



Aakash Sasikumar
MSc Computer Science



Bokang Jia
MSc Machine Learning



Gokul Karthik Kumar
MSc Computer Vision



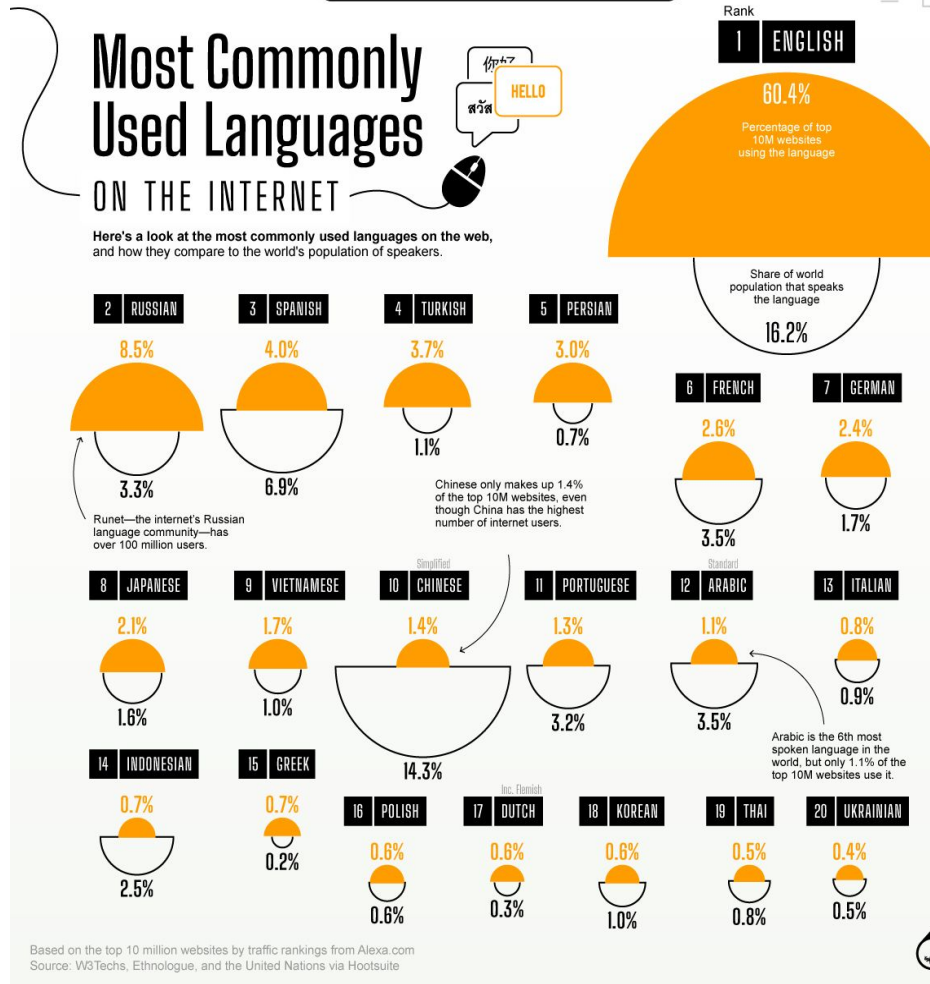
Mentor
Yifan Peng
PhD Electrical and Computer Engineering



Motivation

- Most video content are in english.
- Human dubbing for localization is expensive
- Existing end-to-end AI video dubbing has **accuracy, synchronization and style** issues.

- **Autodub goal:** Create a world free of language barriers.
- **Impact:** Education / Culture exchange / Media representation.



Autodub is a Human-in-the-loop AI dubbing platform that integrates Transcription, Translation, Voice-over, and Background Audio Separation.

✓ At the moment, **Autodub** supports Arabic, Chinese, English, and major Indian languages (Tamil, Telugu, Kannada, Malayalam, and Hindi)



Illustration source: DALL-E 2

How *Autodub* Works?

1. **Import** your Video into the Autodub editor.
2. **Transcribe** to generate subtitles using our AI instantly.
3. Select a target language and **Translate** the subtitles.
4. **Voice-over** to replace the source speech with speed-adjusted AI speech in the target language. Autodub also extracts and **mixes the background music** from the source video.
5. **Interact** with the editor using Autodub's visual guidance.
6. **Export** the dubbed video.



