

Perangkat Saklar Aktivasi Melalui Pengenalan Suara Manusia (The Switching Activation Device by Recognizing of The Human Voice)

Dedi Sugiharto, Dewi Astuti, Mohammad Mujirudin, Harry Ramza

Program Studi Teknik Elektro, Fakultas Teknik
Universitas Muhammadiyah Prof.Dr.Hamka
Jalan Tanah Merdeka No. 6, Kp Rambutan, Jakarta13830
Telp : +62-218400941, Faks : +62-2187782739

Email.: dewiastuti0107@gmail.com, dedysugiharto12@gmail.com, mujirudin@uhamka.ac.id, hramza@uhamka.ac.id

Abstrak

Pengenalan suara atau Voice Recognition sebagai sebuah teknologi yang berkembang sangat cepat dan dinamis. Perintah suara atau pengenalan suara adalah suatu teknologi untuk mempermudah pekerjaan manusia dalam beraktivitas sehari – hari secara efisien. Pengenalan suara dapat membantu pengguna atau manusia mencocokkan suara yang telah validasi sebelumnya dan memverifikasi kesesuaian suara dengan suara pengguna untuk melakukan proses identifikasi suara. Tujuan dari penelitian ini adalah untuk mengetahui persentase keberhasilan suara dengan menggunakan perangkat hardware dan software atau aplikasi dengan menggunakan modul Voice Recognition V3 Elechouse. Dalam penelitian ini tiap manusia dapat menguji pengenalan suara dengan output 7 perintah suara.

Setiap suara manusia dengan modul ini menghasilkan maksimal 80 command suara. Dengan begitu 7 output suara akan diuji dengan menggunakan perintah suara secara langsung dan menganalisis dengan menggunakan aplikasi untuk mengetahui nilai pitch, frekuensi nada, kunci nada dan nilai ct dengan merekam tiap suara manusia tersebut, yang nantinya akan mengetahui identifikasi tiap suara yang diuji dan seberapa tepat serta sensitif dengan alat perancangan sistem pengenalan suara. Penelitian ini menggunakan jarak antara microphone dengan suara manusia dan speaker rekaman dengan jarak range 0-5 cm, 5-10 cm, 10-15 cm, 15-30 cm, dengan maksimal jarak 120 cm. Hasil penelitian menunjukkan bahwa dengan menggunakan perintah suara manusia secara langsung bahwa persentase keberhasilan terbaik pada jarak 0-5 cm dengan tingkat keberhasilan 95%, kemudian dengan analisis menggunakan software pada aplikasi tingkat keberhasilan lebih besar yaitu 96.95%, Jadi kesimpulannya Jika volume perintah suara lebih besar, maka persentase tingkat keberhasilan suara akan semakin tinggi, ketika volume suara perintah kecil dan menurun persentase tingkat suara akan semakin sedikit.

Kata Kunci : Pengenalan Suara, Analisis Identifikasi Suara, Voice Recognition V3 Elechouse

Abstract

Voice recognition is a rapid and dynamic technology. Voice command or voice recognition is a technology to facilitate human work in daily activities efficiently. The voice recognition helps the user or matches the voice that has been validated previously and it can verify the conformity of the voice with the user's voice to conduct the identification process. The aim of this study was to see the percentage of succession by using the hardware and software or the application by implementing the V3 Elechouse voice recognition module. In this study, every human being can recognize speech recognition by outputting 7 voice – commands with each human voice. Every human voice with this module producing an amount of 80 voice commands, so that 7 commands output will be tested using direct commands and analyzed using the application to determine the pitch value, tone frequency, key tone, and ct value by recording each human voice. Finally, it will be determined the sound identification that is tested and accurate also sensitive with the voice recognition system tool. This research uses the distance between the microphone with the human voice and the recorded speaker with a distance of 0 – 5 cm, 5 – 10 cm, 10 – 15 cm, 15 – 30 cm, with a maximum distance of 120 cm. The results showed that using human voice commands directly represented a distance of 0-5 cm with a 95% confidence level, and then by using the software analysis the application increased rapidly, namely 96.95%. In the end, it is true if the volume of voice commands is greater, then the proportion of the sound level will be higher when the volume of the command voice is small and the proportion of the voice level will decrease.

Keywords: Voice Recognition, sound identification analysis, Voice Recognition V3 Elechouse

Received 22 Januari 2021; Accepted 18 Februari 2021

1. Pendahuluan

Perintah suara atau pengenalan suara (*voice recognition*) adalah suatu teknologi untuk mem permudah pekerjaan manusia dalam beraktivitas sehari – hari , dengan adanya teknologi pekerjaan manusia yang sebelumnya dikerjakan

dalam waktu yang lama dan membutuhkan banyak tenaga, maka pekerjaan ataupun aktivitas kita menjadi lebih cepat dan dapat menghemat tenaga kita, khususnya teknologi pengenalan suara (*voice recognition*) ini [1].

Voice Recognition atau pengenalan suara ini yaitu suatu teknik pengenalan suara yang fungsinya mengubah sinyal analog menjadi data digital yang akan di kontrol dengan sistem untuk mendeteksi perintah suara yang diucapkan. Pengenalan ucapan yaitu terdiri dari proses pemasukan informasi berupa ucapan manusia ke dalam komputer dan pengenalan dari komputer atas ucapan manusia itu sendiri [2].

