

Identifikasi Emosi Dari Sinyal Suara Secara Real Time Menggunakan Linear Predictive Coding dan Backpropagation

Yamina Azmi¹, Dr. Esmeralda C. Djamal, Ridwan Ilyas
Program Studi Informatika Fakultas Sains dan Informatika
Universitas Jenderal Achmad Yani
Cimahi
¹yaminaazmi01@gmail.com

Abstrak— *Kepuasan pelanggan adalah salah satu faktor penting suatu perusahaan yang menghasilkan produk dan jasa. Banyak perusahaan saling bersaing untuk meningkatkan kualitas pelayanan salah satunya dengan melakukan evaluasi terhadap customer service dalam melayani konsumen. Penelitian ini telah membuat sistem identifikasi emosi dari sinyal suara. Kalimat yang digunakan berasal dari rekaman percakapan antara customer service dengan pelanggan. Sistem identifikasi ini dibangun dari pelatihan suara terhadap 10 naracoba dengan lima kali perulangan pada setiap kata kunci yang diucapkan, sehingga digunakan 150 set data latih. Sinyal suara setiap enam detik dilewatkan praproses untuk mengurangi noise dan dilakukan frame blocking dan windowing yang kemudian diekstraksi menggunakan Linear Predictive Coding (LPC) dengan nilai poles 8. Koefisien yang dihasilkan LPC kemudian digunakan sebagai fitur dalam Jaringan Syaraf Tiruan menggunakan algoritma pembelajaran Backpropagation. Hasil dari penelitian menunjukkan sistem identifikasi emosi menghasilkan akurasi sebesar 91% untuk data latih dan 76% untuk data baru.*

Kata kunci— *identifikasi emosi, sinyal suara, Linear Predictive Coding, Backpropagation, customer service*

I. PENDAHULUAN

Kepuasan pelanggan adalah salah satu faktor penting suatu perusahaan yang menghasilkan produk dan jasa. Banyak perusahaan saling bersaing untuk meningkatkan kualitas pelayanan salah satunya dengan melakukan evaluasi terhadap customer service dalam melayani pelanggan. Oleh karena itu penting bagi perusahaan menyediakan sarana keluhan pelanggan melalui *call center*, *help desk*, atau *counter service center*. Evaluasi ini sangat berpengaruh terhadap kemajuan perusahaan yang salah satunya telah dianalisis dalam penelitian sebelumnya mengenai pengaruh kepuasan pelanggan terhadap citra perusahaan [1].

Call center merupakan salah satu layanan keluhan pelanggan yang perlu dievaluasi kinerjanya, evaluasi ini dapat dilakukan dalam berbagai bentuk salah satunya dengan memperhatikan emosi customer service pada saat menanggapi keluhan. Emosi seseorang dapat diidentifikasi melalui citra wajah, sinyal otak, dan sinyal suara. Dalam konteks *call center* sinyal suara lebih cocok digunakan dalam identifikasi emosi [2]. Beberapa penelitian terdahulu telah melakukan identifikasi emosi untuk melihat perilaku berdasarkan emosi [3][4], mendeteksi emosi

manusia dari percakapan pembicara [5][6], pengenalan emosi untuk interaksi manusia dan komputer, pengenalan emosi pada *agent virtual* [7] dan membantu tim medis dalam pengobatan penyakit stress [3][8].

Dalam pengenalan suara, sinyal suara diekstraksi terlebih dahulu. Beberapa penelitian ekstraksi sinyal suara dapat menggunakan Mel-Frequency Cepstrum Coefficients (MFCC) ataupun Linear Predictive Coding (LPC) sedangkan untuk diidentifikasi menggunakan K-Nearest Neighbor (KNN), Backpropagation, dan Learning Vector Quantization (LVQ). Penelitian terdahulu menghasilkan tingkat keberhasilan yang didapatkan saat pengenalan suara menggunakan ekstraksi LPC dan Backpropagation adalah 100% [9]. Penelitian lain mengenali emosi dari sinyal suara dengan tingkat keberhasilan rata-rata 95,00% menggunakan metode LPC [10] dan membandingkan fitur LPC dengan MFCC untuk pengenalan emosi dari sinyal suara yang memperoleh fitur LPC memiliki tingkat akurasi 94,23% sedangkan untuk MFCC memiliki tingkat akurasi 89,14% [11][12].

