Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK) p-ISSN: 2355-7699
Vol. 4, No. 3, September 2017, hlm. 168-172 e-ISSN: 2528-6579

PENGENALAN EMOSI BERDASARKAN SUARA MENGGUNAKAN ALGORITMA HMM

Barlian Henryranu Prasetio[1], Wijaya Kurniawan[2], Mochammad Hannats Hanafi Ichsan[3]

[1,2,3]Laboratorium Sistem Komputer dan Robotika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: [1]barlian@ub.ac.id, [2]wjaykurnia@ub.ac.id, [3]hanas.hanafi@ub.ac.id

(Naskah masuk: 16 Mei 2017, diterima untuk diterbitkan: 19 Agustus 2017)

Abstrak

Penelitian ini bertujuan mengenali emosi seseorang melalui ucapan menggunakan algoritma HMM. Sistem dibangun dapat mengenali 3 jenis emosi yaitu marah, bahagia dan netral. Fitur yang digunakan dalam sistem ini adalah *pitch*, energi dan *formant*. *Database* yang digunakan adalah suara dari rekaman film. Dari hasil obeservasi probabilitas emosi marah sebesar 0.196, bahagia 0.254 dan netral 0.045. Sistem memiliki tingkat akurasi rata-rata sebesar 86.66%. Rata waktu eksekusi sistem dalam mendeteksi dan mengklasifikasikan emosi sebesar 21.6ms.

Kata kunci: *suara, emosi, HMM, klasifikasi*

Abstract

This research aims to recognize human emotions through voice using HMM algorithm. The system can confirm three types of emotions: anger, happiness and neutrality. The features used in this system are pitch, energy *and formant. From the results, the emotional probability of angry is 0.196, happy is 0.254 and neutral is 0.045. Base on testing result, the system has an average accuracy of 86.66% and average execution time of the system in detecting and classifying emotions of 21.6ms.*

Keywords: *voice, emotion, HMM, classification*

1. PENDAHULUAN

Emosi adalah perasaan intens yang ditujukan kepada seseorang atau sesuatu (N. H. Frieda, 1993). Selain itu, emosi dapat diartikan sebagai reaksi yang timbul akibat perbuatan seseorang atau pun kejadian tertentu. Jenis-jenis emosi dapat dikategorikan seperti kecemasan, kebosanan, ketidakpuasan, dominasi, depresi, jijik, frustrasi, takut, kebahagiaan, ketidakpedulian, ironi, sukacita, netral, panik, larangan, kejutan, kesedihan, stres, rasa malu, shock, kelelahan, stres beban tugas dan kuatir.

Beberapa penelitian menunjukkan bahwa beberapa parameter statistik memiliki korelasi yang tinggi antara ucapan dengan keadaan emosional pembicara (B. Heuft, 1996). Parameter tersebut adalah *pitch*, energi, artikulasi dan bentuk spektral. Misalnya, emosi kesedihan memiliki standar deviasi *pitch* yang rendah dan tingkat berbicara lambat, sementara emosi marah biasanya memiliki standar deviasi *pitch* yang lebih tinggi dan berbicara cepat (A. Nogueiras, 2001).

Bentuk emosional sesorang juga dapat dipengaruhi oleh tempat atau budaya tempat tinggalnya. Sebagai contoh, kalimat interogatif biasanya menyiratkan kontur *pitch* yang lebih luas daripada kalimat afirmatif, sehingga standar deviasi *pitch* mereka biasanya akan lebih tinggi. Namun hal ini tidak ada kaitannya dengan gaya emosional, hanya dengan sifat kalimat. Keterbatasan lain dari menggunakan statistik global adalah kenyataan bahwa pengolahan hanya dapat dilakukan setelah seluruh ucapan telah diucapkan. Fakta ini membatasi kemampuan membangun recogniser real time dan merupakan kelemahan utama ketika emosi bervariasi sepanjang ucapan (A. Nogueiras, 2001).

