

アクセント付きの音声をアクセントのない音声に変換する

日本の空港における航空コミュニケーションの強化

Authors
Li Tang & Julián Villegas Ph.D.
{d8261109, julian}@u-aizu.ac.jp

This research aims to improve aviation communication by addressing challenges specific to non-native English speakers.

1. Developing an **accent neutralization system** for Japanese-accented English inspired by L2-ARCTIC and reference-free conversion methods [1,2].
2. Creating an **ASR system** for Japanese airport communication using a dataset constructed from LiveATC recordings.

1. **Data Collection:** Recordings are gathered from LiveATC channels of Japanese airports, focusing on major hubs like Narita (RJAA), Haneda (RJTT), and Kansai (RJBB).
2. **Preprocessing:** Audio is segmented into individual utterances, with noise reduction and speaker diarization applied. Approximately 5–10 hours are manually transcribed for validation.
3. **Automated Transcription:** Pre-trained ASR systems (e.g., Whisper, Kaldi) transcribe the remaining data, fine-tuned using the manually transcribed subset [3].
4. **Metadata Annotation:** Transcriptions are enriched with speaker roles, airport codes, and contextual information [3].
5. **ASR Model Training:** The model is pre-trained on general datasets (e.g., LibriSpeech), then fine-tuned on the Japanese airport data with a domain-specific language model for improved aviation accuracy.
6. **Evaluation:** Performance is tested on unseen recordings using Word Error Rate (WER), Sentence Error Rate (SER), and aviation-specific metrics.

This study aims to develop a reference-free Japanese Accent Conversion (JAC) system tailored to Japanese-accented English. Inspired by the L2-ARCTIC [1] methodology and the paper “*Converting Foreign Accent Speech Without a Reference*,” [2] the process involves constructing a Japanese-accented English speech corpus modeled after L2-ARCTIC. The steps include:

- Recruit native Japanese speakers to record phonetically balanced prompts derived from the CMU ARCTIC dataset.
- Record speech in a controlled environment to ensure high-quality audio.
- Annotate phoneme boundaries and common pronunciation errors manually for a subset of sentences, targeting phonetic challenges unique to Japanese speakers (e.g., /l/ vs. /r/, vowel epenthesis).

- Train an L2 speech synthesizer using speech embeddings extracted from Japanese-accented utterances, mapping them to their Mel-spectrogram representations.
- Generate golden-speaker utterances by replacing L2 embeddings with embeddings from native English utterances.
- Train a pronunciation-correction model using Japanese-accented utterances as input and golden-speaker utterances as targets.

This research is sponsored by **JAXA Aeronautical Innovation Challenge FY2024 Feasibility Study** (JAXA 航空イノベーションチャレンジ FY2024 フィージビリティスタディ).

[3] Zuluaga-Gomez, J., Veselý, K., Szöke, I., Blatt, A., Motlícek, P., Kocour, M., ... & Klakow, D. (2022). Atco2 corpus: A large-scale dataset for research on automatic speech recognition and natural language understanding of air traffic control communications. *arXiv preprint arXiv:2211.04054*.

