
ABSTRACT

Currently, the development of speech processing rising rapidly. From the voice results obtained does not sound natural at all, then headed toward an increasingly good prosody. Speech synthesizer is on the last block of the Text-to-Speech system. Speech synthesizer is a system that able to produce artificial human voice with speech synthesis. There are some method of speech synthesizer, that is formant synthesis, articulatory synthesis, and concatenative synthesis.

The method used in this final assignment is diphone concatenation method. Speech synthesis was initially formed by the voice recording, then the results are stored in database. Then the voice recording parsed into some of the diphone, that is transition between two adjacent sounds (adjacent phones), that will be more stable when coupled with other diphone. Frequency Domain Pitch Synchronous Overlap-Add (FD PSOLA) algorithm used to assemble the speech diphone unit. By using FD PSOLA algorithm, the process of arranging the inter-containing diphone transition between two adjacent sounds (adjacent phones), became smooth without any noise.

The result that can be obtained from combining diphone concatenation method with FD PSOLA algorithm is sound of speech synthesis more natural and can be clearly understood. In addition, the FD PSOLA algorithm provide an easy way in modifying pitch. The better the quality of diphone units contained in the database will facilitate the formation of speech synthesis and will improve the quality of synthesis results. Based on Mean Opinion score (MOS), the intelligibility, fluidity, and naturalness parameter index has reach 3,24; 3,04 and 3,25. So that the system is fair enough in synthesizing human speech.

Keywords : speech synthesizer, diphone concatenation, pitch, FD PSOLA algorithm.