Sebuah pendekatan yang berbeda untuk statistik global ialah dengan mempertimbangkan bahwa jenis pemodelan hanya merupakan refleksi dari perilaku seseorang dalam waktu singkat (tertentu). Misalnya, kita menggunakan parameter mean dan standar deviasi dari fitur baku waktu singkat seperti energi atau *pitch*, maka kita bisa menghubungkan langsung parameter tersebut dengan *Probability Distribution Function* (pdf). Jika kita menambahkan pemodelan pdf dengan distribusi *Gaussian*, sama dengan menggunakan *Hidden Markov Mode*l (HMM) satu *state*. Yang harus diketahui bahwa statistik suara tidak stasioner. HMM memodelkan suara menjadi rangkaian *state*, yang berbeda untuk masing-masing model suara atau kombinasi suara, dan memiliki sifat statistik yang berbeda pula.

Hidden Markov Model (HMM) terdiri dari rantai markov pada bagian pertama yang menyembunyikan *state* oleh karena itu perilaku internal model tetap tidak terlihat. *State-state* yang tersembunyi dari model menangkap struktur temporal data. HMM merupakan model statistik yang menggambarkan urutan peristiwa. HMM memiliki keuntungan bahwa dinamika temporal fitur

ucapan dapat terdeteksi oleh Matrik *state* transisi. Selama *clustering*, sinyal ucapan diambil dan probabilitas untuk setiap sinyal suara dihitung. Output klasifikasi didasarkan pada probabilitas maksimum yang dimiliki model yang telah dihasilkan sinyal tersebut (B. Schuller, 2003). Untuk pengenalan emosi menggunakan HMM, pertama-tama yang dilakukan adalah *database* memilah sesuai dengan mode klasifikasi dan kemudian mengekstraksi fitur dari input gelombang yang diambil. Fitur-fitur ini kemudian ditambahkan ke *database*. Matrik transisi dan matrik emisi telah dibuat sesuai dengan mode, yang menghasilkan random urutan *state* dan emisi dari model (A. B. Ingale, 2012). Algoritma HMM memiliki tingkat akurasi yang relative lebih rendah dibandingkan dengan algoritma pengenalan suara yang lainnya, namun HMM lebih baik dalam pengenalan suara dengan noise yang tinggi (T. L. Pao, 2008).

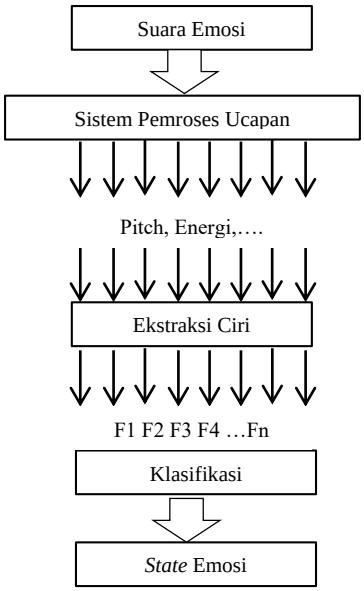
2. PENGENALAN EMOSI

Pengenalan emosi berbasis suara bertujuan untuk secara otomatis mengidentifikasi keadaan emosional manusia dari suaranya. Hal ini didasarkan pada analisis mendalam dari mekanisme generasi sinyal suara, penggalian beberapa fitur yang berisi informasi emosional dari suara pembicara, dan mengambil metode pengenalan pola yang tepat untuk mengidentifikasi keadaan emosi. Seperti sistem pengenalan pola yang khas, sistem pengenalan emosi terdiri dari empat modul utama: masukan ucapan, ekstraksi fitur, klasifikasi dan emosi keluaran (Y. Pan, 2012). Arsitektur umum sistem Pengenalan emosi berbasis suara memiliki tiga langkah yang ditunjukkan pada Gambar 1:

- Sebuah sistem pemroses ucapan, mengekstrak beberapa jumlah sinyal yang sesuai, seperti *pitch* atau energi,
- Jumlah ini diringkas (*summarized*) menjadi beberapa fitur yang sesuai atau dibutuhkan saja,
- Sebuah classifier learning dengan cara mengambil data sampel dan bagaimana menghubungkan fitur ke emosi.