Penelitian ini telah membuat sistem untuk identifikasi emosi dari sinyal suara secara *real time*. Emosi yang diidentifikasi adalah marah, netral, dan senang. Suara yang meninggi dan cepat dapat diidentifikasi dengan keadaan emosi marah, sedangkan suara yang lambat agak tinggi dapat diidentifikasi dengan keadaan emosi senang dan suara yang agak cepat diidentifikasi dengan keadaan emosi netral. Sistem identifikasi yang dirancang menggunakan praproses dengan Pre Emphasize untuk mengurangi noise yang muncul saat perekaman yang kemudian masuk tahap ekstraksi dengan LPC yang sebelumnya melalui proses Frame Blocking dan Windowing terlebih dahulu kemudian Backpropagation untuk mengidentifikasi emosi tertentu. Agar dapat melakukan identifikasi emosi, dilakukan pembelajaran mesin atau pelatihan terlebih dahulu menggunakan sinyal suara yang telah didefinisikan kelasnya yaitu sebanyak tiga kelas dengan kondisi “Marah”, “Netral”, dan “Senang”. Pelatihan menggunakan Backpropagation dengan 150 set data latih yang diambil dari 10 naracoba dengan 5 kali perulangan untuk setiap kondisi emosi. Generalisasi dari pembelajaran mesin merupakan koefisien atau disebut bobot yang disimpan dalam *database*, dan kemudian akan digunakan secara *real time* atau dalam teknologi informasi dikenal dengan istilah waktu nyata yang berarti kondisi

pengoprasian sistem yang dibatasi oleh rentang waktu dan memiliki tenggat waktu yang jelas untuk mengidentifikasi emosi seseorang.

II. METODE PENELITIAN

Tahap identifikasi emosi dari sinyal suara dapat dilihat pada Gambar 1. Pengambilan data sinyal suara untuk data latih dilakukan dengan kalimat yaitu, “Baik, Saya akan menjelaskan proses yang sebenarnya terjadi” dan untuk data uji dilakukan dengan kalimat, “Oh iya Bu... Ada yang bisa Saya bantu?”. Penentuan kalimat kunci tersebut dikarenakan kalimat tersebut seringkali diucapkan oleh *customer service*. Perekaman dilakukan menggunakan Microphone Condenser Taffwer BM8000 dengan SNR > 10dB, dengan nilai frekuensi sampling yang digunakan yaitu 8000 Hz menggunakan *channel* mono dengan resolusi 16bit. Perekaman dilakukan dengan durasi selama enam detik. Data diperoleh dari 10 naracoba, dengan masing – masing mengucapkan kalimat dengan lima kali perulangan pada setiap emosi, sehingga terbentuk 150 set data latih. Pengucapan harus dilakukan dengan artikulasi yang jelas serta durasi pengucapan harus mendekati enam detik, hal ini dilakukan untuk memperoleh data latih yang baik. Perekaman dilakukan pada waktu pagi hari dimulai dari jam 08.00 sampai 10.00 WIB untuk mendapatkan emosi senang, kemudian siang hari pada pukul 11.00 sampai 13.00 untuk mendapatkan emosi marah dan sore hari untuk dimulai dari jam 13.00 sampai 15.00 WIB untuk mendapatkan emosi netral.

Perekaman dilakukan ketika naracoba dalam keadaan sehat, suara normal dan sedang tidak dalam keadaan suara parau, tidak melakukan perekaman sesaat setelah bangun tidur, perekaman dilakukan dengan posisi naracoba duduk dengan badan tegak lurus dan perekaman dilakukan di tempat sunyi untuk menghindari besarnya *noise*. Kemudian sinyal suara masuk ke dalam praproses dengan Pre Emphasize untuk mempertahankan frekuensi ciri yang ada. Kemudian dimasukkan ke dalam Frame Blocking untuk membagi sinyal audio ke dalam beberapa *frame* dan Windowing untuk meminimalisir diskontinuitas sinyal menggunakan Metode Hamming Window.