Autodub Editor

The screenshot displays the Autodub Editor interface. At the top left is the Autodub logo, and at the top right is a 'Home' button with a user profile 'GK'. The main workspace is divided into several sections:

- Video Preview:** Shows a video of the Burj Khalifa. The project name 'abudhabi-bg' is visible in the background.
- English Subtitles:** A list of English subtitles with a vertical scrollbar. The visible text includes: 'the nation's capital.', 'And it serves as the administrative centre for the entire country', 'It is usually overshadowed by its biggest sister city Dubai', 'and if you are interested to go to Dubai,', 'check out my video guide here.', 'But what is not as big as Dubai or', and 'as famous as Dubai'. A timecode '00:00:07 / 00:01:15' is shown at the bottom of this section.
- Arabic Subtitles:** A list of Arabic subtitles with a vertical scrollbar. The visible text includes: 'عاصمة الأمة', 'وهي بمثابة المركز الإداري للبلد بأكمله', 'وعادة ما تطغى عليها أكبر مدينة شقيقة دبي', '، وإذا كنت مهتمًا بالذهاب إلى دبي', 'تتحقق من دليل الفيديو الخاص بي هنا', 'لكن ما ليس بحجم دبي أو', and 'مشهورة مثل دبي'. Each subtitle has a circular refresh icon to its right.
- Audio and Export Controls:** Includes buttons for 'Import Video', 'Import Subtitle', 'English' (dropdown), 'Transcribe', 'Arabic' (dropdown), 'Translate', 'Male1' (dropdown), 'Voice-over', 'Clear', 'Undo', 'Export Video', 'Export Subtitle', 'Space: Play / Pause', and 'Ctrl + Z: Undo'. It also features an 'Audio Speed Scale' with a slider from 'Very Slow' to 'Very Fast' and options for 'Lengthen text', 'Shorten text', 'Shorten time', and 'Lengthen time'.
- Timeline:** A horizontal timeline at the bottom shows time markers from 00:00:03 to 00:00:12. Below the timeline is an audio waveform with three segments labeled with timecodes: 4.26, 3.01, and 2.59.



Autodub's Unique Features

- Model agnostic.
 - Switch cutting edge models.
 - Extensible to new languages.
- Practical Human-in-the-loop UI design.
 - Error cascading (ASR -> MT -> TTS)
 - Synchronization without training time-constrained MT.
 - Improvement though user interaction.
- Background audio extraction.
 - Nonvocal audio applied to TTS outputs
 - Retain original quality.
- Speaker profile matching. (Under development)
 - Language-agnostic speaker embeddings to match speaker profiles.