Pada Gambar 1 dapat dilihat bahwa setelah melakukan *pre-processing*, suara dimodelkan berdasarkan cirinya. Ekstraksi ciri berdasarkan pada partisi ucapan dalam interval kecil yang dikenal sebagai *frame*. Untuk memilih fitur yang sesuai yang membawa informasi tentang emosi dari sinyal suara merupakan langkah penting dalam sistem pengenalan emosi berbasis suara.

Energi adalah fitur dasar dan paling penting dalam sinyal suara. Untuk mendapatkan nilai statistik dari fitur energi, kita menggunakan fungsi *short-term* untuk mengekstrak nilai energi di setiap *frame* ucapan. Kemudian kita dapat memperoleh nilai statistic energi dalam keseluruhan sampel dan menghitung energi, seperti nilai mean, nilai maks, varian, range variasi, kontur energi (D. Ververidis, 2004).



Gambar 1. Sistem Pengenalan Emosi Berbasis Suara

Tingkat getaran vokal disebut frekuensi fundamental F0 atau frekuensi *pitch*. Sinyal *pitch* memiliki informasi tentang emosi, karena tergantung pada ketegangan pita suara dan *sub-glottal* tekanan udara, sehingga nilai rata-rata dari *pitch*, varian, variasi *range* dan kontur yang berbeda dalam tujuh status emosional dasar (Y. Pan, 2012).

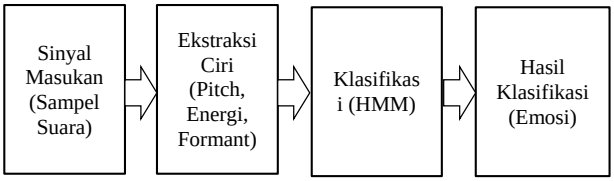
Nilai statistik berikut, dihitung dari *pitch* dan digunakan dalam vektor fitur *pitch* (F. Yu, 2001):

- Mean, Median, Variance, Maksimum, Minimum (untuk vektor fitur *pitch* dan turunannya)
- Energi rata-rata suara dan tanpa suara
- Kecepatan ucapan (kebalikan dari panjang rata-rata bagian ucapan).

Mel-Frequency cepstrum Coefficient (MFCC) adalah fitur yang paling penting dari ucapan. MFCC memiliki resolusi frekuensi yang baik, pada frekuensi rendah. Selain itu, MFCC juga memiliki ketahanan terhadap kebisingan yang baik. MFCC mengambil rata-rata atau nilai mean logaritmik *spektrum* setelah *Mel Filter Bank* dan Frekuensi wrapping (Y. Pan, 2012). LPCC mengandung karakteristik tertentu pada saluran bicara. Seseorang ketika sedang dalam emosional yang berbeda akan memiliki karakteristik saluran bicara yang berbeda pula, sehingga kita dapat mengekstraksi koefisien fitur ini untuk mengidentifikasi emosi yang terkandung dalam ucapannya.

3. METODE

Dalam perancangan, sistem mampu mendetesi dan mengklasifikasikan emosi berdasarkan suara yang diterima. Sistem pengenalan emosi biasanya mengenali sejumlah 3-5 jenis emosi. Penelitian ini menggunakan 3 jenis emosi yaitu marah, bahagia, dan netral. Secara umum, teknik pengenalan emosi dapat dilihat pada Gambar 2.