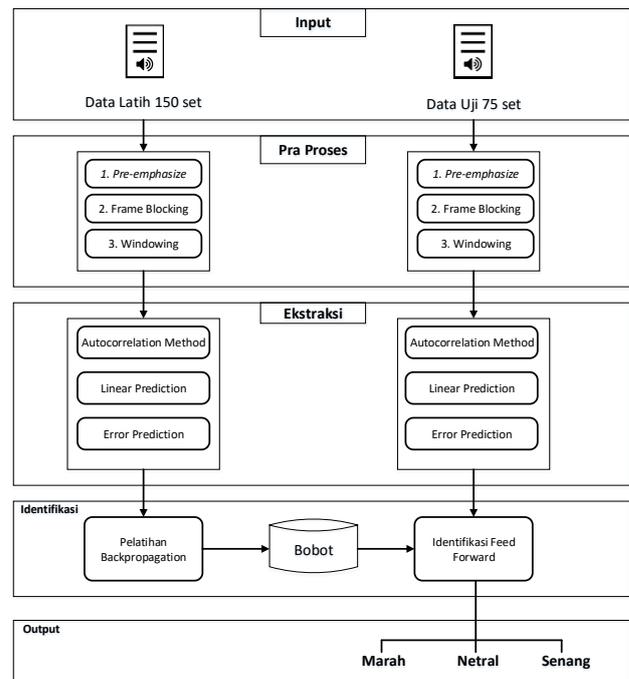
Sinyal suara yang telah melalui praproses selanjutnya diekstraksi untuk mendapatkan fitur atau ciri pembeda yang terdapat dalam sinyal suara. Ekstraksi dalam penelitian ini menggunakan metode Linear Predictive Coding (LPC). Penggunaan metode tersebut dalam ekstraksi sinyal suara dalam emosi sudah sering digunakan [13]. Terdapat tiga langkah dalam ekstraksi menggunakan LPC yaitu Autocorrelation Method, Linear Prediction dan Error Prediction. Setelah mendapatkan fitur dari hasil perhitungan ekstraksi, langkah selanjutnya adalah melakukan pelatihan menggunakan Backpropagation. Pelatihan ini dimulai dengan komputasi setiap neuron *input* terhadap setiap neuron pada *hidden layer*, yang diteruskan pada setiap neuron dalam *output layer*.

Penelitian ini menggunakan fungsi aktivasi Sigmoid Biner untuk membuat *output* berada dalam rentang 0 sampai 1. Inisialisasi bobot dilakukan secara *random* pada rentang -0.5 sampai 0.5. Luran dari tahap ini adalah bobot-bobot yang telah diperbaharui dan akan digunakan pada tahap identifikasi.

A. Praproses

Data masukan dilakukan praproses melalui tahapan Pre-Emphasize *filter*. Pre Emphasize berfungsi untuk mempertahankan frekuensi – frekuensi tinggi pada spektrum sehingga dihasilkan data *input* suara dengan *noise* yang rendah. Pre Emphasize termasuk salah satu jenis *filter* yang sering digunakan pada sinyal suara asli. *Filter* ini akan mempertahankan nilai frekuensi ciri dari sebuah spektrum.

Setelah melalui tahap Pre Emphasize kemudian dilakukan proses untuk mendapatkan fitur sinyal suara dengan Frame Blocking dan Windowing kemudian dengan LPC, Frame Blocking berfungsi untuk membagi sebuah sinyal suara ke dalam bentuk *short segmen* atau *short frame*, agar sinyal suara mengalami perubahan dalam jangka waktu tertentu. Sinyal suara selama enam detik, kemudian *disampling* setiap 0,02 detik. Oleh karena itu, satu perekaman menghasilkan 595 *frame*/detik dengan *overlapping* sebanyak 50%. Windowing berfungsi untuk mengurangi diskontinuitas sinyal yang dihasilkan dari proses sebelumnya, Windowing dilakukan dengan metode Hamming Window, dikarenakan Hamming Window menghasilkan *sidelobe* dan *noise* yang tidak terlalu tinggi.



Gambar 1. Perancangan sistem identifikasi emosi

B. Ekstraksi

Ekstraksi yang dilakukan adalah Frame Blocking dan Windowing terlebih dahulu yang kemudian dilakukan proses LPC. LPC merupakan cara yang paling sering digunakan untuk ekstraksi sinyal suara terhadap emosi, karena memiliki komponen yang sederhana.

Proses Frame Blocking dilakukan karena sinyal suara terus mengalami perubahan yang tidak teratur akibat adanya pergeseran artikulasi dari mulut. Panjang dari suatu *frame* biasanya digunakan untuk memproses sinyal antara 10-30 ms. Proses ini pada umumnya dilakukan secara *overlapping* pada tiap *frame*. Penelitian terdahulu menggunakan *overlapping* 50%

dari panjang *frame* keseluruhan [14]. Waktu pengambilan data sampling (T_s) ditetapkan setiap 20 ms dikarenakan karakteristik sinyal yang berubah dalam selang waktu tertentu untuk merefleksikan suara yang berbeda pada orde 0,2 detik atau lebih. Dengan Persamaan 1 maka didapat jumlah frame adalah 99 frame/detik;

$$\left(\frac{(I-N)}{M} + 1\right) \quad (1)$$

Dengan $I = \text{sample rate} : \frac{F_s}{T_s} = \frac{48000}{2} = 24000$

$N = \text{sample point} = 24000 \times 0.02 = 480$

$M = \frac{N}{2} = \frac{480}{2} = 240$

$f = \frac{24000-480}{240} + 1 = 99 \text{ frame}$

Proses Windowing merupakan proses pembobotan terhadap setiap *frame* berfungsi untuk mengurangi efek diskontinuitas pada ujung *frame* yang dihasilkan oleh Frame Blocking. Proses ini dilakukan dengan menggunakan Persamaan 2.

