Arun Baby | Speech Researcher

🏠 www.arunbaby.com ● www.linkedin.com/in/arunbaby0

About

I am passionate about research, technologies, have a broad engineering background, and love developing myself, the people around me, and leading teams.

Professional Interests: Machine Learning, Deep Learning, Conversational AI, Speech Synthesis, Automatic Speech Recognition, Text to Speech Systems, Speech tech for Indian Languages, Data Science.

Professional Experience - 12+ years

Samsung Research(SRI-B) - Staff Engineer(Senior Chief Engineer)

On-Device ASR(Acqui-hired from Zapr Media Labs)   **Apr 2022 - present**
Speech Research Team(ASR)   *Samsung Research*

- Galaxy AI team: leading the tech development of on-device ASR models for S24 devices
- ASR model compression.
- On-device model optimisations.
- Streaming and hybrid ASR models.
- On-device TTS research and development
- Research submissions in ICASSP, Interspeech and NCC
- Environment: Tensorflow, Python, Bash.

Zapr Media Labs(acquired by Samsung) - Research Scientist(Speech)

Conversational AI   **Jan 2020 - Apr 2022**
Speech Research Team   *Zapr Media Labs*

- VoiceBots - Speech based bot for automatic customer assistance.
- Achieved around **80% call success** for feedback and ratings voicebots with preliminary deployments.
- VoiceAnalytics - Customer care data analytics for better user engagement.
- Achieved **10-30% relative WER improvements** compared to commercial ASR systems.
- Speech research training and deployment framework architect.
- Created a **scalable kubeflow framework for parallel training/inference** for 100s of models.
- Context-based ASR systems for Indian Languages: **50% NER improvement.**
- Controllable expressive text-to-speech: speed and pitch controllablity.
- **Bilingual models** for Indian Languages.
- Indian English optimizations for TTS and ASR models.
- Automated Phonetic rules for Indian English models.
- A **20% relative improvement in MOS scores** for Indian English and Hindi TTS models.
- NLP for grapheme to phonemes.
- **3 speech research article** submissions in international conferences.
- 1 accepted **patent.**
- Mentored 2 juniors.
- Environment: ESPnet, Kaldi, PyTorch, ParallelWaveGAN, Huggingface, VOSK, IndicNLP,
- FastAPI, docker, kubeflow, kubernets, gRPC.
- AWS, GCP.

Cisco Systems - Software Engineer II

Data Center debug analytics for routing products   **Aug 2018 - Jan 2020**
Enterprise Networking   *Cisco Systems*

- Lossless logging, improvements in debuggability, and analytics for debugging.
- Up to **80% memory footprint reduction** for debuger logs.
- Code optimisations for **10% faster code execution**.
- IPv4, IPv6 code maintenance.
- Environment: C, Linux, Python.

IIT Madras - Project Officer(Research)

Development of text-to-speech synthesis systems for Indian languages (Research) Jan 2015 - Jun 2018
MeitY, Govt. of India *IIT Madras*

- Development of a **language independent parser** across 13 Indian languages.
- **Segmentation of phonemes** using signal processing cues and machine learning (deep learning) techniques.
- **Development of voice models** for 13 Indian languages using HTS (a parametric speech synthesis toolkit) and Merlin (a neural network based speech synthesis toolkit).
- Integration of the unified parser to the IndicTTS application in **Android platform**.
- **Knowledge transfer** to Samsung Research Institute - Noida, IndusOS and ShinanoKenshi.
- **Lead and managed** IITM TTS team along with **collaborating with 11 organisations** across the country.
- Environment: C, Python, Perl, Shell, HTK, Kaldi, HTS, Merlin (TTS) and Android.

Ministry of Electronics and
Information Technology
Government of India

HCL Technologies - Software Engineer

Web and Mobile Applications - Development and Maintenance Jan 2013 - Jul 2014
KnowledgePoint360 Group *HCL Technologies*

KnowledgePoint360 Group LLC provides multi-channel health-care communication and advisory services for pharmaceutical industries, industries related to biotechnology and health-care professionals.

