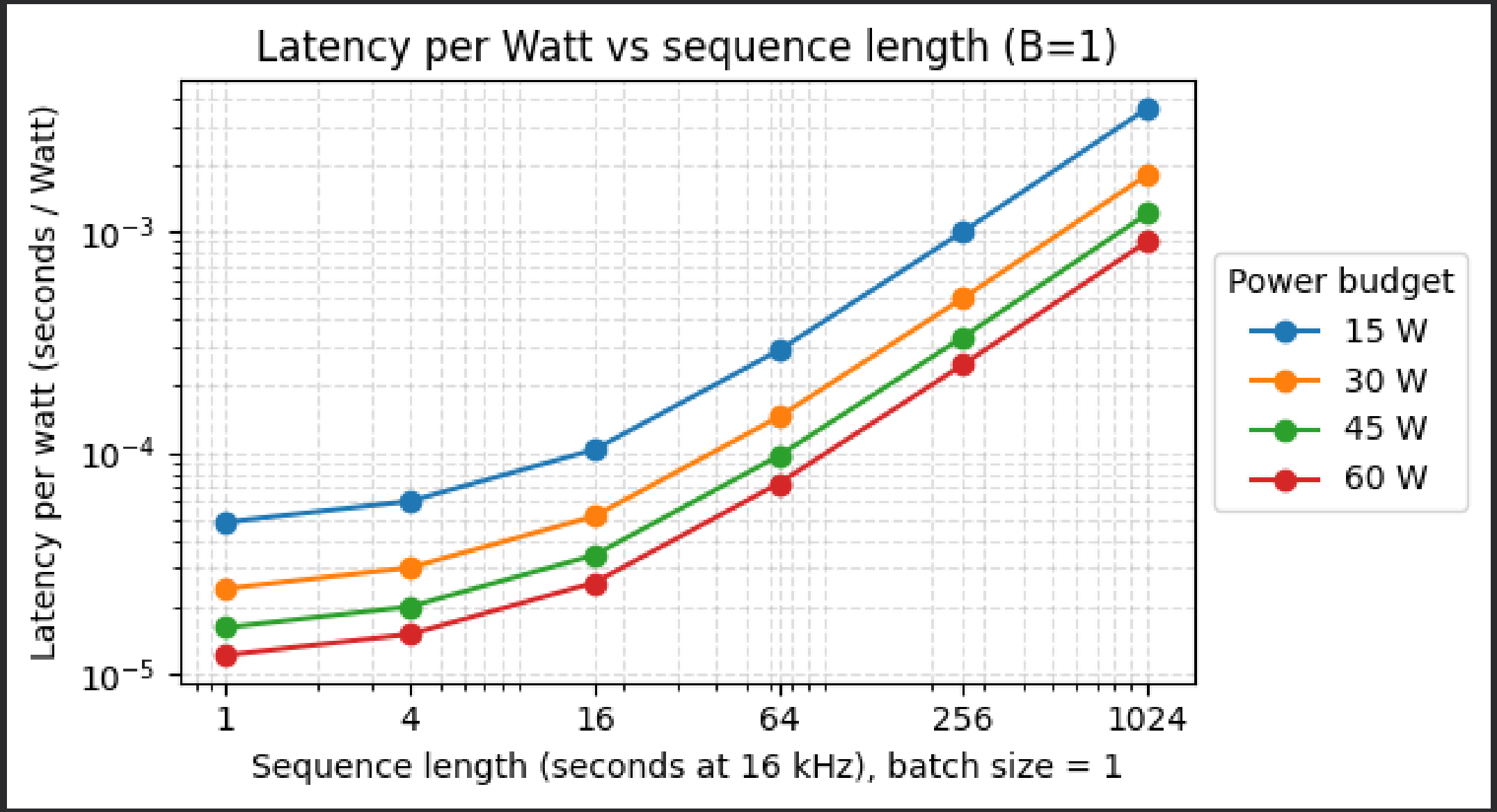
Here Are Our Metrics



Here Are Our Metrics



Here Are Our Metrics



The Next Phase Needs $10K

To Prototype On Medical-Grade Hardware (IEC 60601-certified)

01

Computer

Rugged AI computer with shielded medical I/O and ESD-safe chassis so plugging in cameras and sensors is safe for staff and patients.

$5,800

02

Monitor

High-brightness, disinfectant-safe screen that stays readable under hospital lighting and survives constant cleaning without damage.

$1,200

03

Chassis

Stable, low-tip medical cart with smooth hospital grade casters and cable routing so the whole system moves safely between patients.

$1,500

04

Battery

Fault-tolerant medical battery pack with thermal and overload protection, built to power the cart for hours without risky DIY wiring.

$1,500

Considerations TODO

Many practical challenges need to be considered.

RNNs on Edge Impulse?

RNNs aren't supported, so it needs to be hacked or CNN is needed instead.

Testing On Jetson AGX Orin

Hardware optimization needs to be re-run on the different AGX constraints.

Applied Example

An example articulatory report needs to be made for the medical technician.

Propose HW Configs

Give some options for low, medium, and high end recommended configs.

Real-Time Demonstration

Optimal hardware usage is likely not real-time, but RT would improve demo.

Training Improvement

Close the overfitting gap and consider a contrastive loss for robustness.

Investment Thesis Recap

(The Elevator Pitch)

- *FOR* MRI technologists and researchers
- *WHO* need reliable articulatory data without wasting scanner time
- *THE* ALFORM system *IS A* real-time articulatory tracking tool
- *THAT* shows which articulators are being used and how often, while you scan
- *UNLIKE* expensive trial-and-error experiments that only discover missing articulators after the fact
- *OUR PRODUCT* gives instant feedback so you can focus on the right movements and collect better data, faster

Thank You!

Acknowledgements

- MRI images are from USC 75-Speaker Speech MRI Database
- Gender and Dialect region data are from TIMIT
- Source link to the articulation video is embedded
- Financial estimates are approximate
- A language model was used to facilitate drafting