Penelitian ini bertujuan untuk membuat sistem yang dapat mengenali suara dalam bentuk kalimat agar kedepannya bisa digunakan dalam teknologi listrik. Proses pengolahan suarapun perlu melewati beberapa proses seperti: sampling, ekstraksi dan pembelajaran. Dengan proses ekstraksi suatu sinyal suara dapat diketahui karakteristiknya [3]. Terdapat beberapa macam metode ekstraksi ciri yang biasa digunakan, tetapi pada penelitian kali ini menggunakan metode *Linear Predictive Coding (LPC)*. LPC digunakan karena sistem ekstraksinya yang mengadopsi sistem pendengaran manusia sebagai filter pengambilan informasi.

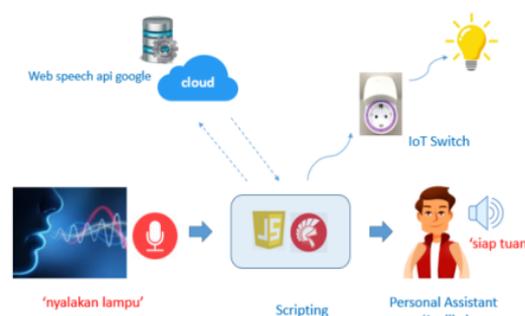
Oleh karena itu penulis membuat judul “Perancangan Sistem Pengenalan Suara Manusia Dan Analisis Ucapan Berbasis Mikrokontroler sebagai judul tugas akhir, sistem ini bermula dari pencuplikan sinyal analog menjadi sinyal digital, untuk proses ekstraksi parameter suara digunakan metode LPC (*Linier Predictive Coding*) untuk mendapatkan koefisien *cepstral*, menentukan jarak, respon sensitifitas, akurasi, serta frekuensi dan amplitudo dengan menggunakan perintah suara secara langsung maupun menggunakan aplikasi software dengan perintah rekaman (*record*).

2. Landasan Teori

2.1. Pengenalan Suara

Pengenalan suara adalah proses identifikasi suara berdasarkan kata yang diucapkan dengan melakukan sebuah konversi sinyal akustik, yang kemudian ditangkap oleh *audio device* (perangkat input suara). Pengenalan suara ini juga merupakan sistem yang digunakan untuk mengenali perintah kata dari manusia dan kemudian diterjemahkan menjadi suatu data yang dimengerti oleh komputer. Tetapi pelafalan kata intonasi dan kejelasan dalam mengenali perintah suara harus jelas, supaya komputer dapat menerima data suara dengan tepat [4]. Keuntungan dari sistem ini adalah pada kecepatan dan kemudahan dalam penggunaannya.

Kata – kata yang ditangkap dan dikenali bisa jadi sebagai hasil akhir, dan untuk sebuah aplikasi seperti *command & control*, penginputan data, dan persiapan dokumen. Parameter yang dibandingkan ialah tingkat penekanan suara yang kemudian akan dicocokkan dengan *template database* yang tersedia. Sedangkan sistem pengenalan suara berdasarkan orang yang berbicara dinamakan *speaker recognition*. Terdapat Gambar 1 Arsitektur Desain Perintah Suara :



Gambar 1. Arsitektur Desain Perintah Suara

2.2 Linier Predictive Coding (LPC)

LPC (*Linier Predictive Coding*) adalah salah satu metode parametrik yang digunakan untuk merepresentasikan sinyal. Umumnya LPC digunakan karena menyediakan pemodelan yang baik untuk sinyal suara, LPC dapat dengan mudah dan langsung diterapkan baik secara perangkat lunak maupun perangkat keras karena perhitungan matematis yang dilibatkan relatif lebih singkat dari metode-metode yang dikenal sebelumnya. Secara garis besar, cara kerja sistem pengenalan suara ini ialah mula-mula sinyal suara manusia yang diterima dengan menggunakan microphone (sinyal analog) dicuplik sehingga menjadi sinyal digital dengan bantuan sound card pada komputer.

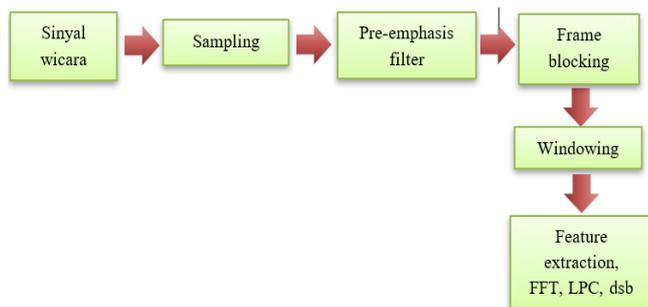
Sinyal digital hasil cuplikan ini terlebih dulu dinormalisasi kemudian diproses awal menggunakan metode LPC sehingga didapat beberapa koefisien LPC yang merupakan feature (ciri) dari suara pembicaraan. Kemudian koefisien LPC tersebut diproses dengan FFT (*Fast Fourier Transform*) untuk mendapatkan sinyal pada domain frekuensi. Hal ini bertujuan agar perbedaan antar pola kata yang satu dengan yang lain terlihat lebih jelas sehingga ekstraksi parameter sinyal memberikan hasil yang lebih baik. Hasil keluaran FFT ini akan di compare sebagai fungsi utama dari sistem untuk proses pengenalan. Hal ini dapat direpresntasikan menggunakan persamaan berikut :

$$s(n) \approx a_1s(n-1) + a_2s(n-2) \dots \dots + a_ps(n-p) \quad (1)$$

Dalam suatu teknik ekstraksi LPC ini ada beberapa langkah2 yang harus dilakukan yaitu :

