

# Rancang Bangun Sistem Automatic Speech Recognition untuk Bahasa Indonesia Berbasis Wav2Vec 2.0 = Development of Automatic Speech Recognition System for Indonesian Language Based on Wav2Vec 2.0

Mohammad Salman Alfarisi, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20523943&lokasi=lokal>

---

## Abstrak

<p>Salah satu permasalahan yang terdapat pada sistem Automatic Speech Recognition (ASR) yang sudah ada adalah kurangnya transparansi dalam penanganan data suara, yang tentunya membuat adanya keraguan terhadap privasi data tersebut. Di sisi lainnya, untuk mengembangkan sebuah sistem ASR yang memiliki akurasi memadai dan dapat bekerja secara luring membutuhkan jumlah data yang banyak, khususnya data suara yang sudah diiringi dengan transkripnya. Hal ini menjadi salah satu hambatan utama pengembangan sistem pengenalan suara, terutama pada yang memiliki sumber daya minim seperti Bahasa Indonesia. Oleh karena itu, dalam penelitian ini dilakukan perancangan sistem pengenalan suara otomatis berbasis model wav2vec 2.0, sebuah model kecerdasan buatan yang dapat mengenal sinyal suara dan mengubahnya menjadi teks dengan akurasi yang baik, meskipun hanya dilatih data dengan label yang berjumlah sedikit. Dari pengujian yang dilakukan dengan dataset Common Voice 8.0, model wav2vec 2.0 menghasilkan WER sebesar 25,96%, dua kali lebih baik dibandingkan dengan model Bidirectional LSTM biasa yang menghasilkan 50% namun membutuhkan jumlah data dengan label 5 kali lipat lebih banyak dalam proses pelatihan. Namun, model wav2vec membutuhkan sumber daya komputasi menggunakan 2 kali lebih banyak RAM dan 10 kali lebih banyak memori dibandingkan model LSTM</p><hr /><p>One of the main problems that have plagued ready-to-use Automatic Speech Recognition (ASR) Systems is that there is less transparency in handling the user's voice data, that has raised concerns regarding the privacy of said data. On the other hand, developing an ASR system from scratch with good accuracy and can work offline requires a large amount of data, more specifically labeled voice data that has been transcribed. This becomes one of the main obstacles in speech recognition system development, especially in low-resourced languages where there is minimal data, such as Bahasa Indonesia. Based on that fact, this research conducts development of an automatic speech recognition system that is based on wav2vec 2.0, an Artificial Model that is known to recognize speech signals and convert it to text with great accuracy, even though it has only been trained with small amounts of labeled data. From the testing that was done using the Common Voice 8.0 dataset, the wav2vec 2.0 model produced a WER of 25,96%, which is twice as low as a traditional Bidirectional LSTM model that gave 50% WER, but required 5 times more labeled data in the training process. However, the wav2vec model requires more computational resource, which are 2 times more RAM and 10 times more storage than the LSTM model.</p>