ABSTRAK

Ucapan adalah cara yang mudah untuk mendeteksi keadaan pembicara, dan pengenalan emosi ucapan telah diteliti secara ekstensif di bidang interaksi manusia-mesin. Saat ini, sebagian besar metode pengenalan emosi ucapan menggunakan parameter dari OpenSmile. Namun, metode ini tidak dapat menangkap keadaan dinamis pembicara karena parameternya diambil dari data regresi. Metode kami mencoba menangkap keadaan dinamis emosional dari sinyal suara dan mengekstraksi beberapa parameter darinya. Penelitian ini mengusulkan Discrete Wavelet Transform (DWT) untuk menguraikan sinyal suara menjadi beberapa level sehingga dapat dipilih level dengan noise yang minimal. Hasil dekomposisi DWT menunjukkan bahwa level 4 sampai 6 memberikan sinyal dengan level noise paling kecil. Diasumsikan bahwa informasi emosional berada pada segmen suara sehingga pada penelitian ini diusulkan untuk melakukan segmentasi suara untuk memisahkan segmen suara dari yang tidak bersuara untuk proses ekstraksi parameter. Metode yang diusulkan mengadopsi parameter emosional yang ada seperti Zero-Crossing Rate, Energy, Peak, Fourier Coefficients, dan Cepstrum. Penelitian ini menggunakan Neural Networks untuk klasifikasi. Hasil eksperimen menunjukkan bahwa DWT level 6 mencapai delapan klasifikasi emosi dengan akurasi 98%.

Kata kunci: Pengenalan Emosi; Discrete Wavelet Transform; Segmentasi Suara; Jaringan syaraf; Ekstraksi Parameter.