1. Pre-emphasis Filter (penapisan).

Dalam pengolahan sinyal suara ini *preemphasis* filter diperlukan setelah proses sampling. Tujuan dari pemfilteran ini adalah untuk mendapatkan bentuk spectral frekuensi sinyal suara yang lebih halus. Dimana bentuk spectral yang relatif bernilai tinggi untuk daerah dan cenderung turun secara tajam untuk daerah frekuensi di atas 2000 Hz. Berikut terdapat gambar di bawah ini :



Gambar 2. Posisi filter pre-emphasis pada sistem pengolahan suara

2. Frame Blocking

Tahap ini merupakan sinyal kata yang telah diratakan dibagi sebanyak *T* frame dengan masing – masing frame memuat *N* cuplikan dan frame – frame yang berdekatan dipisahkan sejauh *M* cuplikan. Ukuran sampel tiap frame dihitung dari rata – rata cuplik tiap detik dengan waktu cuplik tiap detik dengan waktu cuplik tiap periode, umumnya digunakan 30 ms. Jika rata – rata cuplik adalah 8000Hz, maka jumlah sampel tiap frame adalah $3.10 \cdot 3 - \text{detik} \cdot 8000 \text{ sampel/detik}$, sama dengan 240 sampel.

Dengan menggunakan aturan *N* point DFT, bahwa jumlah sampel *N* harus bernilai 2^p , dimana $240 < 2^p$, sedangkan $240 < 28$, maka $N = 256$ [Rabiner,1993]. Untuk melakukan pembagian blok ini dapat mengacu pada *linier prediction*. Parameter – parameter yang biasanya digunakan antara lain ; *N* adalah jumlah sampel pada analisa *frameblocking*, dan *M* adalah jarak antara satu frame dengan frame selanjutnya.

3. Windowing (penjendelaan)

Windowing digunakan digunakan untuk mengurangi *discontinuitas* sinyal pada awal dan akhir frame. Untuk pengukuran nilai energi pada sinyal wicara kita harus melibatkan fungsi window. Hal ini karena dalam pengukuran energi sinyal wicara kita harus menyusunnya dalam frame –frame tertentu. Ini merupakan standar dalam teknologi *speech processing*, 28 sebab secara umum dalam pengolahan sinyal wicara kita terlibat dengan sinyal dengan durasi yang terlalu panjang, bila dihitung dalam total waktu pengukuran. Fenomena ini juga dikenal sebagai *short term speech signal energy*.

4. Analisis Auto Korelasi

Setiap frame dari sinyal setelah melalui proses *windowing*. Kemudian dilakukan analisis autokorelasi sebagai berikut :

$$r_1(m) = \sum_{n=0}^{N-1-m} \overline{X_1}(n) \overline{X_1}(n+m), = 0,1, \dots, p \dots (2)$$

2. 3. Mikrokontroler Arduino AtMega 328

Arduino Uno adalah papan mikrokontroler berdasarkan ATmega328 (datasheet). Dalam bahasa Itali "Uno" berarti satu, maka jangan heran jika peluncuran Arduino 1.0 diberi nama Uno. Arduino ini berisi semua yang diperlukan untuk mendukung mikrokontroler, untuk mengaktifkan cukup menghubungkannya ke komputer dengan kabel USB dengan adaptor AC-DC atau baterai.

Arduino merupakan kit elektronik atau papan rangkaian elektronik open source yang di dalamnya terdapat komponen utama yaitu sebuah chip mikrokontroler dengan jenis AVR dari perusahaan Atmel [16]. Berikut terdapat gambar seperti dibawah ini :



Gambar 3. Arduino Uno R3 Atmega 328

2. 4. Voice Recognition V3 Elechouse

Dalam Modul ini Voice Recognition Elechouse V3 suatu modul yang digunakan sebagai inputan untuk mematikan atau menghidupkan suatu alat listrik seperti lampu, kipas angin, tv dan lain lain, dengan menggunakan perintah suara secara langsung. Beberapa cara lain untuk menerapkan pengenalan suara dalam bentuk lain seperti, langsung dari ponsel android ke Alexa tahu Rashberry pi, menggunakan bluetooth atau beberapa teknologi lainnya.

Tetapi modul yang satu ini sangatlah efisien, karena dengan menggunakan perintah suara secara langsung, dengan berbicara langsung mengarahkan perintah dengan pengenalan suara untuk mematikan atau menghidupkan sebuah lampu contohnya. Modul ini merupakan pengenalan suara yang paling ringkas dan mudah dikendalikan. Berikut gambar Modul *Voice Recognition V3 Elechouse* :