$$y(m) = x(m)w(m), 0 \leq m \leq N - 1 \quad (2)$$

Proses Windowing dilakukan karena proses *framing* dapat mengakibatkan kebocoran spektral atau *aliasing*. Terdapat banyak fungsi window, salah satunya Hamming Window. Hamming Window memiliki fungsi untuk menghasilkan *sidelobe* yang tidak terlalu tinggi dan noise yang dihasilkan pun rendah, sehingga banyak penelitian terdahulu menggunakan fungsi ini [2][3][7][14]. Dalam penelitian ini Hamming Window digunakan agar hasil yang muncul dapat lebih baik dalam pembatasan sinyal yang akan dianalisis. Fungsi Hamming Window didefinisikan pada Persamaan 3.

$$w(m) = \frac{1}{2} + 0,46 \cos \left[\frac{2\pi m}{N-1} \right], 0 \leq m \leq N - 1 \quad (3)$$

Linear Predictive Coding digunakan untuk fitur ekstraksi sinyal suara yang menyatakan ciri-ciri penting pada emosi tersebut dalam bentuk koefisien - koefisien LPC dengan cara memprediksi kombinasi linear sinyal data baru. Penelitian terdahulu telah mengenali emosi dari suara dengan tingkat keberhasilan rata - rata 95,00% menggunakan metode LPC dan metode Dynamic Time Warping untuk pencocokan pola [10].

Tujuan dari Analisa Linear Predictive Coding adalah untuk mendapatkan serangkaian koefisien (a_k) langsung dari sinyal suara. Koefisien harus diestimasi dari segmen-segmen pendek sinyal suara karena dasar sinyal suara yang bervariasi terhadap waktu. Pendekatan ini berfungsi untuk mendapatkan koefisien prediksi yang meminimalisir kuadrat rata-rata kesalahan pada satu segmen dari gelombang suara.

Tahap pertama yang dilakukan LPC adalah Autocorrelation Method dengan mencari korelasi antara data sebelum dengan data sesudahnya yang disusun berdasarkan waktu seperti pada Persamaan 6. Untuk mendapatkan banyaknya jumlah orde LPC (p) dapat dilakukan perhitungan seperti pada rumus seperti pada Persamaan 4.

$$p = \left(\frac{F_s}{1000\text{Hz}} \right) + (2 \text{ to } 3) \quad (4)$$

$$r_n(k) = \sum_{m=0}^{N-1-k} s(n)s(n-k) \quad (5)$$

$$\begin{bmatrix} r_n(0) & r_n(1) & r_n(p-1) \\ r_n(1) & r_n(0) & r_n(p-2) \\ \vdots & \vdots & \vdots \\ r_n(p-1) & r_n(p-2) & r_n(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} r_n(1) \\ r_n(2) \\ \vdots \\ r_n(p) \end{bmatrix} \quad (6)$$

Kemudian mencari kesalahan nilai linear prediksi dengan rumus perhitungan seperti pada Persamaan 7.

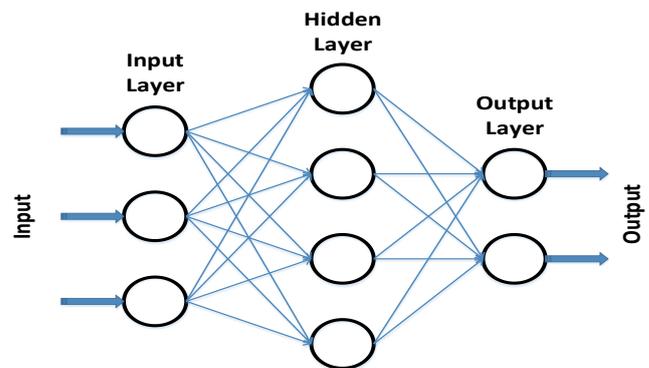
$$\hat{s}(n) = \sum_{k=1}^p a_k s(n-k) \quad (7)$$

Tahap ketiga adalah mencari nilai prediksi error seperti pada Persamaan 8.

