

PENGENALAN BENTUK BENDA BERDASARKAN SINYAL SUARA DENGAN TRANSDUCER MIKROFON DAN TEKNOLOGI KINECT

Siska Aulia*, Lifwarda dan Yustini

Program Studi Teknik Telekomunikasi, Jurusan Teknik Elektro, Politeknik Negeri Padang

*Corresponding author, e-mail: siska.auliaa@gmail.com

Abstrak— Pengolahan suara atau pengenalan kata berkembang pesat sehingga dapat digunakan untuk berbagai aplikasi seperti menggerakkan suatu sistem atau kontrol gerak dan media pembelajaran berbasis multimedia. Implementasi pengenalan suara dan deteksi citra pada penelitian ini menggunakan *transducer mikrofon* dan teknologi *kinect*. Penelitian ini bertujuan untuk menghasilkan sistem yang dapat mengidentifikasi dan mengenali suatu objek dengan perintah kata, seperti lingkaran, segitiga, segiempat dan segibanyak. Dalam pengolahan suara dilakukan ekstraksi ciri suara dengan *Mel-Frequency Cepstrum Coefficient* (MFCC). Pemodelan kata dilakukan dengan menggunakan pemodelan statistik yaitu *Hidden Markov Model* (HMM). HMM mampu memberikan mekanisme yang efisien untuk memodelkan secara statistik keragaman dalam ucapan atau kata. Pengambilan data sampel dengan *transducer mikrofon* secara *offline* dan *online*. Pada penelitian ini pencocokan pola kata melalui proses pelatihan dan pengujian kata. Keluaran sistem ini berupa kata yang dikenali berdasarkan probabilitas tertinggi dan menampilkan bentuk benda berdasarkan kata yang dikenali. Prosesnya setelah kata dikenali, sistem akan tracking citra benda berdasarkan bentuk benda kemudian menampilkan bentuk benda yaitu lingkaran, segitiga, segiempat dan segibanyak. Hasil pengujian dengan transducer mikrofon, untuk sumber terlatih 85%, sumber tidak terlatih 81,5%, dan pengujian dengan Kinect sumber tidak terlatih 84% sehingga sistem pengenalan kata dapat diimplementasikan dengan teknologi Kinect.

Kata Kunci : *speech processing, HMM, MFCC, kinect*

Abstract— Voice processing or speech recognition is growing rapidly hence it can be used for various applications such as moving a system or motion control and multimedia-based learning media. Implementation of speech recognition and image detection in this study using microphone transducer and kinect technology. This study aims to produce a system that can identify and recognize an object with word commands, such as circles, triangles, rectangles and many. In sound processing, sound feature extraction is carried out with *Mel-Frequency Cepstrum Coefficient* (MFCC). Word modeling was done using statistical modeling, namely the *Hidden Markov Model* (HMM). HMM is able to provide an efficient mechanism for statistically modeling diversity in words or words. Data were collected with offline and online microphone transducers. This study matches the pattern of words through training and testing process. The output of this system is a recognizable word based on the highest probability and displaying the object shape based on the recognized word, namely circle, triangle and quadrilateral. Test results with microphone transducers, for 85% trained sources, 81.5% untrained sources, and 84% untrained Kinect source testing hence that word recognition systems can be implemented with Kinect technology.

Keywords : *speech processing, HMM, MFCC, Kinect*

1. Pendahuluan

Kemajuan teknologi semakin meningkat pada berbagai bidang. Bidang pengolahan suara dan bidang pengolahan citra diimplementasikan dalam perangkat lunak maupun perangkat keras. Teknologi Kinect merupakan suatu perangkat keras yang memanfaatkan bidang pengolahan suara dan citra. Kinect dikembangkan oleh Microsoft sebagai sensor gerak untuk video game

dalam operasi windows. Pada penelitian ini teknologi Kinect digunakan untuk mengenali bentuk benda berdasarkan suara atau kata yang diucapkan dan pengolahan citra bentuk.

Pengenalan ucapan atau kata digunakan supaya manusia dan mesin dapat berkomunikasi dengan media suara. Penelitian menggunakan suara yang berperan dalam teknologi Human machine yaitu Automatic Speech Recognition (ASR) adalah suatu proses dimana computer

Received date 2018-10-15, Revised date 2018-10-31, Accepted date 2018-11-17

<https://doi.org/10.25077/jnte.v7n3.600.2018>

This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

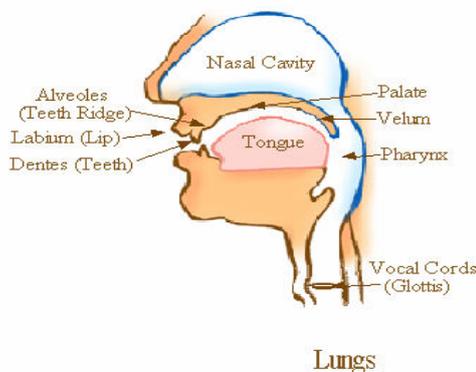
mengenal kata-kata yang diucapkan manusia [1]. Bidang pengolahan suara salah satunya dimanfaatkan sebagai media pembelajaran dalam mengucapkan kata. Penderita tunarungu merupakan pengguna yang membutuhkan media pembelajaran pengucapan kata[2,3]. Media pembelajaran yang memanfaatkan pengolahan citra yaitu berbasis multimedia. Media pembelajaran ini memudahkan guru dalam menyampaikan informasi ketika sulit disampaikan melalui perkataan.

Penelitian ini kombinasi pengolahan suara dan pengolahan citra. Penelitian ini bertujuan untuk menghasilkan sistem yang dapat mengidentifikasi dan mengenali suatu objek dengan perintah kata, seperti lingkaran, segitiga, segiempat dan segibanyak. Pencocokan pola kata dengan Hidden Markov Model dan deteksi benda berdasarkan bentuk dan ukuran benda. Output dari sistem ini dimanfaatkan sebagai media pembelajaran berbasis multimedia dan sebagai referensi modul pengolahan suara untuk teknologi Kinect.

2. Tinjauan Pustaka

2.1 Proses Pembentukan Sinyal Suara pada Manusia

Proses berbicara dapat terjadi jika manusia mampu mengeluarkan sinyal akustik yang kita sebut dengan suara. Suara manusia dihasilkan dengan adanya interaksi antara organ-organ pembentuk sinyal. Alat-alat pembentuk sinyal suara pada manusia dapat dilihat pada Gambar 1[1].



Gambar 1. Organ penghasil suara manusia [1].