Gambar 4. Voice Recognition V3 Elechouse

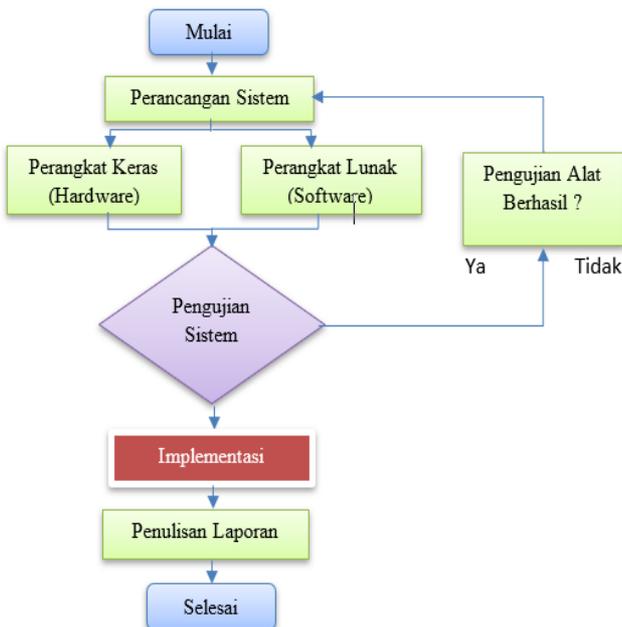
Terdapat beberapa cara itu mnghidupkan modul ini, yang pertama menggunakan port serial atau melalui pin *GPIO built-in*. Dalam module ini papan V3 memiliki kapasitas untuk menyimpan hingga 80 perintah suara masing – masing dengan durasi 1500 milidetik. Yang satu ini tidak akan mengubah perintah manusia menjadi teks tapi akan membandingkannya dengan kumpulan suara yang sudah terekam. Jadi secara teknis tidak ada hambatan bahasa untuk menggunakan produk ini. Kita bisa merekam perintah manusia dalam bahasa apapun atau secara harfiah suara apapun dapat direkam dan digunakan sebagai

perintah. Jadi kita perlu melatihnya terlebih dahulu sebelum kita membiarkannya mengenali perintah suara apa pun.

Dengan menggunakan modul dengan pin GPIO, modul akan mengirimkan output hanya 7 perintah dalam 80 percobaan pengetesan suara, Untuk metode ini modul ini dapat memilih dan memuat 7 perintah ke *recognizer* dan *recognizer* akan mengirimkan output ke masing – masing GPIO pin jika salah satu perintah suara ini dikenali. Perangkat ini bekerja pada kisaran tegangan masukan 4,5 – 5 volt, dan akan menghasilkan arus kurang dari 40 mA. Dalam modul ini dapat bekerja dengan akurasi 99% jika digunakan dalam kondisi ideal. Faktor pilihan mikrofon dan kebisingan di lingkungan memainkan peran penting dalam mempengaruhi kinerja modul.

3. Metodologi

Pada penelitian kali ini perancangan ini memanfaatkan metode *Linier Predictive Coding* (LPC), seperti yang telah dijelaskan sebelumnya, teknik ini yaitu salah satu metode yang digunakan untuk mempresentasikan sinyal, lebih tepatnya *Linier Predictive Coding* (LPC) di sini adalah sebagai pemfilteran suara yang dijadikan ciri khas untuk membedakan suara setiap manusia yang sudah ada di dalam *database* dengan suara yang di inputkan. Dapat dilihat bentuk *flowchart* (diagram alir) dibawah ini :



Gambar 5. Flow Chart Perancangan Sistem

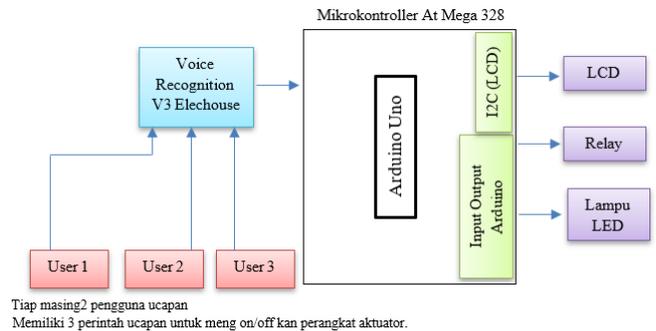
Berikut penjelasan diagram alir tahapan – tahapan metode perancangan sistem :

3. 1. Perancangan Sistem

Perancangan design dan rangkaian ini dimulai dengan perancangan sistem yang terbagi atas dua macam yaitu, perancangan sistem perangkat lunak (*software*) dan perangkat keras (*hardware*).

3. 2. Perangkat Keras (Hardware)

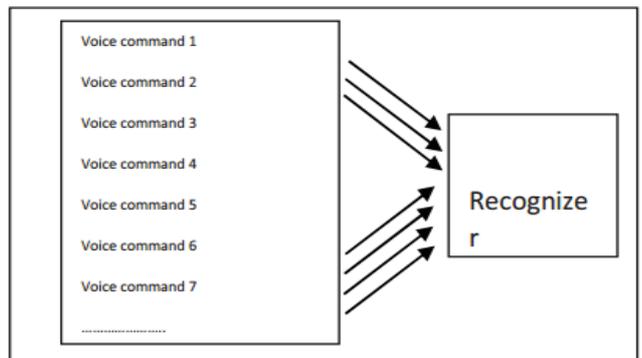
Perangkat keras (*hardware*) meliputi perancangan alat pengenalan suara, yang terdiri dari beberapa komponen penting, khususnya inputan mic yang digunakan untuk menjalankan sebuah perintah suara, dalam perancangan ini sistem mampu membaca perintah masukan ucapan dan mengenalinya sebagai perintah pengendali perangkat aktuator. Pada Perancangan alat ini dimulai dengan proses perancangan modul voice recognition v3 elechouse, dimana modul ini sebagai input suara untuk menangkap atau mendeteksi suara dari tiap personal manusia, karena setiap suara yang dihasilkan oleh manusia memiliki nilai pengenalan suara yang berbeda-beda. Kemudian Arduino sebagai mikrokontroller pemrosesan dan outputnya berupa LCD, perangkat aktuator (lampu), dan relay.



