ABSTRAK

Proses anotasi suara berupa sinkronisasi berkas audio dengan transkrip pada level kata dengan hasil informasi waktu mulai dan berakhir setiap kata dapat dilakukan otomatis dengan *forced alignment.* Salah satu alat yang dapat melakukan *forced alignment* adalah Gentle. Gentle yang menerapkan model akustik *deep neural network* (DNN) untuk proses anotasi suara perlu diteliti karena pada bidang *automatic speech recognition* (ASR) penggunaan DNN menunjukkan peningkatan performa dibandingkan dengan model *monophone* maupun *triphone* yang biasanya digunakan pada alat *forced alignment* lainnya.

Penelitian ini menganalisis implementasi *forced alignment* dengan model akustik DNN pada tools Gentle sehingga diketahui tingkat akurasi dan *robustness* dibandingkan dengan proses *manual alignment*. Pembuatan model akustik diawali dengan pelatihan model *monophone* dan *triphone* hingga diakhiri dengan pelatihan model DNN. Hasil penelitian menunjukkan model akustik DNN pada Gentle lebih akurat mengolah data tanpa *noise* (*median* 0.82 dan *mean* 0.78) dibanding data *noise* (*median* 0.77 dan *mean* 0.73). Selain itu, model juga mempercepat proses anotasi hingga 87 kali lebih cepat. Namun, model belum cukup *robust* pada data tanpa *noise* (*median* 25.96 ms dan *mean* 48.99 ms) dan data *noise* (*median* 28.07 ms dan *mean* 67.91 ms) untuk ambang batas 20 ms dibandingkan *manual alignment.* Hal ini berarti peninjauan ulang secara manual harus tetap dilakukan meskipun dapat mempercepat proses anotasi.

Kata kunci: *forced alignment;deep neural network; voice annotation*