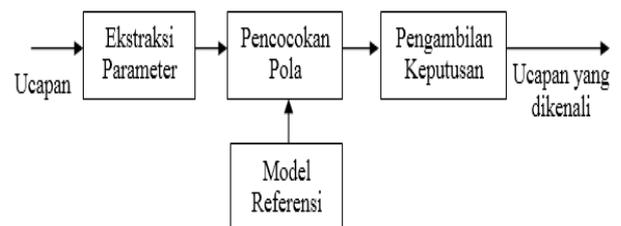
Suara yang dihasilkan pada organ-organ suara dibangkitkan karena adanya tekanan udara yang dihasilkan dari paru-paru yang kemudian dialirkan ke batang tenggorokan. Udara yang

mengalir melewati batang tenggorokan akan masuk ke *larynx*. Disana terdapat pita-pita suara sehingga ketika udara melewatinya akan bergetar sehingga akan menghasilkan pulsa-pulsa udara. Pulsa-pulsa udara ini akan masuk ke *vocal tract*. Pada bagian ini terjadi resonansi pulsa-pulsa udara yang kemudian membentuk pola-pola suara yang menghasilkan sinyal suara.

Teknologi pengenalan ucapan bertujuan menciptakan suatu cara yang efisien agar suatu mesin (komputer) mampu menerima informasi lewat perkataan serta beraksi dengan cepat dan tepat sesuai dengan informasi tersebut. Penelitian tentang pengenalan ucapan adalah bagian dari pengembangan teknologi komputer dengan kecerdasan buatan (*artificial intelligent*) yang dapat “mendengar”, ”mengerti”, dan “beraksi” sesuai dengan informasi perkataan yang diberikan [1].

2.2 Sistem Pengenalan Kata

Pengenalan kata atau ucapan merupakan upaya agar manusia dan mesin dapat berkomunikasi dengan media suara. Secara umum suatu sistem pengenalan kata terdiri atas dua proses utama. Proses pertama adalah ekstraksi parameter dan proses kedua adalah pencocokan pola. Proses ini dapat dilihat pada gambar 2 [1,5]:



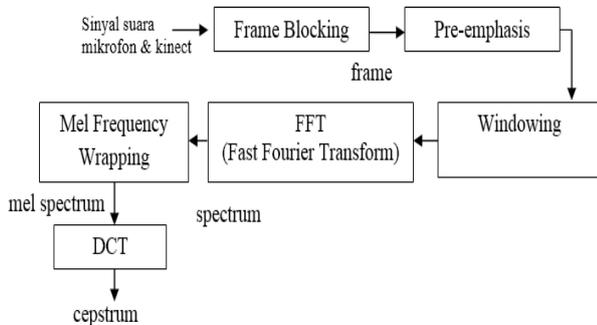
Gambar 2. Blok diagram sistem pengenalan kata

2.2.1 Ekstraksi Sinyal Suara

Banyak metoda yang digunakan dalam proses ekstraksi sinyal suara. Ekstraksi sinyal suara atau ekstraksi parameter digunakan untuk mendapatkan parameter dari kata yang diucapkan. Ekstraksi parameter ini dapat dilakukan dengan berbagai cara diantaranya dengan Linear Predictive Coding (LPC), Mel Frequency Cepstrum Coefficient (MFCC) dan lain-lain.

Pada penelitian ini digunakan *Mel-Frequency Cepstrum Coefficient* yang bekerja berdasarkan prinsip sistem pendengaran manusia dalam skala mel, sehingga karakteristik parameter sinyal input yang dipilih sesuai dengan parameter yang

diinginkan. MFCC ini menggunakan *filter bank* berbentuk segitiga untuk menentukan koefisiennya, dengan frekuensi tengah berjarak linier terhadap frekuensi dibawah 1000 Hz dan logaritmis frekuensi diatasnya [5]. Dari metode ini didapatkan parameter yang mewakili karakteristik gelombang suara. Diagram blok MFCC (*Mel Frequency Cepstrum Coefficient*) seperti gambar 3.



Gambar 3. Blok diagram ekstraksi ciri MFCC[5]

2.3.2 Hidden Markov Model

Hidden Markov Model (HMM) merupakan suatu rantai Markov dimana setiap state akan membentuk observasi. Disini deret observasi diketahui untuk menduga deretan state yang tersembunyi. Sebagai contoh, state yang tersembunyi tersebut bisa saja merupakan sebuah kata atau fonem, dan observasi merupakan suatu sinyal akustik[1].

Hidden Markov Model memiliki 5 parameter yaitu[1,5]: N adalah jumlah state pada model. Tiap state diberi nama {1, 2, 3,..., N} dan tiap state pada saat t disimbolkan q_t . M adalah jumlah dari simbol observasi per state. Jumlah simbol observasi yang berbeda tiap state yang dilambangkan dengan $V = \{V_1, V_2, V_3, \dots, V_M\}$. Matrik probabilitas transisi antar state adalah $(A = a_{ij})$, dimana $a_{ij} = P[q_{t+1} = j | q_t = i]$, $1 \leq i, j \leq N$ Distribusi probabilitas *symbol Observation* pada sebuah state yaitu $B = \{b_j(k)\}$, dimana $b_j(k) = P[O_t = V_k | q_t = j]$, $1 \leq k \leq M$ *symbol distribution in state j*, $j = 1, 2, 3, \dots, N$. Probabilitas state awal, $\pi = \{\pi_i\}$ dimana $\pi_i = P(q_1 = S_i)$ $1 \leq i \leq N$

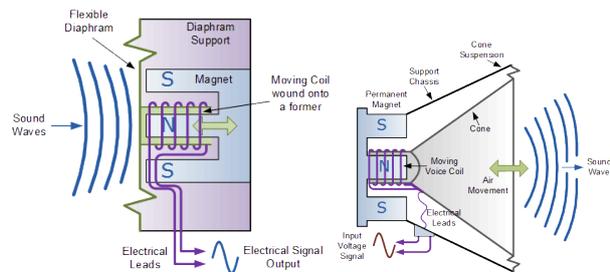
Pelatihan kata menggunakan HMM yaitu metoda Baum-Welch atau metoda *Expectation-Maximalization Method* (EM). Algoritma EM dirancang untuk menangani kasus dimana terdapat data yang hilang (data yang tidak lengkap)[1,5]. Dalam kasus Hidden Markov ini data yang hilang adalah deretan state.

Metoda EM memilih $\lambda = (A, B, \pi)$ yang memiliki $P(O|\lambda)$ yang maksimal dengan menggunakan prosedur iterasi. Hasil akhir dari pengenalan sinyal dengan HMM yaitu model sinyal, adalah membandingkan model sekarang dinyatakan sebagai $\lambda = (A, B, \pi)$ dan model re-estimasi dinyatakan sebagai $(\bar{\lambda} = \bar{A}, \bar{B}, \bar{\pi})$. Jika $P(O|\bar{\lambda}) > P(O|\lambda)$ maka terdapat model baru yaitu $\bar{\lambda}$. Model baru ini memiliki kemungkinan yang lebih besar sebagai model yang dihasilkan oleh deret observasi yang diamati. Iterasi akan berhenti sampai kondisi konvergen[1].