- **Lead the PoC development** of web based app for improving the efficiency of physicians in USA.
- Lead major back-end development/maintenance of the product development for a team of 4.
- Development of a **multi-platform** and **cross browser web application** from scratch for managing physicians and speakers (medical representatives).
- Development of an **interface to verify** speaker's and physician's data by Verisys with **automatic syncing** to the web application.
- **Maintenance** of core applications, new portal, mobile applications and services across different projects.
- **Performance optimization** of front-end and back-end services.
- Management of **project kick-off** activities.
- Environment: Visual Studio, C#, ASP.net, MVC, HTML5, JavaScript, jQuery, CSS and SQL Server.

Database Query Optimization for Statistical Data Jun 2012 - Dec 2012
Thomson Reuters *HCL Technologies*

Thomson Financial, a popular financial management company, handles stock market data.

- **Optimization of SQL jobs** to perform statistical data analysis.
- Development of a **multiple job scheduler** to improve the efficiency of the system.
- Environment: SQL Server and SSIS.

Delphi to .Net Migration Oct 2011 - Jun 2012
Triple Point Technology *HCL Technologies*

The company provides real-time, innovative solutions to the complex issues that arise in the management of the commodity supply chain.

- Development of a **PoC application** for multi-market commodity and enterprise risk management.
- Environment: .NET framework 3.5, C#, WPF and XML.

Education

M.S. (By Research) 2015 - 2018
Computer Science and Engineering *Indian Institute of Technology Madras*

- Supervisor: Prof. Hema A Murthy
- Thesis: A Unified Approach to Building Speech Synthesis Systems in Indian Languages
- CGPA: 7.8/10
- Key courses: Pattern Recognition, Kernel Methods for Pattern Analysis, Speech Technology, Linear Algebra and Random Processes, Advanced Data Structures and Algorithms

B.Tech   **2007 - 2011**
Computer Science and Engineering   *Rajagiri School of Engineering and Technology*

- Thesis: Parallel C - A Serial to Parallel C Code Converter
- Percentage: 75/100

Publications: Google Scholar link

- **Arun Baby**, George Joseph, Shatrughan Singh **Robust Speaker Personalisation Using Generalized Low-Rank Adaptation for Automatic Speech Recognition**, ICASSP 2024, April 2024.
- Dhanya Eledath, **Arun Baby**, Shatrughan Singh **Robust Speech Recognition Using Meta-Learning for Low-Resource Accents**, National Conference on Communications(NCC), February 2024.
- **Arun Baby**, Pranav Jawale, Saranya Vinnaitherthan, Sumukh Badam, Nagaraj Adiga, Sharath Adavanne **Non-native English lexicon creation for bilingual speech synthesis**, Speech Synthesis Workshop(SSW), August 2021.
- **Arun Baby**, Jeena J Prakash, Aswin Shanmugam S and Hema A Murthy, **Significance of Spectral Cues in Automatic Speech Segmentation for Indian Language Speech Synthesizers**, Speech Communication, Oct 2020.
- **Arun Baby**, Saranya Vinnaitherthan, Nagaraj Adiga, Pranav Jawale, Sumukh Badam, Sharath Adavanne, Srikanth Konjeti **An ASR Guided Speech Intelligibility Measure for TTS Model Selection**, arXiv pre-print, May 2020.
- **Arun Baby**, **A Unified Approach to Building Speech Synthesis Systems in Indian Languages**, MS Thesis, IIT Madras, 2019.
- Anju Leela Thomas, Anusha Prakash, **Arun Baby**, and Hema A Murthy, **Code-switching in Indic Speech Synthesisers**, INTERSPEECH 2018, 1948-1952, Hyderabad, India.
- **Arun Baby**, Karthik Pandia DS and Hema A Murthy, **Signal Processing Cues to Improve Automatic Speech Recognition for Low Resource Indian Languages**, SLTU 2018, 25-29, Gurugram, India.
- **Arun Baby**, Jeena Prakash, Rupak Vignesh and Hema A Murthy, **Deep Learning Techniques in Tandem with Signal Processing Cues for Phonetic Segmentation for Text to Speech Synthesis in Indian Languages**, INTERSPEECH 2017, 3817-3821, Stockholm, Sweden.
- Atish Ghone, Rachana Nerpagar, Pranaw Kumar, **Arun Baby**, Aswin Shanmugam, Sasikumar Mukundan and Hema A Murthy, **TBT (Toolkit to Build TTS): A High Performance Framework to build Multiple Language HTS Voice**, INTERSPEECH 2017 (Show and Tell), 3427-3428, Stockholm, Sweden.
- **Arun Baby**, Anju Thomas, Jeena Prakash, Anusha Prakash and Hema A Murthy, **Speech Synthesis in Indian Languages and Future Perspectives**, Global Conference on Cyberspace (GCCS), November 2017.
- **Arun Baby**, Jeena Prakash and Hema A Murthy, **A Hybrid Approach to Neural Networks based Speech Segmentation**, Frontiers of Research in Speech and Music (FRSM), December 2017.
- **Arun Baby**, Nishanthi N L, Anju Leela Thomas, and Hema A Murthy, **A Unified Parser for Developing Indian Language Text to Speech Synthesizers**, International Conference on Text, Speech, and Dialogue, pp. 514-521, Brno, Czech Republic, September 2016.
- **Arun Baby**, Anju Leela Thomas, Nishanthi N L, and TTS Consortium, **Resources for Indian Languages**, CBBLR Workshop, International Conference on Text, Speech, and Dialogue, pp. 37–43, Brno, Czech Republic, September 2016.
- Anusha Prakash, **Arun Baby**, Aswin Shanmugam S, Jeena J Prakash, Nishanthi N L, Raghava Krishnan K, Rupak Vignesh Swaminathan and Hema A Murthy, **Blizzard Challenge 2015 : Submission by DONLab, IIT Madras**, Blizzard Challenge, Berlin, Germany, September 2015.