Technologies used by *Autodub*



React

NEXT.js



Hugging Face



Firebase



pyannotate



steroid



FastAPI



SpeechBrain



Google Cloud Platform



OpenAI



AI4BHARAT

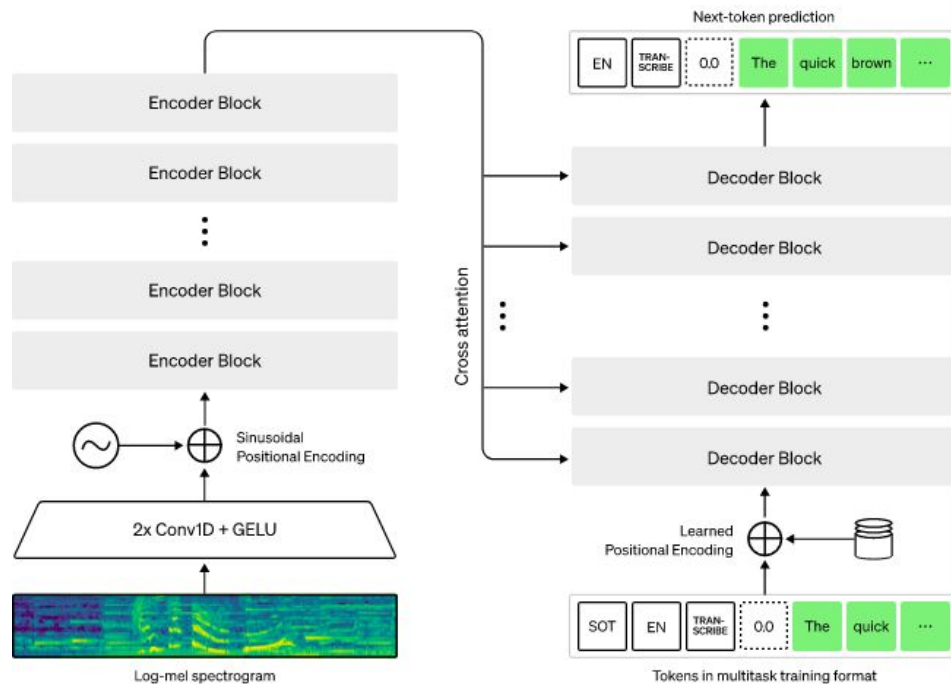


Autodub's Transcription API

- OpenAI's Whisper ASR for English
- GCP's ASR for all languages

Under Development:

- Integration of AI4Bharat's IndicASR for Indian languages
- Fine-tuning of Whisper ASR for Arabic and Indian languages



Source: <https://openai.com/blog/whisper/>



Autodub's Translation API

- Meta's NLLB MT for English to X
- GCP's MT for all combinations

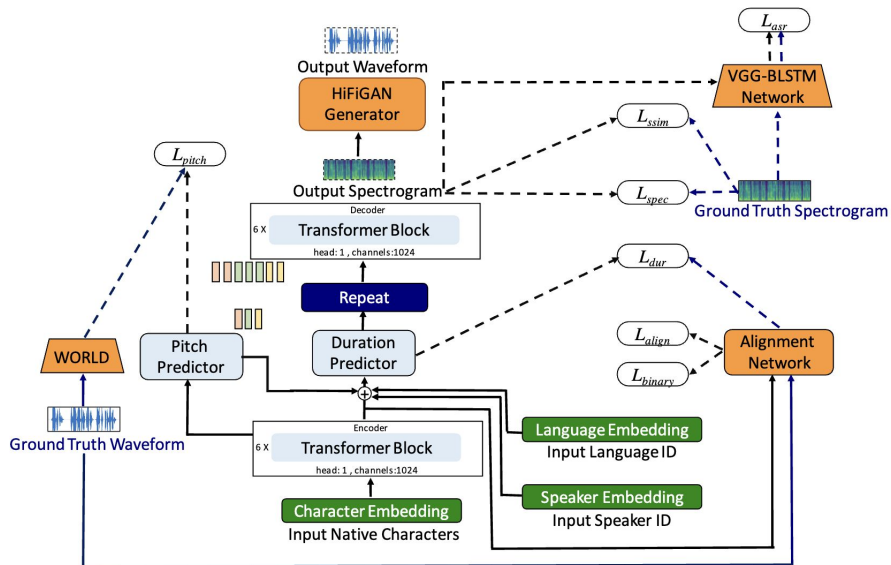
Under Development:

- Integration of AI4Bharat's IndicTrans for Indian languages



Autodub's Voice-Over API

- AI4Bharat's TTS for Indian languages (FastPitch + HiFiGAN)
 - Co-led by Gokul at Microsoft
- GCP's TTS for all languages



Towards Building Text-To-Speech Systems for the Next Billion Users

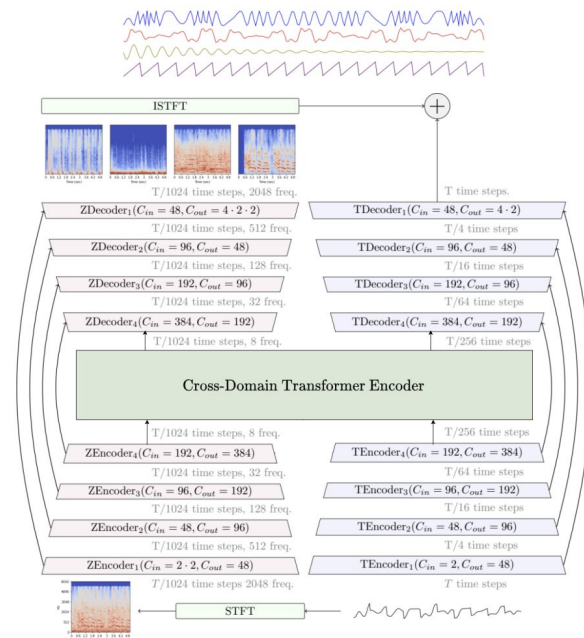
Gokul Karthik Kumar*, Praveen S V*, Pratyush Kumar,
Mitesh M. Khapra, Karthik Nandakumar