2.3 Transducer Mikrofon

Mikrofon berfungsi sebagai transducer yaitu dapat mengubah gelombang suara manusia menjadi sinyal listrik. Mikrofon selalu dihubungkan dengan alat penguat suara (*amplifier*), agar keluaran mikrofon dalam bentuk sinyal listrik yang masih lemah tersebut dapat diperkuat semaksimal mungkin sesuai kebutuhan dan hasilnya dapat didengar melalui *speaker*.

Mikrofon dinamis terdiri dari sebuah membran yang lentur, sebuah magnet dan sebuah lilitan sebagaimana yang ditunjukkan gambar 4 suara yang berupa tekanan udara yang bervariasi dari mulut akan mengenai membran fleksibel sehingga membran tersebut akan bergerak-gerak sesuai dengan tekanan udara dari suara. Dibelakang membran menempel lilitan, lilitan tersebut juga akan bergerak sesuai dengan gerakan membran. Gerakan lilitan akan memotong garis medan magnet sehingga akan dihasilkan energi listrik. Energi listrik yang dihasilkan sebanding dengan variasi tekanan udara yang dikeluarkan oleh mulut.



Gambar 4. Prinsip Kerja Mikrofon [12]

2.4 Kinect

Kinect merupakan sensor gerak yang dikembangkan oleh Microsoft untuk *video game*

console Xbox dan computer dengan sistem operasi Windows. Kinect dapat menginterpretasikan *gesture* secara spesifik, sehingga pengguna dapat melakukan kontrol tanpa menyentuh *game controller* [3]. Saat ini, Kinect sudah dilengkapi dengan *speech recognition* atau pendeteksian suara. Namun, bahasa yang dapat dideteksi masih meliputi bahasa Inggris, Perancis, Spanyol, Italia, dan Jepang. Sensor Kinect memiliki tiga bagian [3], seperti ditunjukkan gambar 5, yaitu: kamera RGB, sensor kedalaman dan mikrofon.

Untuk menjalankan fungsi sebagai sensor suara, Kinect memiliki *multi-array microphones* yang terdiri dari empat mikrofon untuk menyimpan suara dari pengguna. Mikrofon inilah yang nantinya akan menangkap suara untuk kemudian diproses lebih lanjut. Kinect memiliki *Analog to Digital Converter* (ADC)[8]. Dengan adanya ADC, maka tahap pengolahan sinyal suara yang masuk akan berkurang. Karena pada umumnya proses pengenalan suara dilakukan dengan mikrofon biasa dan menghasilkan sinyal analog.

Kinect SDK merupakan sebuah perangkat lunak yang digunakan untuk membantu pengembang perangkat lunak dalam mengembangkan aplikasi yang dikembangkan menggunakan sensor Kinect. Kinect SDK memiliki beberapa kemampuan mendengar baik secara keseluruhan maupun fitur [3], sebagai berikut:

a. Raw sensor stream

Akses *low level* dari sensor sensor pada Kinect yang terdiri dari kamera RGB, sensor kedalaman, dan mikrofon.

b. Skeletal tracking

Akses untuk mendeteksi gambar kerangka dari satu atau dua pemain yang ada pada jangkauan Kinect. Dengan adanya *skeletal tracking* ini Kinect telah dapat mendeteksi sendi-sendi pada kerangka manusia.

c. Advance audio capabilities

Memiliki integrasi dengan Windows *speech recognition* API. Di mana Kinect telah dapat mengenali beberapa bahasa, seperti Inggris, Jepang, Korea, Italia, Spanyol, dan Jerman. Selain itu Kinect memiliki kemampuan *Acoustic echo cancellation*, dan *noise suppression*. *Acoustic echo cancellation* (AEC) dapat menghilangkan *echo* yang muncul ketika pengguna mengucapkan suara. Sedangkan

noise suppression sendiri akan meningkatkan tingkat kejernihan suara[8].



Gambar 5. Konfigurasi Teknologi Kinect[8]

2.5 Geometri Bentuk Benda dan Warna

2.5.1 Geometri Bentuk Benda

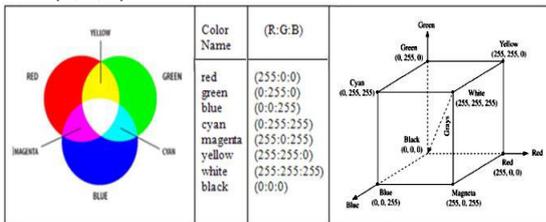
Lingkungan memberikan berbagai materi yang dapat kita pelajari. Pada lingkungan sekitar dapat dijumpai berbagai macam bentuk yang mempunyai nama, ciri, ukuran, fungsi, dan sebagainya. Definisi bentuk dalam kamus besar Bahasa Indonesia adalah rupa atau wujud yang ditampilkan. Bentuk adalah wujud, bangun atau rupa yang mempunyai pola dasar lingkaran, kotak, dan segitiga [9].

Pada pembelajaran anak di Taman Kanak-kanak (TK), materi tentang bentuk sangat perlu diajarkan sejak dini. Materi tentang bentuk yang diajarkan pada anak usia TK adalah bentuk-bentuk geometri [9]. Geometri adalah ilmu yang menjabarkan tentang bentuk dua dimensi (persegi, segitiga, lingkaran, segi empat, heksagon, trapesium) dan tiga dimensi (kubus, balok, tabung). Pemahaman bentuk geometri adalah kemampuan untuk mengingat, mengenal, mengidentifikasi, hingga mengaplikasikan berbagai wujud seperti bentuk dua dimensi dan tiga dimensi melalui informasi yang konkret. Adapun pengertian pemahaman bentuk geometri pada penelitian ini diartikan sebagai kemampuan menunjuk dan mencari 4 benda yang berbentuk geometri (segiempat, segitiga, lingkaran, segibanyak) yang mempunyai ciri-ciri tertentu. Tampilan bentuk benda berdasarkan kata atau suara yang dikenali [9].

Salah satu jenis bentuk geometri yaitu bangun datar. Bangun datar adalah bangun yang rata dan mempunyai dua dimensi yaitu panjang dan lebar tetapi tidak mempunyai tinggi dan tebal [9]. Bangun datar ditinjau dari segi sisinya dapat digolongkan menjadi dua, yakni bangun datar bersisi lengkung (lingkaran dan elips) dan lurus (segitiga, segiempat, segilima, segienam).

