



INTISARI

Penelitian ini berfokus pada perancangan dan pengembangan aplikasi berbasis web yang memfasilitasi pengguna dalam merekam dan mengelola korpus ucapan untuk sistem *Text-to-Speech* (TTS) yang mengubah teks tertulis menjadi ucapan yang terdengar alami. Korpus ucapan adalah komponen dataset pelatihan yang kritis untuk mengembangkan model TTS berkualitas tinggi. Aplikasi ini bertujuan untuk menyederhanakan proses pembentukan korpus ucapan yang secara tradisional manual dan kompleks sehingga dapat diakses oleh khalayak yang lebih luas.

Pengembangan aplikasi mengikuti pendekatan *Feature-driven Development* (FDD) dengan penyampaian fitur yang bertahap. Aplikasi ini mencakup alur kerja otomatis untuk pembentukan korpus ucapan, yang dibagi menjadi tiga sesi utama: praperekaman untuk manajemen set ucapan, perekaman untuk perekaman sampel suara, dan pascaperekaman untuk mengeksport data dengan fleksibel ke dalam paket korpus ucapan standar. Pengujian fungsional dilakukan menggunakan metode pengujian *black box* yang berhasil mengonfirmasi bahwa aplikasi memenuhi semua persyaratan fungsional yang ditentukan.

Untuk memvalidasi efektivitas korpus ucapan yang dihasilkan, kerangka kerja TTS berbasis VITS Piper digunakan untuk melatih dua model TTS, satu menggunakan korpus ucapan bahasa Indonesia dan lainnya menggunakan korpus ucapan bahasa Inggris yang dihasilkan oleh aplikasi. Model-model ini kemudian dievaluasi dalam aspek kealamian dan kualitas keseluruhan dengan hasil mencapai *Mean Opinion Score* (MOS) di atas 3,75 poin yang menyimpulkan hasil model yang berkualitas tinggi dan terdengar alami.

Sebagai kesimpulan, aplikasi yang dikembangkan berhasil menghasilkan korpus ucapan yang memenuhi standar dan cocok untuk melatih model TTS personal dengan kualitas suara yang baik. Alat ini secara signifikan menyederhanakan proses pembuatan korpus ucapan, mendorong partisipasi yang lebih luas dan berpotensi meningkatkan keragaman dan kualitas sistem TTS. Penelitian ini menunjukkan bahwa aplikasi dapat mendukung pengembangan model TTS personal yang dapat menghasilkan ucapan sintesis berkualitas tinggi dan lebih mudah diakses.

Kata kunci : interaksi berbasis web, pengenalan ucapan, sumber daya bahasa, pengembangan perangkat lunak lincah



ABSTRACT

This research focuses on designing and developing a web-based application that facilitates users in recording and organizing speech corpus for Text-to-Speech (TTS) systems which convert written text into natural-sounding speech. A speech corpus is a critical training dataset component for developing high-quality TTS models. The application aims to simplify the traditionally manual and complex process of speech corpus formation, making it accessible to a broader audience.

The development followed the Feature-driven Development (FDD) approach with incremental feature delivery. The application includes an automated workflow for speech corpus formation, divided into three main sessions: pre-recording for utterance set management, recording for streamlining the voice sample recording process, and post-recording to flexibly export the data into a standard speech corpus package. Functionality testing was conducted using black box testing methods confirming that the application met all specified functional requirements.

To validate the effectiveness of the generated speech corpus, the VITS-based Piper TTS framework was used to train two TTS models, one using Indonesian speech corpora and the other using English speech corpora produced by the application. These models were then evaluated for both naturalness and overall quality, achieving a Mean Opinion Score (MOS) above 3.75 points, indicating high-quality, natural-sounding speech.

In conclusion, the developed application successfully produces speech corpora that meet industry standards, suitable for training personal TTS models with good sound quality. This tool significantly simplifies the process of creating speech corpora, encouraging broader participation and potentially enhancing the diversity and quality of TTS systems. The research demonstrates that the application can effectively support the development of personalized TTS models, making high-quality synthetic speech more accessible.

Keywords : web-based interaction, speech recognition, language resources, agile software development