Academic Projects

Signal Processing and Machine Learning Projects   **July 2015 - May 2016**
Pattern Recognition, Kernel Methods for Pattern Analysis, Speech Technology   *IIT Madras*

- Classification techniques: Bayes classifier, GMM, HMM, DNN, CNN and SVM.
- Clustering techniques: k-means and Support Vector Clustering.
- Regression techniques: simple linear regression, polynomial regression and Support Vector Regression.
- Learning methods: supervised, semi-supervised and unsupervised.

- Applications: Speaker Recognition using signal processing cues and acoustic modeling.
  - Environment: Matlab, C++, Perl, Shell and Torch.

Course Work

Key Courses   **January 2015 - May 2016**
Core and electives   *IIT Madras*

  - Pattern Recognition   - Data Structures and Algorithms
  - Kernel Methods for Pattern Analysis   - Linear Algebra and Random Processes
  - Speech Technology

Skills and Tools

  - **Languages** - C, C++, C#, Perl, Shell and Python
  - **Operating Systems** - Linux and Windows
  - **Applications and Tools** - Visual Studio, HTK, Festival, Kaldi, Merlin (TTS), ESPnet MATLAB and Pytorch
  - **Database** - MYSQL and SQL Server
  - **Web development** - HTML5, PHP, XML, CSS, JavaScript, jQuery and Flask

Positions of Responsibility

  - Coordinator for speech research publishing at Zapr Media Labs.
  - **Point of Contact** for lossless logging for Netstack team.
  - **Point of Contact** for industrial collaboration for TTS resources at IIT Madras.
  - **Technology Expert/Point of Contact** for ASP.NET MVC 4 and jQuery across the projects at HCL Technologies.
  - **Point of Contact** of speaker verification project (Web and Mobile Application Development and Maintenance, KnowledgePoint360 Group, HCL Technologies).

Co - Curricular Activities

  - Member of International Speech Communication Association (**ISCA**).
  - Reviewer for **ICON** 2018, 2019.
  - Presented 2 papers in **INTERSPEECH** 2017.
  - Presented 2 papers in TSD 2016.
  - Attended Winter School on Speech and Audio Processing (**WiSSAP**), 2016.
  - Finalist in **Bootcamp 17** - a startup bootcamp event at IIT Madras.
  - Participated in Microsoft **code.fun.do** in 2016 and 2017.
  - Won awards for various Inter-school and college **maths quizzes**.
  - Won awards for **coding and debugging** competitions at various **technical fests**.

Extra - Curricular Activities

  - Participated in the 8th **edition of Terry Fox Run Chennai 2016.**
  - Member of **Disaster Management Team**, IIT Madras 2016-17.
  - Member of **IIT Madras Volleyball** team.
  - Member of Volleyball, Table Tennis, Tennis, Basketball, Badminton and Squash **Hostel** teams at IIT Madras.
  - **Hapkido** practitioner since 2016 (3rd Kup).