2.5.2 Ruang Warna RGB

Sebuah jenis warna, dapat kita bayangkan sebagai sebuah vektor di ruang dimensi-3. Yang biasanya dipakai dalam matematika, koordinat-ujungnya dinyatakan dalam bentuk tiga bilangan, yaitu komponen-x, komponen-y dan komponen-z. Misalnya, sebuah vektor dituliskan sebagai $r = (x,y,z)$. Untuk warna, komponen-komponen tersebut digantikan oleh komponen R(Red), G(Green), B(Blue). Jadi, sebuah jenis warna ditulis misalnya sebagai: **warna** = RGB(30, 75, 255). Ruang wana RGB dapat dilihat pada gambar 6. Dalam ungkapan ini warna putih = RGB(255,255,255), sedangkan warna **hitam** = RGB(0,0,0).



Gambar 6. Ruang warna RGB[1]

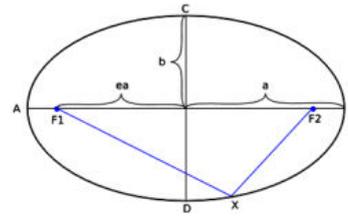
2.5.3 Ekstraksi ciri citra

Ekstraksi ciri citra adalah tahapan mengekstrak ciri informasi dari objek didalam citra yang ingin dibedakan atau dikenali dengan objek lainnya. Ciri yang telah diekstrak kemudian digunakan sebagai parameter nilai masukan untuk membedakan antara objek satu dengan lainnya pada tahapan identifikasi atau klasifikasi.

1. Ekstraksi Ciri Bentuk

Untuk membedakan bentuk objek satu dengan objek lainnya, dapat menggunakan parameter yang disebut dengan 'eccentricity'. *Eccentricity* merupakan nilai perbandingan antara jarak *foci ellips minor* dengan *foci ellips mayor* suatu objek. *Eccentricity* memiliki rentang nilai antara 0 hingga 1. Objek yang berbentuk memanjang/mendekati bentuk garis lurus, nilai *eccentricity*-nya mendekati angka 1, sedangkan objek yang berbentuk bulat/lingkaran, nilai *eccentricity*-nya mendekati angka 0. Penghitungan *eccentricity* diilustrasikan pada gambar 7. Persamaan *Eccentricity*:

$$e = \sqrt{1 - \frac{b^2}{a^2}} \quad (1)$$

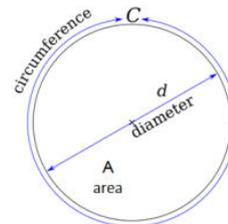


Gambar 7. Ilustrasi Penghitungan eccentricity

Parameter lainnya yang dapat digunakan untuk membedakan bentuk suatu objek yaitu 'metric'. *Metric* merupakan nilai perbandingan antara luas dan keliling objek. *Metric* memiliki rentang nilai antara 0 hingga 1. Objek yang berbentuk memanjang/mendekati bentuk garis lurus, nilai *metric*-nya mendekati angka 0, sedangkan objek yang berbentuk bulat/lingkaran, nilai *metric*-nya mendekati angka 1. Penghitungan *metric* diilustrasikan pada gambar 8. Persamaan perhitungan *metric*:

$$M = \frac{4\pi A}{C^2}$$

(2)



Gambar 8. Ilustrasi Penghitungan metric

3. Metode Penelitian

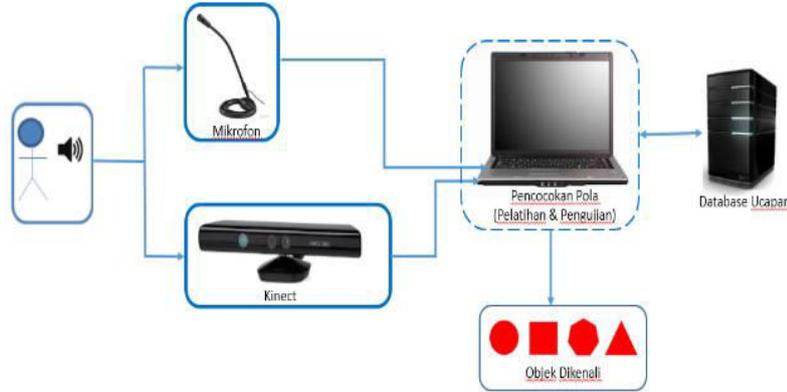
Sistem dan simulasi ini bertujuan untuk mengimplementasikan pengenalan bentuk benda berdasarkan kata dikenali didalam *Personal Computer* menggunakan transducer mikrofon dan teknologi Kinect. Pada gambar 9 Blok diagram implementasi pengenalan bentuk benda dan gambar 10 blok diagram sistem pengenalan kata untuk identifikasi bentuk benda.

Berdasarkan gambar 10 proses implementasi pengenalan bentuk benda berdasarkan suara dengan Mikrofon dan Kinect dijelaskan sebagai berikut. Pengenalan kata dimulai dengan mengucapkan kata "lingkaran", "segitiga", "segiempat" dan "segibanyak" menggunakan transducer mikrofon dan Kinect. Sinyal suara yang dirubah menjadi sinyal akustik melalui transducer mikrofon dan Kinect merupakan sinyal analog. Kemudian dirubah menjadi sinyal digital

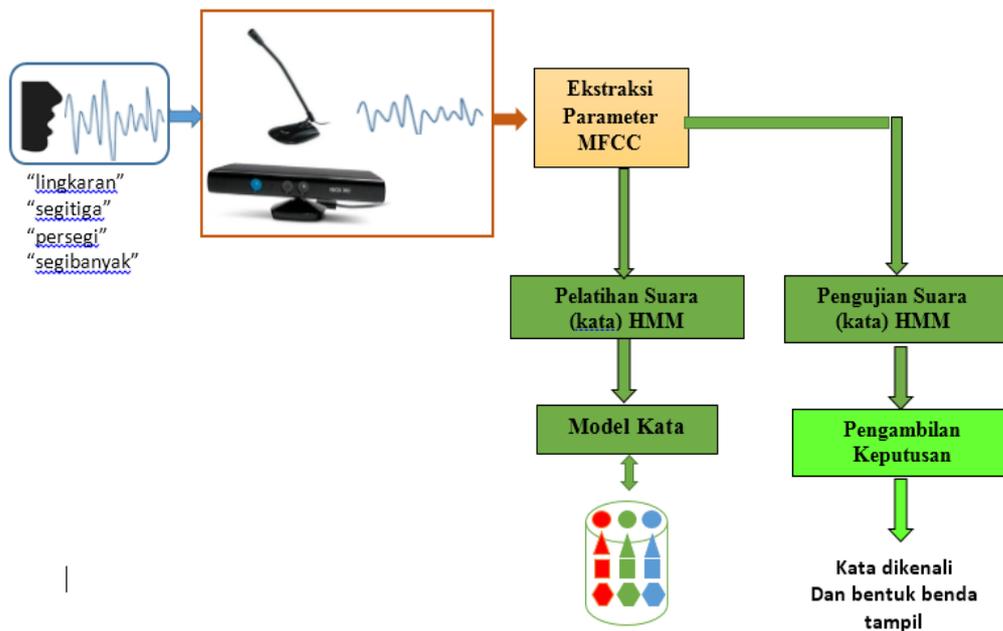
melewati *sound card* komputer dan disimpan dikomputer. Perangkat lunak pencocokan pola kata (pelatihan dan pengujian) menggunakan HMM dan pengolahan citra bentuk benda.