Gambar 2. Teknik Pengenalan Emosi

Pada Gambar 3 dapat dilihat bahwa sistem terdiri dari 2 bagian utama yaitu ekstraksi fitur dan klasifikasi. Bagian awal sistem adalah input suara yang dicacah menjadi *frame* yang merupakan sampel suara. Kemudian sampel suara ini dilakukan ekstraksi ciri berdasarkan *pitch*, energi dan frekuensi *formant*. Dari hasil perhitungan ekstraksi ciri, diklasifikasikan menggunakan teknik HMM.

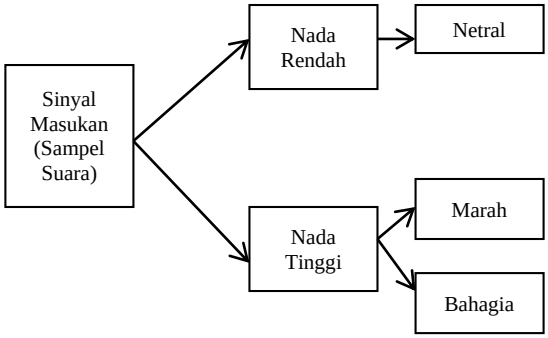
3.1. Ekstraksi Ciri

Ekstraksi ciri berdasarkan pada partisi ucapan dalam interval kecil yang dikenal sebagai *frame*. Untuk memilih fitur yang sesuai yang membawa informasi tentang emosi dari sinyal suara merupakan langkah penting dalam sistem pengenalan emosi berbasis suara. Terdapat dua jenis fitur: prosodic fitur energi, *pitch* dan fitur spektral termasuk MFCC, MEDC, LPCC.

Proses ekstraksi ciri dilakukan untuk mendapatkan nilai dari parameter *pitch*, energi dan *formant*. Proses ini dimulai dengan menyaring sinyal suara menggunakan 2 filter yaitu *Low Pass Filter* (LPF) dengan frekuensi *cut-off* 3.5kHz dan *Finite Impulse Response* (FIR). Sinyal suara keluaran kedua filter ini berupa sinyal diskrit dan flatten. Kemudian, amplitude dari sinyal tersebut dikuadratkan sehingga mendapatkan energi. Selain itu, sinyal keluaran filter juga disegmentasi tiap 10ms. Setelah dilakukan segmentasi, sinyal suara dimaksukkan pada bagian *Linear Prediction Coding* (LPC). Pada LPC sinyal diumpan balikkan menggunakan filter adaptif sehingga memiliki 2 keluaran, yaitu sinyal *predicted* $\hat{S}(n)$ dan *predicted error* $e(n)$. Setelah itu melakukan transformasi pada sinyal $\hat{S}(n)$ dan $e(n)$ menggunakan *Discreate Fourier Transform* (DFT). Hasil DFT dari sinyal $\hat{S}(n)$ diambil nilai puncak pada tiap *picking*-nya. Nilai puncak *picking* sinyal ini disebut *formant*. Sedangkan hasil DFT dari sinyal $e(n)$ dilakukan proses logaritmik pada *absolute magnitude*-nya. Sinyal tersebut kemudian dilakukan transformasi kembali menggunakan DFT. Hasil sinyal ini disebut *cepstrum*. Nilai puncak tiap *picking* pada sinyal *cepstrum* disebut *pitch*.