Gambar 6. Diagram Perancangan Perangkat Keras Identifikasi Suara

Berdasarkan blok diagram diatas cara kerja dari perancangan keras alat ini adalah Arduino Uno bekerja sebagai pusat pengendali atau kontrol, Voice recognition V3 Elechouse yaitu sebagai pengolah perintah ucapan dari tiap manusia, modul ini memiliki maksimal output 7 perintah ucapan dari 80 perintah suara yang diujikan.

Untuk outputnya yaitu Relay, LCD dengan I2C, 3 Buah Lampu LED. Tiap output lampu memiliki 3 perintah yang berbeda, yaitu untuk menyalakan Lampu, Kipas, Dan Televisi. Relay akan berperan sebagai switch on/off untuk mengaktifkan alat ketika pengenalan suara terdeteksi .



Gambar 7. Pengaturan Perintah Suara

Gambar diatas adalah contoh pengaturan pada perintah suara yang dimiliki oleh modul *voice recognition v3 elechouse*. Percobaan akan dilakukan sebanyak 3 kali dari tiap masing – masing pengguna suara. Setiap satu manusia memiliki 3 perintah ucapan percobaan, untuk menyalakan

Lampu, Kipas Angin, dan Televisi. Dari setiap 3 kali percobaan yaitu untuk menentukan nilai pengenalan dari setiap users tersebut, yang nantinya dapat teridentifikasi nilai pengenalan suara tiap users yang akan diuji. Berikut gambar skematik perancangan perangkat keras (*hardware*) :



Gambar 8. Skematik Alat Pengenalan Suara

Perintah Ucapan Suara yang akan dilakukan adalah :

1. Mendeteksi sinyal suara oleh computer, saat suara dari manusia memancarkan bunyi melalui mulut, kemudian komputer (*software*) akan membaca dari tiap perintah suara tersebut. Setiap pelafalan kata dan intonasi yang diucapkan akan dikirimkan melalui komputer (*software*) melalui udara.
2. Pengujian suara dilakukan oleh 1 – 3 orang laki – laki, dengan tiap manusia melakukan perintah untuk menghidupkan atau mematikan sebuah lampu, dengan maksimal output 7 perintah suara. dimana output perintah yang digunakan adalah :
 - Menghidupkan Lampu
 - Mematikan Lampu
 - Menghidupkan Kipas
 - Mematikan Kipas
 - Menghidupkan Televisi
 - Mematikan Televisi
 - Matikan semua perangkat secara bersamaan.
3. Tiap manusia melakukan percobaan sebanyak 10 – 40 perintah ucapan, maksimal ucapan tiap manusia 80 perintah ucapan. secara berulang untun mengetahui nilai pengenalan setiap manusia.
4. Setiap manusia akan diketahui nilai frekuensi, amplitudo, respon waktu aktif alat, tingkat keberhasilan, dan kegagalan, dan nilai pengenalan masing – masing manusia.
5. Kemudian sinyal suara akan memberikan informasi ke komputer (*software*), dan salah satu metode yang digunakan adalah dengan Linier Predictive Coding yaitu metode ini adalah pemodelan suara yang didasari bahwa suara kita atau suara manusia terbentuk dari suatu getaran sinyal yang bergetar melalui pada rongga tenggorokan. Dituliskan sebagai berikut :

$$s(n) \approx a_1 s(n-1) + a_2 s(n-2) \dots \dots \dots + a_p s(n-p) \quad (3)$$

6. Setelah semuanya selesai kemudian informasi yang telah diolah semuanya dikirimkan ke Arduino sebagai pusat kendali untuk melakukan perintah.
7. Kemudian semua sinyal akan diolah oleh mikrokontroler, setelah itu sinyal siap dipancarkan, dan outputnya berupa Lampu yang siap dijalankan.

3. 3. Perangkat Lunak (*Software*)

Perangkat lunak (*software*) yaitu meliputi pembuatan aplikasi pengenalan suara, untuk membuktikan dari tiap personal manusia memiliki nilai pengenalan yang berbeda – beda dengan di buktikan dengan program *software* audacity dan juga melodyne std 4, dengan aplikasi *software* ini untuk menentukan nilai karakteristik tiap manusia dengan bantuan aplikasi, yang terdiri dari nilai frekuensi, nilai kunci nada, dan nilai kestabilan CT atau *Compressor Threshold*.

Karena dalam setiap manusia pasti mempunyai nilai suara yang berbeda – beda, dan setiap manusia diciptakan dengan karakteristik ciri atau suara yang berbeda – beda, ada yang memiliki jenis suara yang rendah ataupun suara yang tinggi, dari penjelasan tersebut bahwa disetiap nada suara atau dalam musik memiliki beberapa kategori suara sebagai berikut :

1. Frekuensi Nada.

Nada suara sederhana dalam hanya memiliki satu frekuensi meskipun intensitasnya dapat bervariasi, dalam hal ini tinggi rendahnya nada berdasarkan frekuensi menunjukkan pada persepsi atau frekuensi satu nada. Jadi Frekuensi *Pitch* suara manusia sangat berbeda – beda.