Proses pengenalan kata atau pencocokan pola kata dengan HMM dijelaskan sebagai berikut. Pertama melakukan ekstraksi parameter atau ciri dengan MFCC. Outputnya merupakan model

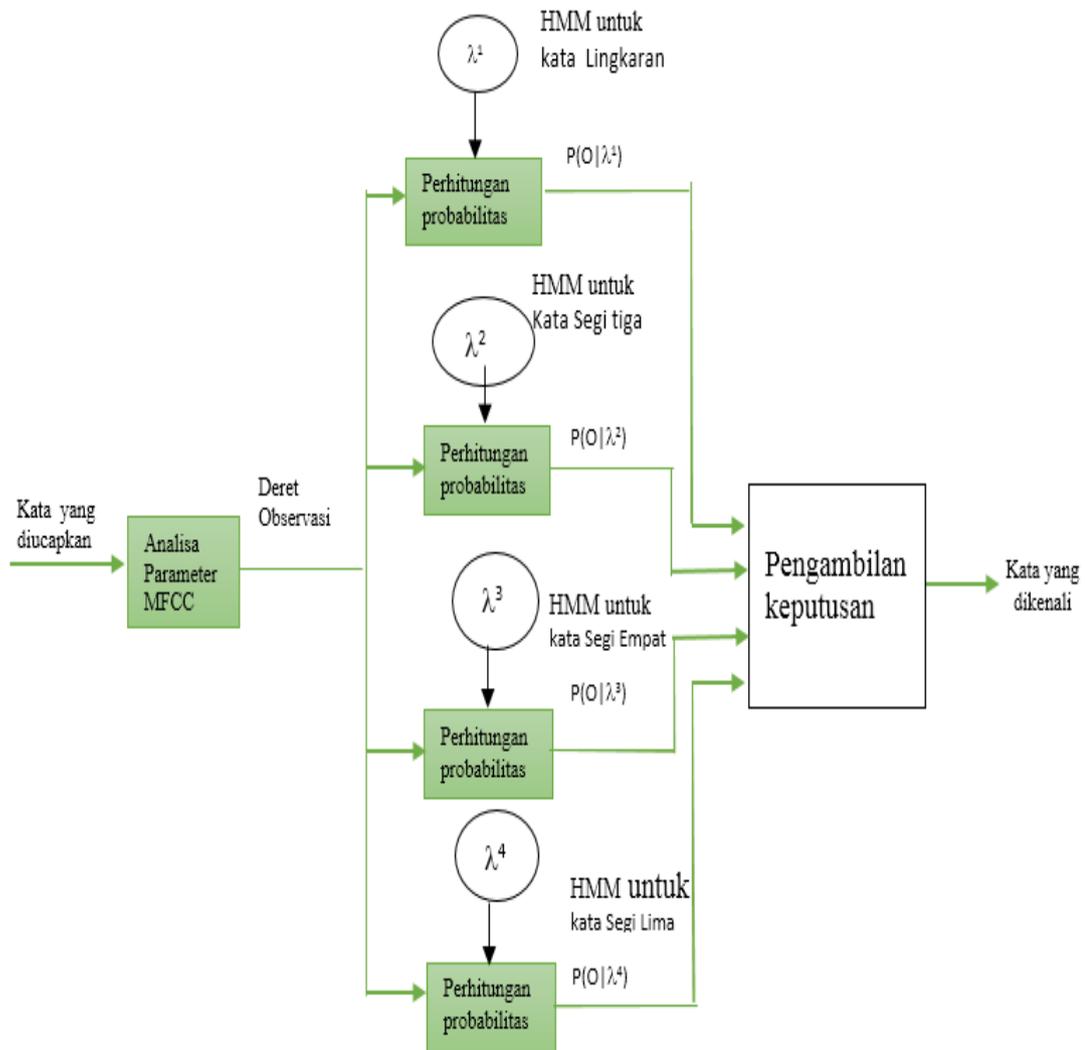
akustik yang berisikan vector parameter. Pelatihan kata dengan Hidden Markov Model. Proses pelatihan ini akan membentuk model kata. Model kata ini disimpan pada database. Pengujian kata menggunakan HMM tipe Isolated Word dapat dilihat pada gambar 11.



Gambar 9. Blok diagram implementasi pengenalan bentuk benda berdasarkan suara dengan mikrofon dan kinect



Gambar 10. Sistem pengenalan kata dengan metode hidden markov model



Gambar 11. Blok diagram pengenalan kata dengan hmm tipe isolated word [1,6]

Proses pengujian sistem atau simulasi pengenalan bentuk benda berdasarkan suara sebagai berikut:

- a. Pertama proses pengenalan kata dimulai dengan mendeteksi awal dan akhir suara yang diucapkan sumber.
- b. Proses ekstraksi parameter untuk pelatihan dan pengujian dilakukan dengan cara yang sama.
- c. Sinyal akustik berupa vector parameter suara dilakukan pemodelan HMM, dan dibandingkan dengan model-model kata yang disimpan dalam database.
- d. Proses perbandingan ini akan menghasilkan suatu perbedaan ukuran kemiripan (kemungkinan atau probabilitas) dari masing-masing model kata. Kemudian

untuk pengambilan keputusan dipilih perbandingan yang memiliki probabilitas yang paling besar sebagai kata yang dikenali.

- e. Pemilihan bentuk benda berdasarkan kata dikenali dan ekstraksi citra berdasarkan bentuk. Pada proses ekstraksi citra menggunakan ruang warna RGB.

4. Hasil dan Pembahasan

Pada penelitian ini , proses pengujian sistem yang dibangun ada 2 yaitu pengenalan bentuk beda dengan mikrofon dan pengenalan bentuk beda dengan Teknologi Kinect. Sistem Pengenalan bentuk benda ini terbagi 2 yaitu secara offline dan online. Pada pengujian secara online, sumber langsung mengucapkan kata pada sistem yang dibuat dan langsung dikenali.

Sedangkan cara offline, suara direkam terlebih dahulu.

4.1. Pengujian dengan Transducer Mikrofon

Pada pengujian ini dilakukan secara offline dan online. Pengujian ini dilakukan terhadap sumber terlatih dan sumber tidak dilatih.

