

## ABSTRAK

Perkembangan *speech synthesizer* saat ini meningkat cukup pesat. Berawal dari hasil suara yang didapatkan tidak terdengar alami sama sekali, kemudian menuju ke arah prosodi yang semakin bagus. Salah satu contoh penerapan *speech synthesizer* adalah pada *Text-to-Speech*. *Speech synthesizer* berada pada blok terakhir dalam sistem *Text-to-Speech*. *Speech synthesizer* merupakan sebuah sistem yang mampu menghasilkan suara tiruan manusia dengan sintesa ucapan. Metode *speech synthesizer* yang ada saat ini adalah *formant synthesis*, *articulatory synthesis*, *concatenative synthesis*.

Metode yang digunakan dalam tugas akhir ini adalah metode *diphone concatenation*. Mula-mula sintesa ucapan dibentuk dengan melakukan perekaman suara dan hasilnya disimpan dalam *database*. Kemudian suara rekaman tersebut dipecah menjadi *diphone* yang memiliki transisi antar dua bunyi yang berdekatan (*adjacent phones*) sehingga akan lebih stabil saat digabungkan dengan *diphone* yang lain. Untuk menggabungkan unit ucapan *diphone*, digunakan algoritma *Waveform Similarity Overlap-Add* (WSOLA). Dengan menggunakan algoritma WSOLA, perangkaian antar *diphone* yang mengandung transisi antar dua bunyi yang berdekatan (*adjacent phones*), menjadi halus tanpa bunyi yang bersifat eksplosif.

Dari metode *diphone concatenation* dan penerapan algoritma WSOLA maka sintesis ucapan yang dihasilkan ternyata dapat dimengerti dengan jelas, lancar dalam pengucapan dan datar tanpa intonasi. Semakin beranekaragam unit *diphone* yang terdapat pada *database* akan memudahkan dalam pembentukan sintesis ucapan, sehingga akan meningkatkan kualitas hasil sintesis. Namun semakin besar memori yang dibutuhkan untuk menyimpan unit-unit *diphone* tersebut. Dari hasil penilaian *Mean Opinion Score* (MOS), parameter *intelligibility* mencapai nilai 3,41 dan *fluidity* yang mencapai 3,35 serta *naturalness* mencapai nilai 3,28. Dengan demikian kemampuan sistem dalam mensintesis suara ucapan manusia termasuk dalam kategori cukup.

**Kata kunci** : *speech synthesizer*, *diphone concatenation*, algoritma WSOLA