$$e(n) = s(n) - \hat{s}(n) = s(n) - \sum_{k=1}^p a_k s(n-k) \quad (8)$$

C. Backpropagation

Backpropagation merupakan suatu metode dalam Jaringan Syaraf Tiruan (JST) dengan metode pelatihan *supervised*. Arsitektur jaringan Backpropagation untuk identifikasi emosi dapat dilihat pada Gambar 2. Beberapa penelitian menggunakan Backpropagation untuk analisis suara yaitu membuka dan menutup aplikasi dengan pengujian terbaik menggunakan *hidden neuron* 153 [15], pengenalan digit 0 sampai 9 menggunakan dua *hidden layer* dengan jumlah neuron 35 [16], aplikasi pengenalan ucapan untuk membuka dan menutup pintu menggunakan model neuron 100, 200, 1 dengan iterasi yang digunakan adalah 5000 [14], dan identifikasi bahasa otomatis dengan 2 *hidden layer* dan 30 neuron *hidden* [17]. Penelitian lainnya menggunakan 250 *neuron input*, satu *hidden layer* dengan 16 neuron *hidden*, dan 10 neuron *output* yang digunakan untuk mengenali 10 kata ucapan menghasilkan akurasi 96,332% untuk pembicara yang dikenal dan 92% untuk pembicara yang tidak dikenali [18].



Gambar 2. Arsitektur Multilayer Perceptron

Dalam JST terdapat fungsi aktivasi yang berfungsi untuk menentukan keluaran pada suatu neuron. Fungsi aktivasi yang digunakan harus memenuhi beberapa syarat yaitu kontinu, terdiferensial dengan mudah dan fungsi yang tidak turun [16]. Fungsi yang memenuhi syarat tersebut adalah Sigmoid Biner yang ditunjukkan pada Persamaan 9.

$$f(x) = \frac{1}{1 + e^{-x}} \quad (9)$$

Pelatihan menggunakan Backpropagation terdapat 3 fase. Fase 1 adalah *feed forward* yaitu setiap neuron *input* akan mengirimkan sinyal masukan pada *hidden layer*. Masing-masing neuron di *hidden layer* dikalikan dengan bobot dan

dijumlahkan dengan *bias* seperti ditunjukkan pada Persamaan 10 dan menghitung bobot neuron *hidden* dengan Persamaan 11.

$$z_{net_j} = v_{j0} + \sum_{i=1}^n x_i v_{kj} \quad (10)$$

$$z_j = f(z_{net_j}) = \frac{1}{1 + e^{-z_{net_j}}} \quad (11)$$

Masing-masing neuron *output* ($y_k, k = 1, 2, 3, \dots, m$) dikalikan dengan bobot dan dijumlahkan dengan *bias* seperti pada Persamaan 12 dan menghitung bobot neuron *output* dengan Persamaan 13.

$$y_{net_k} = w_{k0} + \sum_{j=1}^p z_j w_{kj} \quad (12)$$

$$y_k = f(y_{net_k}) = \frac{1}{1 + e^{-y_{net_k}}} \quad (13)$$

Perhitungan kuadrat *error* pada *output layer* menggunakan Mean Square Error (MSE). Nilai MSE merepresentasikan rata-rata kesalahan *error* absolut antara hasil identifikasi dengan nilai sebenarnya (*target*). Jika nilai diperoleh nol maka tidak ada lagi kesalahan, nilai yang lebih rendah dari nilai yang telah ditetapkan berarti hasilnya lebih baik [19]. Menghitung MSE dilakukan dengan menggunakan Persamaan 14.

$$\nabla E = \sum_{k=0}^m (t_k - y_k)^2 \quad (14)$$

Fase 2 adalah *backpropagation*, masing-masing neuron *output* ($y_k, k = 1, 2, 3, \dots, m$) menerima pola target yang telah dicapai sesuai dengan pola masukan saat pembelajaran lalu hitung perubahan *bias* dengan menggunakan Persamaan 17.

$$\delta_k = (t_k - y_k) f'(y_{net_k}) = (t_k - y_k) y_k (1 - y_k) \quad (15)$$

$$\Delta W_{kj} = \alpha \delta_k z_j; k = 1, 2, 3, \dots, m; j = 0, 1, \dots, p \quad (16)$$

$$\Delta W_{k0} = \alpha \delta_k \quad (17)$$

Pada setiap *hidden layer* dilakukan perhitungan kesalahan *hidden layer* kemudian informasi kesalahan tersebut digunakan untuk menghitung besar koreksi bobot dan *bias* antara *input layer* dan *hidden layer*.

$$\delta_{net_j} = \sum_{k=1}^m \delta_k w_{kj} \quad (18)$$

$$\delta_j = \delta_{net_j} f'(\delta_{net_j}) = \delta_{net_j} z_k (1 - z_j) \quad (19)$$