3. 4. Nilai Kunci Nada.

Dalam setiap manusia ketika melakukan perintah suara dengan berbicara langsung, hal ini membuktikan bahwa tiap manusia memiliki nilai ciri suara yang dikategorikan sebagai nilai kunci, contoh : Dicky melakukan perintah suara “kipas nyala” dan dia berada dinada Ab diatas C tengah yang memiliki tangga nada yang diset ekuivalen setara dengan 440 Hz.

1. Nilai CT (*Compressor Threshold*).

Nilai ini adalah nilai ambang batas tingkat yang perlu dinaikkan sinyal agar kompresor dapat mulai bekerja, jika sinyalnya terlalu rendah atau tidak melawati ambang batas, kompresor hanya akan membiarkan sinyal lewat mengenai data latihan tanpa berubah.

2. Pengujian Sistem

Setelah tahap perancangan selesai, selanjutnya adalah melakukan pengujian alat untuk melihat kinerja alat tersebut dan mengamati data yang didapat serta tingkat keberhasilan dalam perancangan sistem tersebut.

3. Implementasi

Selanjutnya setelah tahap pengujian selesai, alat tersebut kemudian di terapkan dan dilaksanakan serta dilakukan pengujian atau eksperimen untuk mendapatkan nilai yang akan ditindak lanjuti menganalisis sistem kerja alat.

4. Hasil dan Pembahasan

Hasil dari perancangan Sistem Pengenalan Suara Manusia dari bahan triplex dan kotak plastik yang berukuran 52 cm

x 41 cm untuk triplex, dan 18,5 cm x 26 cm untuk bahan plastik dengan ketebalan triplek 1cm. Seperti pada gambar 9 berikut ini :



Gambar 9. Alat Sistem Pengenalan Suara

Dalam tahap hasil pengujian perancangan sistem pengenalan suara manusia ini adalah dengan perangkat keras secara langsung dan pengujian menggunakan perangkat lunak software dengan aplikasi Audacity dan Melodyne Std 4. Adapun pengujian yang dilakukan sebagai berikut :

1. Pengujian perangkat keras (hardware) secara langsung, yaitu dengan menggunakan suara manusia secara langsung, pengujian ini nantinya akan menguji tingkat keberhasilan, kepekaan alat dengan mencirikan dari suara manusia yang berbeda – beda, pengujian ini untuk melihat kinerja dari perangkat keras yang berfungsi dengan baik, sebagai percobaan untuk mengidentifikasi tiap masing – masing suara manusia.
2. Pengujian perangkat lunak (software), pengujian berikutnya dengan menggunakan aplikasi perangkat lunak (software) berupa Arduino IDE, dan aplikasi penggabungan Audacity 2.4.2 dengan Melodyne Std 4, pengujian pada aplikasi ini nantinya akan memperlihatkan grafik dari nilai recognition rate tiap manusia berupa frekuensi, amplitudo, tingkat suara yang ideal, dan suara yang tidak ideal. Pengujian melalui aplikasi akan direkam kembali dari tiap manusia menggunakan rekaman *microphone* atau bisa juga dengan Google Voice, yang nantinya akan di analisa melalui perangkat lunak (software) yang telah disebutkan di atas, dari tiap suara manusia akan memiliki nilai ciri yang berbeda beda.
3. Pada gambar 10 - 12 merupakan tampilan monitor serial pada Arduino IDE, tampilan diatas adalah monitor untuk melakukan proses perekaman perintah suara adapun penjelasannya adalah sebagai berikut :

Clear = perintah menghapus semua nilai catatan pengenalan suara.

Record = perintah untuk merekam perintah suara

Train = perintah mencoba melakukan perintah suara dengan merekam untuk mengenali suara tersebut, dalam perintah ini maksimal output 7 perintah suara dengan contoh format coding : *Train 0 1 2 3 4 5 6* (7 Output Perintah)

Load = perintah digunakan untuk memuat atau menyimpan catatan pengenalan suara sebagai contoh load ini mengikuti perintah train, jika *train 0 1 2 3 4 5 6*,

maka untuk menyimpannya kita gunakan perintah *load 0 1 2 3 4 5 6*.

```
Elechouse Voice-----
Elechouse Voice Recognition V3 Module "train" sample.
-----
Usage:
-----
COMMAND      FORMAT          EXAMPLE          Comment
-----
train        train (r0) (r1)...  train 0 2 45     Train records
load        load (r0) (r1) ...  load 0 51 2 3    Load records
clear       clear            clear            remove all records in Recognizer
record     record / record (r0) (r1)...  record / record 0 79  Check record train status
vr         vr              vr              Check recognizer status
getsig     getsig (r)      getsig 0         Get signature of record (r)
sigtrain   sigtrain (r) (sig)  sigtrain 0 ZERO  Train one record(r) with signature(sig)
settings   settings        settings         Check current system settings
help       help            help            print this message
-----
```

Gambar 10. Tampilan Monitor Arduino IDE

```
Train success: 7
Record 0      Trained
Record 1      Trained
Record 2      Trained
Record 3      Trained
Record 4      Trained
Record 5      Trained
Record 6      Trained
```

Gambar 11. Tampilan *Train* / Latih perekaman berhasil dijalankan

```
-----
load 0 1 2 3 4 5 6
-----
Load success: 7
Record 0      Loaded
Record 1      Loaded
Record 2      Loaded
Record 3      Loaded
Record 4      Loaded
Record 5      Loaded
Record 6      Loaded
```

Gambar 12. Tampilan *Load* / pemanggilan perintah berhasil dijalankan

Vr = perintah *vr* digunakan untuk mengecek status perintah suara yang sudah di record sebelumnya, untuk mengetahui nilai pengenalan suara itu sama dengan si pengenalan atau tidak.