3.2. Pemodelan HMM

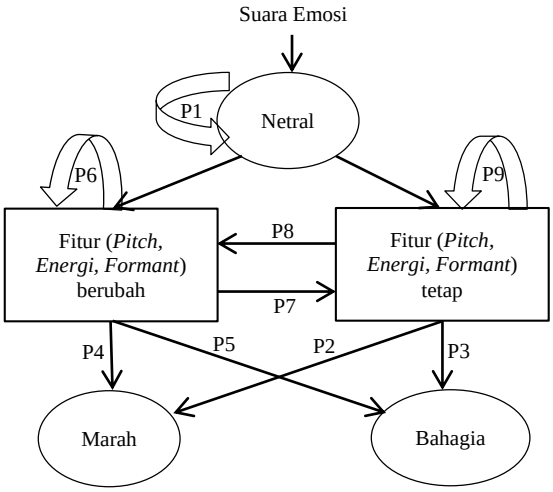
Setiap sinyal suara akan memiliki fitur *pitch*, energi dan *formant*. Ketiga fitur tersebut dihitung nilai mean-nya. Dalam HMM, klasifikasi suara dilakukan sesuai dengan mode klasifikasi sinyal masukan yang diambil. Fitur-fitur ini kemudian ditambahkan ke *database*. Matrik transisi dan matrik emisi dibuat sesuai dengan probabilitas emosi sehingga menghasilkan urutan *state* (A. B. Ingale, 2012). Sistem pengklasifikasikan emosi menggunakan HMM. Teknik klasifikasi HMM dapat dilihat pada Gambar 3.



Gambar 3. Klasifikasi Suara

Pada Gambar 3 dapat dilihat bahwa sinyal masukan suara dikategorikan menjadi 2 bagian yaitu nada rendah dan tinggi. Nada rendah menhasilkan emosi netral. Nada tinggi menghasilkan nada negatif untuk marah atau nada positif untuk bahagia.

Pada bagian awal, setiap suara diasumsikan sebagai emosi netral. Jika diketahui sebuah suara memiliki mean *pitch* (Mx), mean energi (My) dan mean *formant* (Mz), maka berdasarkan training sistem, probabilitas emosi tetap netral adalah P1, emosi marah adalah P2 dan emosi bahagia adalah P3. Namun suara tersebut dapat dipengaruhi oleh perubahan fitur (*pitch*, energi dan *formant*) yang mempengaruhi factor hidden transisi. Jika suara berubah fitur maka probabilitas emosi marah adalah P4 dan emosi bahagia adalah P5. Jikasuara memiliki fitur berubah, maka akan tetap berubah dengan probabilitas P6 dan dapat menjadi fitur tetap dengan probabilitas P7. Sementara jika suara memiliki fitur tetap, maka suara akan menjadi fitur berubah dengan probabilitas P8 dan akan menajdi fitur tetap dengan probabilitas P9. Ilustrasi pemodelan HMM untuk mengenali emosi dapat dilihat pada Gambar 4.



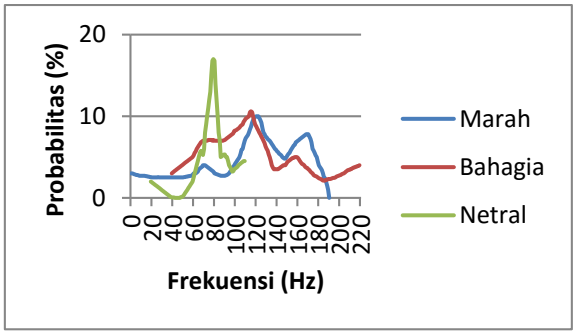
Gambar 4. Diagram *State* Pemodelan HMM

4. HASIL DAN PEMBAHASAN

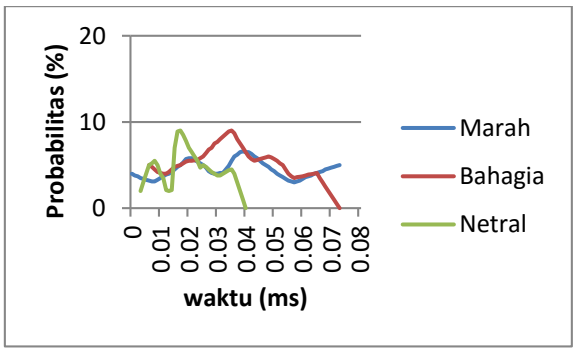
Penelitian ini menggunakan suara pria sebagai data latih dan data uji.