4.1.1 Pengujian dengan Transducer Mikrofon Secara Offline

Hasil simulasi ditunjukkan pada gambar 12 a, 12b, dan 12c. Gambar 12a adalah awal simulasi dimana diupload bentuk-bentuk benda yang akan dipilih. Gambar 12b adalah proses input suara yang menyebutkan bentuk benda yang ingin dipilih. Gambar 12c adalah hasil dari simulasi dimana sistem sudah berhasil memilih benda yang sesuai dengan input suara yang diberikan

Simulasi dilakukan untuk pengujian sumber suara terlatih dan tidak terlatih Pengujian pertama terhadap sumber terlatih dengan 10 kali pengucapan dapat dilihat pada tabel 1. Pengujian kedua terhadap 50 sumber tidak terlatih dengan satu kali pengucapan dapat dilihat pada tabel 2.

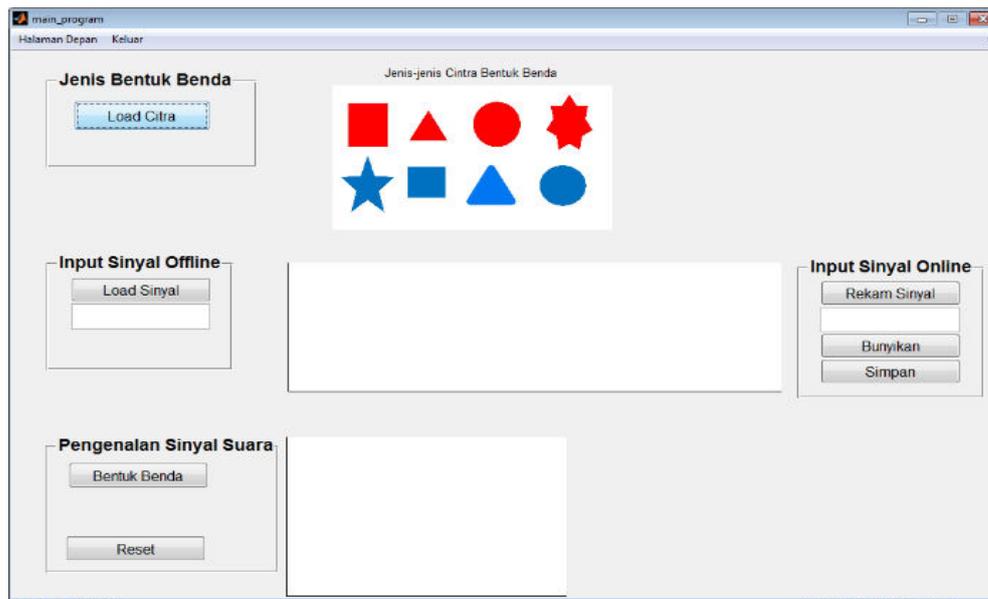
Tabel 1. Hasil pengujian pengenalan bentuk benda sumber terlatih

Kata yang diucapkan	Jumlah ucapan	Jumlah ucapan yang benar	Jumlah ucapan yang salah	Persen tase benar
Lingkaran	10	10	0	100%
Segitiga	10	8	2	80%
SegiEmpat	10	9	1	90%
SegiBanyak	10	7	3	70%
Rata-rata				85%

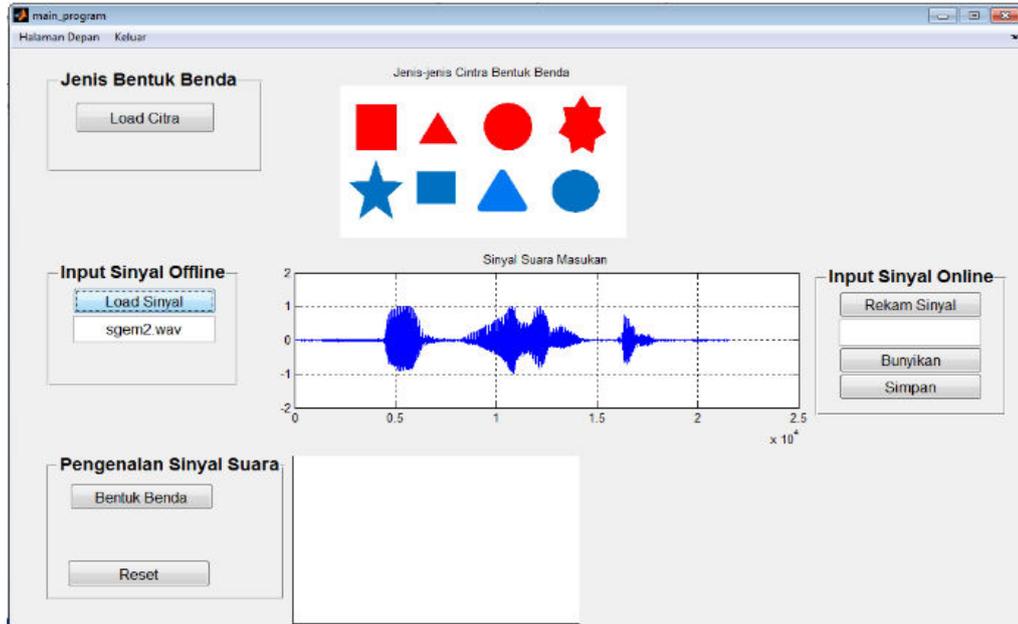
Berdasarkan hasil pada tabel 1, dari 40 ucapan yang diujikan, sistem mampu mengenali 34 ucapan dengan benar atau tingkat keakuratannya 85%.

Tabel 2 . Hasil pengujian pengenalan bentuk benda sumber tidak terlatih

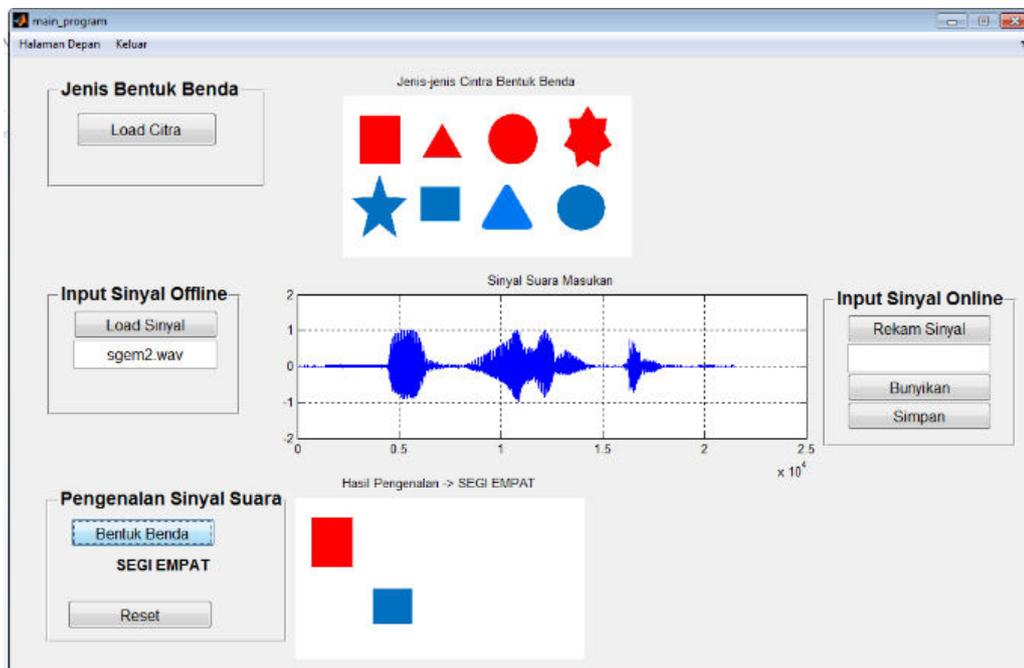
Kata yang diucapkan	Jumlah ucapan	Jumlah ucapan yang benar	Jumlah ucapan yang salah	Persen tase benar
Lingkaran	50	48	2	96%
Segi tiga	50	38	12	76%
Segi empat	50	42	8	84%
Segi banyak	50	35	5	70%
Rata-rata				81,5%