Dari Tabel 1 dibawah telah diuji sampel untuk sistem pengenalan suara manusia pada orang pertama, dengan menguji untuk mengetahui tingkat kepekaan pada perangkat keras (*hardware*) secara langsung, dengan memberikan perintah sebagai "lampu on" dan "lampu off" untuk lampu ke-1, didapatkan hasil bahwa tingkat

Tabel 1 Pengujian Menggunakan Perintah Suara Manusia Secara Langsung

| Perintah ucapan | Maksimal pengulangan ucapan | Jarak perintah (cm) | Jumlah pengulangan | | Delay/Waktu Respon (detik) | Persentase keberhasilan |
|------------------|-----------------------------|---------------------|--------------------|----------------|----------------------------|-------------------------|
| | | | Berhasil | Tidak berhasil | | |
| Nyalakan Lampu | 20 | 0 – 5 | 18 kali | 2 kali | 01,00 – 02,00 | 90 % |
| Matikan Lampu | 20 | 0 – 5 | 16 kali | 4 kali | 01,50 – 02,11 | 80 % |
| Nyalakan Lampu | 20 | 5 – 10 | 17 kali | 3 kali | 01,27 – 01,90 | 85 % |
| Matikan Lampu | 20 | 5 – 10 | 17 kali | 3 kali | 01,90 – 02,22 | 85 % |
| Nyalakan Lampu | 20 | 10 – 20 | 15 kali | 5 kali | 01,50 – 02,00 | 75 % |
| Matikan Lampu | 20 | 10 – 20 | 13 kali | 7 kali | 01,95 – 02,11 | 65 % |
| Nyalakan Lampu | 30 | 20 – 30 | 12 kali | 18 kali | 01,00 – 01,10 | 40 % |
| Matikan Lampu | 30 | 20 – 30 | 11 kali | 19 kali | 01,15 – 02,00 | 36,7 % |
| Lampu Mati Semua | 20 | 0 – 5 | 18 kali | 2 kali | 01,00 – 01,11 | 85 % |

keberhasilan perintah suara dengan persentase tertinggi pada jarak 0 – 5 cm, dengan persentase keberhasilan mencapai 90% dengan persamaan sebagai berikut :

Persentase Keberhasilan

$$= \frac{\text{Jumlah bagian berhasil}}{\text{Jumlah keseluruhan}} \times 100\% \quad (4)$$

Dari Tabel 2 dijelaskan bahwa nilai dibawah adalah nilai dari persentase keberhasilan dengan persentase tertinggi,

dari alat ini juga bisa dianalisa bahwa alat tersebut akan bekerja dengan baik, jika tidak ada *noise* disekitar atau kebisingan, penulis berpendapat juga bahwa jika alat tersebut diuji dengan banyaknya kebisingan dan *noise* yang berlebihan masih tetap bisa berjalan dengan baik, namun dengan si manusia ini berbicara dengan dengan frekuensi yang tinggi atau nada suara yang keras, tetapi akurasi suara dan intonasi suara harus sama dengan sample yang telah direkam si pengenalan suara sebelumnya.

Tabel 2. Persentase Keberhasilan Nilai Tertinggi Suara Manusia

| Pengguna | Total pengulangan ucapan | Jarak perintah (cm) | Jumlah pengulangan | | Persentase keberhasilan (%) |
|---------------------------|--------------------------|---------------------|--------------------|----------------|-----------------------------|
| | | | Berhasil | Tidak berhasil | |
| Manusia ke -1 (Widiyanto) | 80 | 0 – 5 | 70 kali | 10 kali | 87.5 % |
| Manusia ke -2 (Dicky) | 80 | 0 – 5 | 71 kali | 9 kali | 88.75 % |
| Manusia ke -3 (Hafidz) | 80 | 0 – 5 | 69 kali | 11 kali | 86.25 % |
| Manusia ke – 4 (Mutia) | 80 | 0 – 5 | 70 kali | 10 kali | 87.5 % |