4.1. Probabilitas Emosi

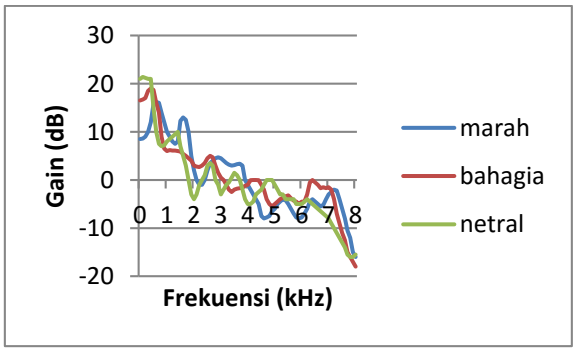
Dalam melakukan ekstraksi ciri, sistem menggunakan 3 fitur yaitu *pitch*, energi dan *formant*. Setiap suara memiliki kontur emosi yang berbeda pada tiap frekuensi. Kontur probabilitas emosi dalam frekuensi untuk fitur *pitch* dapat dilihat pada Gambar 5, fitur energi dapat dilihat pada Gambar 6 dan fitur *formant* dapat dilihat pada Gambar 7, berikut masing-masing warna yang memberikan informasi terkait kondisi emosinya.



Gambar 5. Kontur Probabilitas Emosi Dalam Frekuensi Untuk Fitur *Pitch*



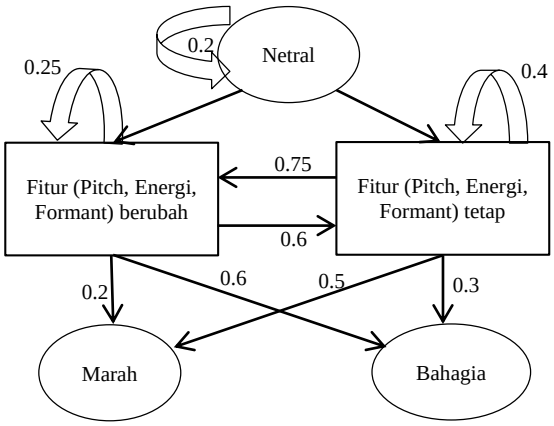
Gambar 6. Kontur Probabilitas Emosi Dalam Frekuensi Untuk Fitur Energi



Gambar 7. Kontur Probabilitas Emosi Dalam Frekuensi Untuk Fitur *Formant*

4.2. Klasifikasi Emosi

Berdasarkan hasil observasi ekstraksi ciri emosi menggunakan suara dapat dibuat diagram *state* untuk 3 emosi dengan asumsi frekuensi dasar F0=250Hz dengan gender pria. Diagram *state* HMM dapat diliha pada Gambar 8.



Gambar 8. Diagram *State* HMM Klasifikasi Suara

4.3. Pengujian

Pengujian dilakukan dalam 2 skenario. Skenario pertama sistem diberikan sinyal masukan suara yang kemudian disimpan dalam *frame* dengan panjang 10 detik. Sistem diberi masukan 3 jenis emosi dan masing-masing emosi terdiri dari 5 suara uji. Hasil klasifikasi sistem dibandingkan dengan dataset untuk dilihat tingkat akurasinya. Data set adalah suara yang telah diketahui jenis emosinya dan disimpan pada *database*. *Database* yang digunakan adalah suara dari rekaman film. Hasil pengujian akurasi sistem dapat dilihat dan Tabel 1.