2022 | ArXiv - Under Review at ICASSP 2023



Autodub's Background Audio separation API

- All non-speech audio.
- HT Demucs:
 - Waveform separation model.
 - Trained on MUSDB music instrument separation dataset.
 - Adapter for background audio separation.
 - Empirically effective



Source: Rouard, Simon, Francisco Massa, and Alexandre Défossez. "Hybrid Transformers for Music Source Separation." *arXiv preprint arXiv:2211.08553* (2022).



Autodub's Timeline

Pre-Hackathon

2 months

- AI4Bharat's Text-To-Speech
- Background Audio Extraction

Hackathon - Online

2 months

- Evaluation of open-source AI tools
- Editor Frontend
- Editor Backend
- User Authentication

Hackathon - Onsite

2 days

- API development
- Integration of Frontend, Backend and APIs





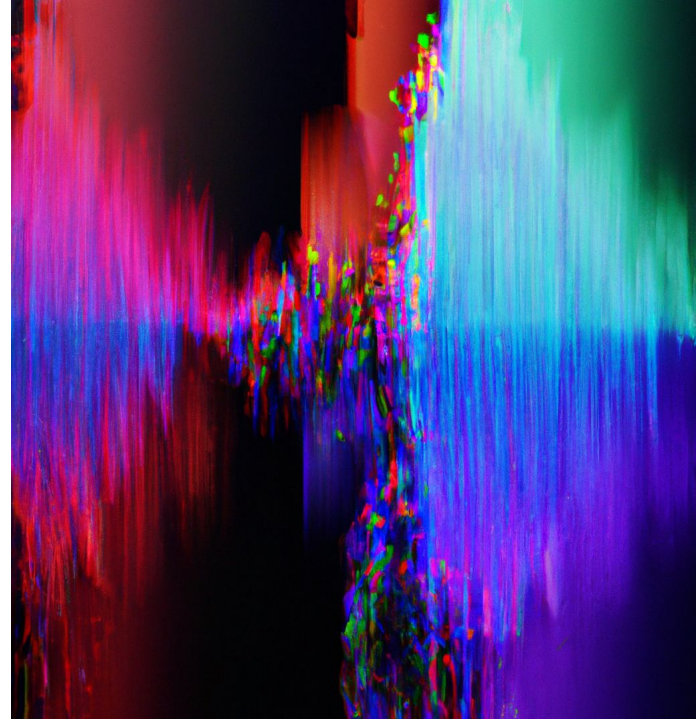
In Summary

- **Autodub**
 - Cross-lingual model-agnostic video **dubbing platform** with **human-in-the-loop** feedback.
- **Benefits:**
 - Critical affordability and scalability vs **expert dubbing**.
 - Highly **customizable** output with human interaction.
 - Improved **synchronization** and **style** matching.
 - Natural sounding video with **background audio retention**
- **Impact:**
 - Increasing **education accessibility** for non-English speakers.
 - Increase **reach of content creators** to new audience.
 - Unlock **economic opportunities** for human voice artists.

Future Goals + Takeaways

- A human-dubbing marketplace with cross-lingual voice artist matching.
- Multivoice support with speaker diarization
- Video lip-syncing
- Additional underrepresented languages

Autodub is a unique dubbing platform that bring us one step closer to a world in which human knowledge is accessible to everyone.



It's Time For ***Autodub*** Demo!

Editor: <https://autodub.io>

APIs: <http://app.autodub.io:8000/docs>

Video Demo: <https://www.autodub.io/demo>

