

ABSTRACT

Currently, the development of speech synthesizer is increasing rapidly. From the speech results obtained that does not sound natural to hear at all, and then go to the better prosody. One of the speech synthesizer applications is on the Text-to-Speech system. Speech synthesizer is a system that is capable of producing sound imitation with human speech synthesis. There are three method of speech synthesizer, that is formant synthesis, articulatory synthesis, concatenative synthesis.

Diphone concatenation is the method used in this final assignment. Initially, speech synthesis formed with voice recording and the results stored in database. Then it brakes into diphone that has a transition between two neighboring sound (adjacent phones) so it will be more stable when combined with another diphone. To combine speech diphone units, Time Domain Pitch Synchronous Overlapping-Add (TD PSOLA) algorithm is used. By using the TD algorithm PSOLA, combination between diphone containing transition between two neighboring sounds (adjacent phones), will be smooth without the explosive sound.

Combining diphone concatenation method with TD PSOLA algorithm would make the speech synthesis sounds clear, smooth and as natural as the voice of human being without prosody. This is the result of using pitch synchronization during the pitch modification process before concatenation. Proven by MOS index that reach value of 3,4. Even though in some particular case, the speech synthesis will produce an explosive sound, such as “click” or “pop”. The more diverse units of the diphone database will facilitate the establishment of speech synthesis, but the larger memory will be needed to store them.

Keywords: diphone concatenation, pitch, TD PSOLA algorithm, speech synthesizer