

Sistem text-to-speech dengan metode unit selection synthesis untuk Bahasa Indonesia = Text-to-speech system with unit selection synthesis method for Bahasa Indonesia

Bayu G. Wundari, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20248992&lokasi=lokal>

Abstrak

Skripsi ini membahas tentang sistem Text-to-Speech (TTS) untuk Bahasa Indonesia dengan Unit Selection Synthesis sebagai metodenya untuk mensintesa ucapan. Unit yang digunakan pada sistem TTS ini berupa suku kata Bahasa Indonesia. Sistem TTS yang dibuat pada skripsi ini memiliki 2 modul utama, yaitu modul Natural Language Processing (NLP) dan modul Digital Signal Processing (DSP). Modul NLP bertugas untuk memroses input teks yang masuk guna mendapatkan informasi dari teks itu berupa unit suku kata dengan pitch dan ToBI (Tone and Break Indices) yang bersesuaian dengan kalimat pada teks masukan, Informasi ini kemudian digunakan oleh modul DSP untuk menghasilkan ucapan. Pada modul DSP ini, metode sintesa ucapan yang digunakan adalah Unit Selection Synthesis yang merupakan generasi ketiga setelah Concatenative Synthesis. Metode Unit Selection Synthesis menggunakan database yang sangat banyak sekali untuk dapat menghasilkan ucapan dengan tingkat kealamian yang tinggi. Untuk tiap unit suku kata memiliki karakteristik seperti pitch, durasi, Mel Frequency Cepstrum Coefficient (MFCC), dan ToBI yang berbeda-beda dengan unit yang lain walaupun suku kata yang digunakan adalah sama. Suku kata dengan karakteristik yang berbeda tersebut diperoleh dari hasil pemotongan file wav suatu rekaman ucapan. Dari segi intelligibilitas, ucapan yang dihasilkan tidaklah baik. Hal ini disebabkan database yang dimiliki sangat kurang dan rekaman ucapan yang dijadikan sumber data memiliki banyak noise sehingga mengganggu proses pemotongan file wav untuk mendapatkan suku kata. Namun tingkat kealamian ucapan yang diperoleh dari sistem TTS ini dapat dikatakan cukup baik karena pitch dari suku kata yang cukup bervariasi sehingga intonasi yang terdengar tidak mendatar saja.

This undergraduate thesis discusses about a Text-to-Speech system with Unit Selection Synthesis as it's method to synthesize speech. Units which are used as the units for the synthesizer are Bahasa Indonesia syllables. In this study, the TTS system uses 2 main modules, they are Natural Language Processing module (NLP) and Digital Signal Processing Module (DSP). The NLP module processes input text for retrieving information from the input in the form of syllables with their pitch and ToBI (Tone and Break Indices) associated with the sentences in the text. The retrieved information then used by DSP module to produce speech. The third generation synthesizer after concatenative synthesis, Unit Selection Synthesis, is chosen as the speech synthesizer in the DSP module. To get speech with high naturalness, the synthesizer must uses a large speech database. Each and every syllable has it's own characteristics such as pitch, duration, Mel Frequency Cepstrum Coefficient (MFCC), and ToBI that are different from other units eventhough the syllables are the same. The author get the syllables by trimming a wav file of recorded speech. From the intelligibility point of view, the quality of the produced speech is not good. It is because the quality of the possessed database is poor and the recorded speech chockablock with noise in such a way that unsettles the process of trimming the wav file in order to get the syllables. Yet, from the naturalness point of view, the quality of the speech could be accepted because of the variety of the pitch of the syllables so that the perceived speech is not monotone.