

## ABSTRAKSI

Perkembangan *Text-to-Speech* saat ini sudah berkembang cukup pesat, awalnya *Text-to-Speech* menghasilkan suara yang tidak natural sama sekali. Kemudian perkembangan *Text-to-Speech* menuju ke arah prosodi yang semakin bagus. Dengan adanya perkembangan animasi, sekarang ini perkembangan *Text-to-Speech* menuju ke arah visualisasinya yang dikenal dengan nama *Text-to-Video*.

Video terdiri dari suara dan gambar yang bergerak. *Text-to-Video* adalah gabungan dari dua sistem pokok yaitu *Text-to-Speech* dan *Facial Animation*. Untuk menghasilkan suara pada *Text-to-Video* menggunakan metode *Text-to-Speech* sedangkan untuk gambar yang bergerak digunakan *Facial Animation*.

Metode sintesis *Text-to-Speech* yang ada saat ini yaitu sintesis *formant*, sintesis *concatenation*, dan sintesis *articulatori*. Sedangkan metode yang digunakan dalam Tugas Akhir ini adalah metode *diphone concatenation*, dengan database *diphone* yang dibuat sendiri. Sedangkan pada gambar Bergeraknya, pada prinsipnya menggunakan *facial animation*. Proses utama dalam video ini yaitu proses *morphing* beberapa gambar yang telah disediakan dalam database. Proses *morphing* yang dipakai dalam Tugas Akhir ini adalah *cross dissolve*.

Berdasarkan hasil rata-rata nilai MOS yang didapat dari 30 responden, menunjukkan hasil dari *Text-to-Video* cukup bagus, bahkan jika dilihat dari kenaturalan animasi wajah nilai MOS yang didapat sudah cukup tinggi (4,0733).