Tabel 1. Hasil Pengujian Akurasi Sistem

Suara	Klasifikasi Sistem	DataSet	Akurasi
1	Marah	Marah	1
2	Marah	Bahagia	0
3	Marah	Marah	1
4	Marah	Marah	1
5	Marah	Marah	1
6	Bahagia	Bahagia	1
7	Bahagia	Bahagia	1
8	Bahagia	Marah	0
9	Bahagia	Bahagia	1
10	Bahagia	Bahagia	1
11	Netral	Netral	1
12	Netral	Netral	1
13	Netral	Netral	1
14	Netral	Netral	1
15	Netral	Netral	1
	Rata-Rata		86.66%

Skenario pengujian kedua menghitung waktu eksekusi pengenalan emosi. Waktu dihitung mulai masuknya sinyal suara sampai sistem memberikan hasil klasifikasi emosi yang terdeteksi. Waktu eksekusi sistem dalam mengenali emosi dapat dilihat pada Tabel 2.

Tabel 2. Waktu Eksekusi Sistem

Suara	Klasifikasi Sistem	Waktu (ms)
1	Marah	20.5
2	Marah	20.1
3	Marah	19.5
4	Marah	19.0
5	Marah	20.1
6	Bahagia	22.3
7	Bahagia	21.9
8	Bahagia	22.4
9	Bahagia	22.9
10	Bahagia	21.0
11	Netral	22.6
12	Netral	22.9
13	Netral	23.0
14	Netral	23.1
15	Netral	22.8
	Rata-Rata	21.6

5. KESIMPULAN

Dari hasil perancangan dan pengujian sistem dapat disimpulkan sebagai berikut:

- Sistem dapat mengenali emosi marah, bahagia dan netral
- Fitur yang digunakan dalam sistem adalah *pitch*, energi dan *formant*
- Sistem mengkalsifikasikan emosi menggunakan HMM
- Dari hasil obeservasi probabilitas emosi marah sebesar 0.196, bahagia 0.254 dan netral 0.045.
- Sistem memiliki tingkat akurasi rata-rata sebesar 86.66%
- Rata waktu eksekusi sistem dalam mendeteksi dan mengklasifikasikan emosi sebesar 21.6ms

6. DAFTAR PUSTAKA

B. INGALE, D. S. CHAUDHARI, 2012, Speech Emotion Recognition Using Hidden Markov Model And Support Vector Machine. *International Journal of Advanced Engineering Research and Studies (IJAERS)*, Vol. I, Issue 3, April-June, 316-318.

A. NOGUEIRAS, A. MORENO, A. BONAFONTE, J. B. MARINO, 2001, Speech emotion recognition using hidden Markov models, EUROSPEECH 2001 Scandinavia, *7th European Conference on Speech Communication and Technology*, 2nd INTERSPEECH Event, Aalborg, Denmark, pp: 2679-2682

B. HEUFT, T. PORTELE, AND M. RAUTH, 1996, Emotions in time domain synthesis, *inProc. of ICSLP*, Philadelphia, pp. 1974–1977

B. SCHULLER, G. RIGOLL, M. LANG, 2003, Hidden Markov model-based Speech emotion recognition, *Proceedings of the IEEE ICASSP Conference on Acoustics, Speech and Signal Processing*, vol. 2, pp. 1-4.

D. VERVERIDIS, C. KOTROPOULOS, AND I. PITAS, 2004, Automatic emotional speech classification, in Proc. 2004 *IEEE Int. Conf. Acoustics*, Speech *and Signal Processing*, vol.1, pp. 593-596, Montreal, May.

F.YU, E.CHANG, Y.XU, H.SHUM, 2001, Emotion detection from speech to enrich multimedia content, *Lecture Notes in Computer Science*, Vol. 2195, 550-557.

N.H. FRIEDA, 1993, Moods, Emotion Episodes and Emotions, *New York: Guilford Press*, hal. 381-403.

T. L. PAO, W. Y. LIAO, Y. T. CHEN, J. H. YEH, 2008, Comparison of Several Classifiers for Emotion Recognition from Noisy Mandarin Speech, *Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Vol. 1, pp: 23-26

Y. PAN, PEIPEI SHEN AND LIPING SHEN, 2012, Speech Emotion Recognition Using Support Vector Machine, *International Journal of Smart Home*, Vol. 6, No. 2, April.