

APOORV VYAS

Mountain View, CA · +1 650 465 5675 · apoorv2904@gmail.com · [Website](#) · [Google Scholar](#) · [LinkedIn](#) · [GitHub](#)

RESEARCH FOCUS

Multimodal representation learning · Large-scale generative model training · Efficient Transformers

EXPERIENCE

Meta — FAIR (Fundamental AI Research)

Menlo Park, CA

Staff Research Scientist (2026–) · Senior Research Scientist (2024–26) · Research Scientist (2022–24)

Nov 2022 — Present

- **Multimodal representation learning (PE-AV).** Proposed and led PE-AV, a family of audio–video–text aligned encoders trained with contrastive learning scaled on caption quality, data types, loss pairs, and model size. Ran all early ablations, scaled the data pipeline to 100M videos, and directed workstreams as the effort grew from 2 to 6 researchers.
- **Multimodal post-training.** Integrated audio and speech recognition into a joint text–perception model; designed and ran the ablations behind the integration approach and drove the work to retain model quality.
- **Large-scale audio generation.** *Movie Gen Audio* (video-conditional): led scaling de-risking and the UNet-vs-Transformer architecture study; owned pre- and post-training data curation, funneling O(2B) videos to a 100M-scale training set. *Audiobox*: co-implemented the model and owned the joint speech and sound-effect training behind its unified generation; built the LLM-captioning, augmentation, and annotation quality-control pipelines; implemented voice prompting. *Voicebox*: implemented the Transformer backbone and duration model; ran multilingual experiments.

Idiap Research Institute / EPFL

Switzerland

Doctoral Researcher — Thesis: *Efficient Transformer-Based Speech Recognition*

Jul 2018 — Aug 2022

- **Compute efficiency.** Developed linear attention, a kernelized formulation expressing autoregressive Transformers as RNNs at linear cost, and clustered attention, which approximates self-attention via input clustering.
- **Data efficiency.** Uncertainty-aware semi-supervised learning via dropout. Cross-domain and cross-lingual analysis of SSL pretraining.

Intel Labs

India

Research / Systems Engineer

Apr 2015 — May 2018

- Ensemble method for out-of-distribution detection in deep networks.
- Low-power supervised semantic hashing for approximate image retrieval.

SELECTED PUBLICATIONS

Vyas, A., Chang, H., Yang, C., Huang, P., Gao, L., Richter, J., Chen, S., Le, M., Dollár, P., Feichtenhofer, C., Lee, A., Hsu, W. **Pushing the Frontier of Audiovisual Perception with Large-Scale Multimodal Correspondence Learning (PE-AV).** *CVPR*, 2026.

Shi, B., Tjandra, A., Hoffman, J., Wang, H., Wu, Y., Vyas, A., Feichtenhofer, C., Dollár, P., Hsu, W., Lee, A. **SAM Audio: Segment Anything in Audio.** *ICML*, 2026.

Meta Movie Gen Team. **Movie Gen: A Cast of Media Foundation Models.** *arXiv:2410.13720*, 2024.

Pratap, V., Tjandra, A., Shi, B., Tomasello, P., Babu, A., Vyas, A., Baevski, A., Adi, Y., Hsu, W., Conneau, A., Auli, M. **Scaling Speech Technology to 1,000+ Languages (MMS).** *JMLR*, 2024.

Vyas, A., Shi, B., Le, M., Tjandra, A., Wu, Y., Guo, B., et al. **Audiobox: Unified Audio Generation with Natural Language Prompts.** *arXiv:2312.15821*, 2023.

Le, M., Vyas, A., Shi, B., Karrer, B., Sari, L., Manohar, V., Adi, Y., Mahadeokar, J., Hsu, W. **Voicebox: Text-Guided Multilingual Universal Speech Generation at Scale.** *NeurIPS*, 2023.

Vyas, A., Katharopoulos, A., Fleuret, F. **Fast Transformers with Clustered Attention.** *NeurIPS*, 2020.

Katharopoulos, A., Vyas, A., Pappas, N., Fleuret, F. **Transformers are RNNs: Fast Autoregressive Transformers with Linear Attention.** *ICML*, 2020.

Vyas, A., Jammalamadaka, N., Zhu, X., Das, D., Kaul, B., Willke, T. **Out-of-Distribution Detection Using an Ensemble of Self Supervised Leave-out Classifiers.** *ECCV*, 2018.

EDUCATION

École Polytechnique Fédérale de Lausanne (EPFL)

2018–2022

PhD, Electrical Engineering

Indian Institute of Technology, Guwahati

2010–2014

B.Tech, Electronics and Electrical Engineering

PATENTS & SERVICE

Patents 9 granted US patents (5 at Intel, 4 at Meta)

Reviewing NeurIPS, ICLR, ICML, CVPR, ECCV, AAAI, TNNLS