Hitung suku perubahan bobot pada *hidden layer* dengan menggunakan Persamaan 20.

$$\Delta v_{ji} = \alpha \delta_j x_i \quad (20)$$

Hitung suku perubahan *bias* pada *hidden layer* dengan menggunakan Persamaan 21.

$$\Delta v_{j0} = \alpha \delta_j \quad (21)$$

Fase 3 adalah modifikasi bobot yaitu setelah semua faktor δ dihitung, semua bobot dimodifikasi bersamaan.

$$W_{kj} = W_{kj}(\text{lama}) + \Delta W_{kj} \quad (22)$$

$$V_{ji} = V_{ji}(\text{lama}) + \Delta V_{ji} \quad (23)$$

III. HASIL DAN DISKUSI

Data latih yang digunakan pada penelitian sistem identifikasi emosi dari sinyal suara sebanyak 150 data yang didapatkan dari 10 naracoba lima diantaranya laki – laki dan lima orang lainnya perempuan, dari 10 naracoba tersebut ditetapkan rentang usia 20–27 tahun, masing-masing naracoba mengucapkan kalimat yang telah ditentukan dengan lima kali pengulangan dengan 150 data latih dan 75 untuk data uji. Pengujian dilakukan menggunakan data yang telah dilatih sebelumnya dan data yang tidak dilatih. Data baru didapatkan dari hasil perekaman naracoba yang sama dengan menggunakan SNR<10db.

Pelatihan pada data latih menggunakan Backpropagation dengan α 0.01 - 0.10 dengan pengurangan learning rate 0.01, serta menggunakan α 0.10 dengan penurunan learning rate 0.10, maksimum epoch 1000, dan minimum learning rate 0.0001. Analisis laju pembelajaran pelatihan Backpropagation dapat dilihat pada Tabel I.

TABEL I. PENGARUH OPTIMALISASI PARAMETER PELATIHAN

Learning rate	Epoch	Minimum learning rate	Waktu (detik)	Akurasi (%)	
				DL	DB
0.01	10000	0.001	13.0	34.58	34.50
		0.005	13.4	35.94	35.89
		0.009	17.8	35.61	35.59

Hasil analisis pada Tabel I menunjukkan bahwa parameter yang memberikan hasil akurasi paling baik adalah parameter dengan *learning rate* 0.01, *epoch* 10000, dan *minimum learning rate* 0.001 waktu pelatihan selama 13.4 detik dan *minimum learning rate* 0.005. Akurasi yang dihasilkan dengan parameter tersebut sebesar 35.94% dikenali untuk data baru dan 35.89% dikenali untuk data yang telah dilatih. Hasil pengujian terhadap 150 data latih dapat dilihat pada Tabel II dan III.

TABEL II. PENGUJIAN DATA LATIH SETIAP NARACOBA

Naracoba	Kelas	Jumlah Data	Jumlah Data Tepat Dikenali %
			Data Latih
Naracoba 1	Marah	5	80
	Normal	5	60
	Senang	5	90
Naracoba 2	Marah	5	90
	Normal	5	93
	Senang	5	50
Naracoba 3	Marah	5	70
	Normal	5	60
	Senang	5	75
Naracoba 4	Marah	5	80
	Normal	5	86
	Senang	5	90
Naracoba 5	Marah	5	91
	Normal	5	66

Naracoba	Kelas	Jumlah Data	Jumlah Data Tepat Dikenali %
			Data Latih
Naracoba 6	Senang	5	30
	Marah	5	95
	Normal	5	85
	Senang	5	87
Naracoba 7	Marah	5	70
	Normal	5	50
	Senang	5	49
Naracoba 8	Marah	5	78
	Normal	5	80
	Senang	5	59
Naracoba 9	Marah	5	80
	Normal	5	93
	Senang	5	70
Naracoba 10	Marah	5	85
	Normal	5	90
	Senang	5	70
Rata – rata akurasi			91

TABEL III. PENGUJIAN DATA BARU SETIAP NARACOBA

Naracoba	Kelas	Jumlah Data	Jumlah Data Tepat Dikenali %
			Data Baru
Naracoba 1	Marah	5	80
	Normal	5	60
	Senang	5	90
Naracoba 2	Marah	5	90
	Normal	5	93
	Senang	5	50
Naracoba 3	Marah	5	70
	Normal	5	60
	Senang	5	75
Naracoba 4	Marah	5	80
	Normal	5	86
	Senang	5	90
Naracoba 5	Marah	5	91
	Normal	5	66
	Senang	5	30
Rata – rata akurasi			76