Gambar 12a. Simulasi upload citra ”jenis-jenis bentuk benda”



Gambar 12b. Simulasi input sinyal suara



Gambar 12c. Simulasi pengenalan bentuk benda berdasarkan sinyal suara dengan mikrofon untuk kata “segiempat”

Pada pengujian kedua dengan sumber tidak terlatih, sistem pengenalan ucapan yang digunakan adalah *single observation*. Pengucapan kata bersifat *isolated word* dimana deret observasi yang digunakan diambil dari satu orang sumber dengan satu kali pengucapan dan dibandingkan dengan model kata referensi

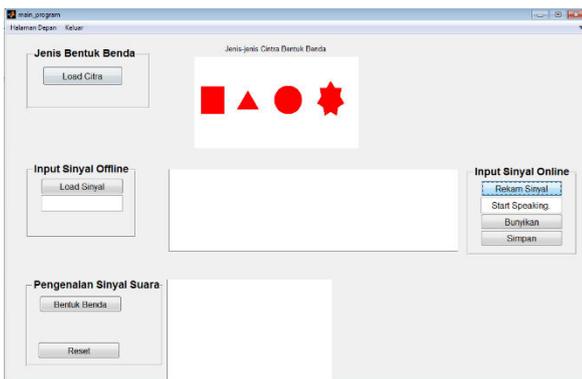
“lingkaran”, “segitiga”, “segiempat” dan “segibanyak”.

Tabel 2 pengujian sumber tidak terlatih dengan 50 orang sumber yang tidak dilatih dan masing-masing sumber mengucapkan masing-masing kata satu kali. Hasil yang diperoleh dari 200 ucapan yang diujikan maka terdapat 163

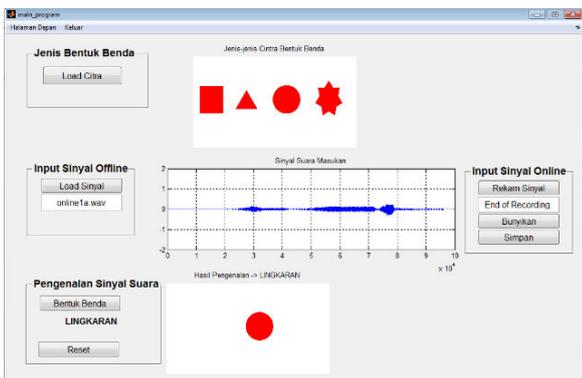
ucapan dapat dikenali sebagai kata yang sesuai. Berarti keakuratan sebesar 81,5%.

4.1.2 Pengujian Pengenalan Bentuk Benda dengan Mikrofon Secara Online

Pengujian pengenalan bentuk benda secara online yaitu pengambilan data langsung, dapat dilihat pada gambar 13a dan 13b. Gambar 13a menampilkan proses rekam suara, dimana sumber mengucapkan kata lingkaran. Sedangkan gambar 13b proses rekam telah selesai, dan menampilkan sinyal suara dan bentuk benda dikenali lingkaran.



Gambar 13a. Simulasi pengambilan sinyal suara secara langsung (on line)



Gambar 13b. Simulasi online dengan mikrofon untuk kata “lingkaran”

Pada pengujian secara online ini tingkat akurasi masih rendah, sehingga pengujian dilakukan pada sumber terlatih untuk masing-masing bentuk benda yaitu lingkaran, segitiga, segiempat dan segibanyak. Sistem sudah berhasil mengidentifikasi bentuk benda dengan pengucapan langsung.

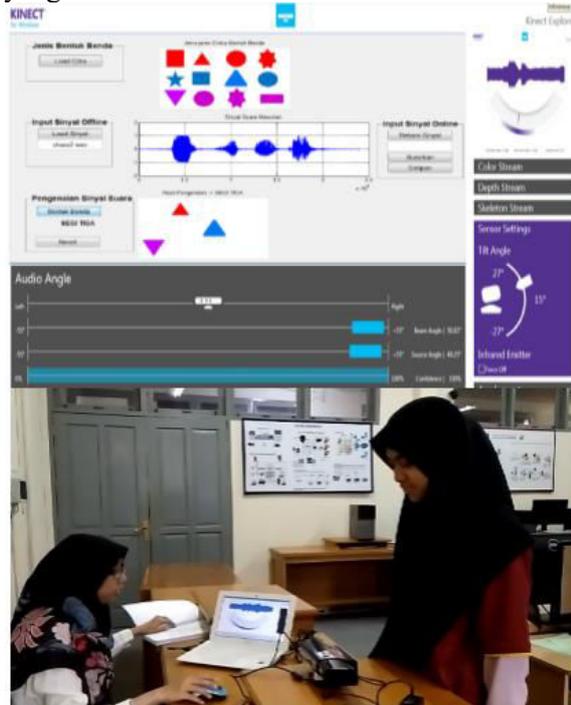
4.2 Pengujian dengan Teknologi Kinect

Pada pengujian ini dilakukan secara langsung dengan sumber tidak dilatih. Pengujian dilakukan di Laboratorium Multimedia Telekomunikasi Politeknik Negeri Padang dengan 25 mahasiswa. Hasil pengujian dapat dilihat pada tabel 3 dan prosesnya pada gambar 14.

Tabel 3 . Hasil pengujian pengenalan bentuk benda dengan kinect sumber tidak terlatih

Kata yang diucapkan	Jumlah ucapan	Jumlah ucapan yang benar	Jumlah ucapan yang salah	Persentase benar
Lingkaran	25	23	2	92%
Segi tiga	25	21	3	84%
Segi Empat	25	21	3	84%
Segi Banyak	25	19	5	76%
Rata-rata				84%

Pada pengujian ketiga dengan kinect dan berdasarkan tabel 3 terdapat 25 orang sumber yang tidak dilatih dan masing-masing sumber mengucapkan masing-masing kata satu kali dengan kondisi tidak bergerak. Hasil yang diperoleh dari 100 ucapan yang diujikan maka terdapat 84 ucapan dapat dikenali sebagai kata yang sesuai. Berarti keakuratan sebesar 84%.