Rata-rata akurasi diperoleh dengan menghitung persentase data yang dikenali dari semua data dengan Persamaan 21.

$$Akurasi = \frac{Jumlah\ data\ dikenal \times 100}{total\ semua\ data} \quad (21)$$

Tabel II dan III merupakan hasil pengujian terhadap data latih dan data baru menggunakan *learning rate* 0.01, minimum *error* 0.005 dan *epoch* 10000. Akurasi rata-rata dari data latih 91% dan untuk data baru 76%. Hasil pengujian untuk data baru terlihat kurang baik, hal ini dikarenakan kurang bervariasinya data latih yang digunakan dalam pelatihan. Sehingga pada saat sistem diuji dengan data baru yang belum dikenali, diperoleh nilai akurasi yang rendah. Selain itu, naracoba yang digunakan tidak semua benar-benar dalam kondisi emosi yang diinginkan sehingga terdapat beberapa kalimat yang diucapkan tidak sesuai dengan emosi yang seharusnya. Sedangkan untuk rendahnya akurasi terhadap data latih, hal tersebut dapat disebabkan karena pada saat pelatihan nilai MSE yang diperoleh masih jauh dari

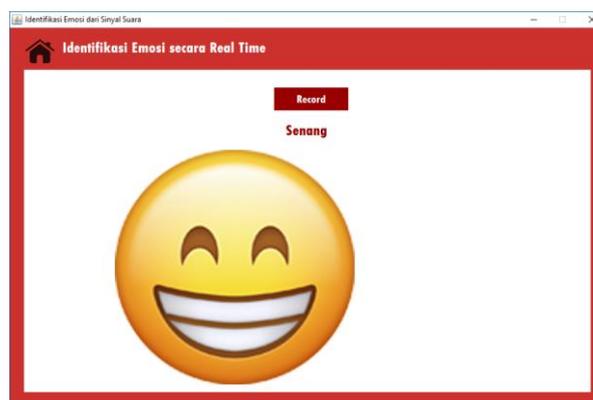
nilai minimum *error* yang ditetapkan sehingga kemungkinan masih banyak bobot yang belum dikoreksi dengan benar dan masih jauh dari nilai target.

Hasil pengujian realtime dengan 30 data masing-masing emosi memiliki 10 data dapat dilihat pada Tabel IV.

TABEL IV. PENGUJIAN DATA REAL TIME

	Marah	Netral	Senang
Marah	8	0	2
Netral	0	7	3
Senang	3	1	6

Hasil dari pengujian data secara *real time* pada Tabel IV menunjukkan bahwa pengujian dari 10 data emosi marah yang dapat dikenali 8 dan 2 lainnya dikenali dengan senang, pada 10 data emosi netral dapat dikenali sebagai netral 7 dan 3 lainnya untuk emosi senang, dan 10 data emosi senang dapat dikenali sebanyak 6 data untuk emosi senang, 3 emosi marah dan 1 emosi netral. Hasil dari pengujian secara *real time* dapat dilihat pada Gambar 2.



Gambar 2. Tampilan Antarmuka Program Identifikasi Emosi Real Time

IV. KESIMPULAN

Penelitian ini telah membangun sebuah sistem identifikasi emosi berbasis desktop dengan menggunakan Bahasa pemrograman Java dan Netbeans sebagai tools. Media penyimpanan data yang digunakan adalah SQLite dari sinyal suara menggunakan Linear Predictive Coding untuk ekstraksi sinyal suara dan menggunakan Backpropagation untuk pelatihan dan pengujian yang sebelumnya dilakukan tahapan pra proses untuk mengurangi *noise* dari data masukan suara, kemudian Frame Blocking untuk mengurangi pergeseran artikulasi, dan Windowing untuk mengatasi terjadinya kebocoran spectral atau *aliasing*. Backpropagation mempunyai kesederhanaan dalam generalisasi sehingga diperlukan untuk mempercepat proses komputasi. Hasil dari pengujian menggunakan maksimum epoch 10000, minimum learning rate 0.005, α 0.1 dan pengurangan α 0.01 dari 150 data uji yang dilatih menghasilkan akurasi sebesar 91% dengan hasil pengenalan 126 data yang dikenali dan 24 data tidak dikenali. Sedangkan dalam pengujian data baru dihasilkan akurasi sebesar 76% dengan hasil pengenalan 68 data yang dikenali dan 82 lainnya tidak dikenali.