Gambar 14. Implementasi pengenalan bentuk benda berdasarkan sinyal suara dengan kinect untuk kata “segitiga”

Pada tulisan ini, pengujian pengenalan bentuk benda telah berhasil dilakukan dengan 2 teknologi yaitu transducer mikrofon dan teknologi Kinect. Hasil pengujian identifikasi atau pengenalan bentuk benda ada 2 yaitu perintah kata yang dikenali (menggunakan pengolahan suara) dan *tracking* objek (menggunakan pengolahan citra). Keberhasilan akurasi kata dikenali dapat dilihat pada tabel 1, tabel 2 dan tabel 3. Sedangkan *tracking* objek mengikuti kata yang dikenali dan dilanjutkan ekstraksi citra bentuk. Jika pengucapan kata segitiga, tetapi yang dikenali lingkaran maka bentuk benda yang akan tampil lingkaran. *Tracking* objek diawali dengan mengkonversi ruang warna citra RGB menjadi grayscale, dimana digunakan dalam ekstraksi citra berbasis aturan (*rule based*) sederhana berdasarkan bentuk benda. Pengujian bentuk benda setelah kata dikenali tidak hanya 4 jenis bentuk benda seperti gambar 13a, akan tetapi dilakukan dengan kombinasi warna dan bentuk atau posisi lain dari masing-masing bentuk benda. Pengujian kombinasi warna, bentuk dan posisi dari bentuk benda dapat dilihat pada gambar 12c dan 14.

5. Kesimpulan

Hasil penelitian ini menunjukkan bahwa implementasi pengenalan suara/kata berhasil dilakukan dengan menggunakan transducer mikrofon dan teknologi kinect secara offline dan online. Pencocokan pola kata dengan HMM berdasarkan probabilitas yang paling tinggi. Keberhasilan sistem pengujian dengan alat mikrofon adalah 85%, dan 81.5% berturut-turut untuk sumber terlatih dan sumber tidak terlatih sementara implementasi dengan Kinect untuk sumber tidak terlatih adalah 84%. Terlihat bahwa teknologi Kinect memiliki tingkat keberhasilan yang lebih baik dalam mengenali bentuk berdasarkan suara.

Ucapan Terima Kasih

Penelitian ini didanai dalam kegiatan penelitian dosen Politeknik Negeri Padang yaitu Penelitian Dana DIPA Politeknik Negeri Padang No. 238/PL9.1.4/LT/2018.

Daftar Pustaka

[1] Aulia, Siska, "Implementasi Pengenalan Kata Dengan Metode Mel Frequency

Cepstrum Coefficient Dan Hidden Markov Model Untuk Mengontrol Gerak Robot Mobil Penjejak Identifikasi Warna", Tugas Akhir. Padang: Teknik Elektro Universitas Andalas. 2011.

- [2] Hardiyanti, Margareta, "Rancang Bangun Aplikasi Pembelajaran Pengucapan bagi Penderita Tunarungu Menggunakan Teknologi Kinect", Institut Teknologi Sepuluh Nopember, Surabaya, 2013.
- [3] Cahyarini Ratri, "Rancang Bangun Modul Pengenalan Suara Menggunakan Teknologi Kinect", *Jurnal Teknik Pomits* Vol. 2, No. 1, ISSN: 2337-3539. 2013.
- [4] Kinect, 2018. [Online]. Available: <http://en.wikipedia.org/wiki/Kinect>.
- [5] Fitrilina, Kurnia Rahmadi, Aulia Siska, "Pengenalan Ucapan Metoda MFCC-HMM untuk Perintah Gerak Robot Mobil Penjejak Identifikasi Warna", *Jurnal Nasional Teknik Elektro*, Vol 2 No.1 Maret 2013.
- [6] Rabiner, "A Tutorial on Hidden Markov Model and Selected Application in Speech Recognition", *Proceedings of the IEEE*, vol 77, No 2, 1989.
- [7] L. R. Rabiner and B-H Juang, *Fundamentals of Speech Recognition*, Prentice Hall, Englewood Cliffs, New Jersey, 1993, chapter 6.
- [8] S Aulia, Lifwardaa dan V. Veronica, "The Implementation of Speech Recognition Using Kinect Technology", *International Conference of Applied Science on Engineering, Business, Linguistics and Information Technology*, pp. 375-383, 2017.
- [9] A. Zunita, *Peningkatan Pemahaman Bentuk Geometri Melalui Pembelajaran Berbasis Multimedia Pada Anak Kelompok B TK KKLKMD Kuwon Bambanglipuro Bantul*, Yogyakarta: Fakultas Ilmu Pendidikan Universitas Negeri Yogyakarta. 2013.
- [10] P. S. Sawai, "Gesture & Speech Recognition using Kinect Device –A Review", *International Conference on*

- Science and Technology for Sustainable Development*, Kuala Lumpur, May, 2016 .
- [11] G. Archana , "Dynamic Hand Gesture Recognition using Hidden Markov Model by Microsoft Kinect Sensor", *International Journal of Computer Applications*, vol.150, no.5, September 2016.
- [12] https://www.electronicstutorials.ws/io/io_8.html
- [13] D. W. J. Stein, "Detection of Random and Sinusoidal Signals in Hidden Markov Noise", *Proceeding of the 30th IEEE Asilomar Conference on Signal Systems and Computers*, 464-468. 1997.
- [14] M. J. Landau, "Simulating Kinect Infrared and Depth Images", *IEEE Transactions On Cybernetics*, Vol. 46, No. 12, December 2016.
- [15] X. Chai, G Li, M. Zhou, "Sign Language Recognition and Translation with Kinect". *In Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, Shanghai, China, April 2013.

Biodata Penulis

Siska Aulia, menyelesaikan pendidikan S1 dan S2 di Universitas Andalas. Saat ini bekerja sebagai dosen di Jurusan Teknik Elektro, Politeknik Negeri Padang.

Lifwarda, menyelesaikan pendidikan S1 di ITS Surabaya dan S2 di UPI YPTK Padang. Saat ini bekerja sebagai dosen di Jurusan Teknik Elektro, Politeknik Negeri Padang.

Yustini, menyelesaikan pendidikan Diploma di Politeknik Unand, Diploma IV di Institut Teknologi Bandung dan S2 di ISTN Jakarta. Saat ini bekerja sebagai dosen di Jurusan Teknik Elektro, Politeknik Negeri Padang.