REFERENSI

- [1] A. Fathoni and A. T. Haryono, "Pengaruh Citra Perusahaan Dan Kualitas Pelayanan Terhadap Loyalitas Pelanggan Dengan Kepuasan Pelanggan Sebagai Variabel Intervening Pada Pt Pelabuhan Indonesia Iii Semarang," *J. Manage.*, vol. 2, no. 2, 2016.
- [2] D. Morrison, R. Wang, and L. C. De Silva, "Ensemble methods for spoken emotion recognition in call-centres," vol. 49, pp. 98–112, 2007.
- [3] A. Tickle, S. Raghu, and M. Elshaw, "Emotional recognition from the speech signal for a virtual education agent," *J. Phys. Conf. Ser.*, vol. 450, no. 1, 2013.
- [4] A. Chamoli, A. Semwal, and N. Saikia, "Detection Of Emotion In Analysis Of Speech Using Linear Predictive Coding Techniques (L.P.C)," pp. 1–4, 2017.
- [5] J. Wu and B. Lin, "Speaker identification based on the frame linear predictive coding spectrum technique," *Expert Syst. Appl.*, vol. 36, no. 4, pp. 8056–8063, 2009.
- [6] H. M. Fayek, M. Lech, and L. Cavedon, "Evaluating deep learning architectures for Speech Emotion Recognition," *Neural Networks*, vol. 92, pp. 60–68, 2017.
- [7] C. F. Aliferis and I. Tsamardinos, "Feed-Forward Artificial Neural Networks," pp. 131–196, 2004.
- [8] D. Kumar, "Voice Stress Analysis using Linear Predictive Coding in MATLAB," vol. 4, no. 7, pp. 210–213, 2016.
- [9] Faradiba, "Pengenalan Pola Sinyal Suara Manusia Menggunakan Metode Back Propagation Neural Network," *EduMatSains*, vol. 2, no. 2, 2017.
- [10] R. Magdlena and L. Novamizanti, "Simulasi dan Analisis Deteksi Emosi Manusia dari Sinyal Suara Percakapan Berbasis Discrete Wavelet Transform dan Linear Predictive Coding," vol. 32, no. 11, pp. 1665–1670, 2014.
- [11] Y. Pan, P. Shen, and L. Shen, "Speech emotion recognition using support vector machine," *Int. J. Smart Home*, vol. 6, no. 2, pp. 101–108, 2012.
- [12] S. Wu, T. H. Falk, and W.-Y. Chan, "Automatic speech emotion recognition using modulation spectral features," *Speech Commun.*, vol. 53, no. 5, pp. 768–785, 2011.
- [13] Z. Liu, M. Wu, W. Cao, J. Mao, J. Xu, and G. Tan, "Neurocomputing Speech emotion recognition based on feature selection and extreme learning machine decision tree R," *Neurocomputing*, vol. 273, pp. 271–280, 2018.
- [14] A. Suroso, Y. Fitri, and S. Fitria, "Aplikasi Pengenalan Ucapan Dengan Ekstraksi Ciri Mel- Frequency Cepstrum Coefficients (MFCC) dan Jaringan Syaraf Tiruan (JST) Propagasi Balik Untuk Buka dan Tutup Pintu," vol. 1, no. 2, pp. 121–132, 2015.
- [15] Z. Ramadhan and S. N. Endah, "Perintah Suara Berbahasa Indonesia untuk Membuka dan Menutup Aplikasi dalam Sistem Operasi Windows Menggunakan Metode Mel Frequency Cepstrum Coefficient dan Metode Backpropagation," in *Seminar Nasional Ilmu Komputer (SNIK 2016)*, 2016, no. Snik, pp. 33–41.
- [16] S. Amalia, "Pengenalan Digit 0 Sampai Digit 9 Menggunakan Ekstraksi Ciri MFCC dan Jaringan Syaraf Tiruan Backpropagation," *Tek. Elektro ITP*, vol. 6, no. 1, pp. 1–14, 2011.
- [17] E. Mansour, M. S. Sayed, A. M. Moselhy, and A. A. Abdelnaiem, "LPC and MFCC Performance Evaluation with Artificial Neural Network for Spoken Language Identification," *Int. J. Signal Process. Image Process. Pattern Recognit.*, vol. 6, no. 3, pp. 55–66, 2013.
- [18] A. H. Md, M. R. Md, U. Kumar Prodhana, and F. K. Md, "Implementation of back-propagation neural Network for isolated bangla speech recognition," *Int. J. Inf. Sci. Tech.*, vol. 3, no. 4, pp. 1–9, 2013.
- [19] M. Singh and K. Verma, "Speech recognition using neural networks," *Int. J. Technol. Eng. Syst.*, vol. 2, no. 1, pp. 108–110, 2011.