Tabel 3 Persentase Keberhasilan Menggunakan Aplikasi

| Pengguna | Total Pengulangan ucapan | Jenis Perangkat | Jumlah | | Persentase keberhasilan (%) |
|---------------------------|--------------------------|------------------|----------|----------------|-----------------------------|
| | | | Berhasil | Tidak berhasil | |
| Manusia ke -1 (Widiyanto) | 30 | Lampu ke-1 | 29 kali | 1 kali | 96.7 |
| | 30 | Lampu ke-2 | 29 kali | 1 kali | 96.7 |
| | 10 | Lampu ke -3 | 9 kali | 1 kali | 90 |
| Manusia ke -2 (Dicky) | 10 | Lampu Mati Semua | 9 kali | 1 kali | 90 |
| | 30 | Lampu ke-1 | 29 kali | 1 kali | 96.7 |
| | 30 | Lampu ke-2 | 28 kali | 2 kali | 93.4 |
| Manusia ke -3 (Hafidz) | 10 | Lampu ke -3 | 9 kali | 1 kali | 90 |
| | 10 | Lampu Mati Semua | 9 kali | 1 kali | 90 |
| | 30 | Lampu ke-1 | 29 kali | 1 kali | 96.7 |
| Manusia ke – 4 (Mutia) | 30 | Lampu ke-1 | 29 kali | 1 kali | 96.7 |
| | 10 | Lampu ke-2 | 10 kali | 0 | 100 |
| | 10 | Lampu ke -3 | 9 kali | 1 kali | 90 |
| Manusia ke – 4 (Mutia) | 30 | Lampu ke-1 | 30 kali | 0 | 100 |
| | 30 | Lampu ke-1 | 29 kali | 1 kali | 96.7 |
| | 10 | Lampu ke-2 | 9 kali | 1 kali | 90 |
| | 10 | Lampu ke -3 | 9 kali | 1 kali | 90 |

Dari tabel 3 dijelaskan bahwa pengujian persentase keberhasilan menggunakan aplikasi dengan me *record* suatu perintah ucapan sangat efektif dan efisien, karena dengan suara dan intonasi yang tepat, alat langsung bekerja dengan baik. Pengujian diatas yaitu setiap manusia memiliki maksimal perintah output suara sebanyak 80 perintah suara.

Total 320 kali pengulangan ucapan suara, dengan keberhasilan 305 kali ucapan dan kegagalan 15 kali ucapan,

dengan pembagian output perangkat seperti ditunjukkan pada tabel diatas, persentase keberhasilan perintah suara dapat dituliskan melalui rumus berikut :

Persentase Keberhasilan

$$= \frac{\text{Jumlah Total Pengulangan Ucapan} - \text{Jumlah kegagalan Ucapan}}{\text{Jumlah Total Pengulangan Ucapan}} \times 100\% \quad (5)$$

$$\text{Persentase Keberhasilan} = \frac{320 - 15}{320} \times 100\%$$

$$\text{Persentase Keberhasilan} = 95.31\%$$

Terjadi total 15 kali kegagalan dalam 320 kali pengujian, dengan maksimal 80 kali perintah ucapan tiap manusia, sehingga persentase keberhasilannya 95.31 %. Persamaan (5) diatas memberikan jumlah bagian yang berhasil terhadap jumlah keseluruhan.

Persamaan (5) juga mempunyai bentuk yang sama dengan persamaan (4), dimana persamaan dihitung dengan memberikan nilai kegagalan dari jumlah total pengujian.

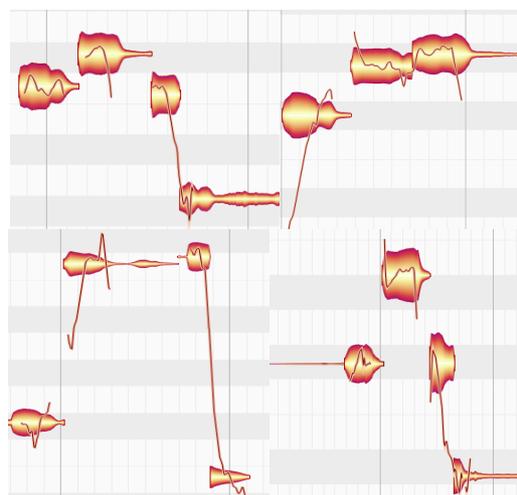
Tabel 4 Nilai Karakteristik Tiap Suara Manusia

| Nama Perintah Suara | Nada Frekuensi (Hz) | Nilai Kunci atau Nada Suara Tiap Intonasi | Nilai CT (Compressor Threshold) Tiap Intonasi |
|---------------------|---------------------|---|---|
| Widiyanto | 131.5 | C3 | +8 ct |
| | 139.8 | Db3 | +15 ct |
| | 129.5 | C3 | -18 ct |
| Dicky | 106.3 | Ab2 | +41 ct |
| | 185.2 | Gb3 | +2 ct |
| Hafidz | 204.4 | Ab3 | -27 ct |
| | 209.1 | Ab3 | +12 ct |
| | 93.3 | Gb2 | +14 ct |
| Mutia | 131.6 | C3 | +11 |
| | 133.6 | C3 | +36 |
| | 82.9 | E2 | +10 |
| | 276.4 | Db4 | -5 ct |
| | 320.1 | Eb4 | +49 |
| | 276.8 | Db4 | -3 |
| | 229.7 | Bb3 | -25 |

Pada Tabel 4 merupakan contoh perbandingan ciri atau karakteristik tiap suara antara suara manusia ke 1 sampai dengan manusia ke-4 dengan percobaan data latih perintah suara “lampu on”, karena setiap manusia memiliki nilai karakteristik yang berbeda beda, dari percobaan diatas setiap manusia memiliki tingkat suara yang berbeda, ada yang suaranya kecil, rendah, tinggi, kasar dan lainnnya. namun nilai karakteristik ini dapat berubah sesuai dengan tinggi rendahnya perintah suara tersebut. Pengujian diatas menunjukkan bahwa setiap manusia memiliki nilai :

1. Frekuensi Nada : Bahwa dari data diatas tiap manusia memiliki jumlah frekuensi nada (*pitch*) di setiap intonasinya.
2. Nilai Kunci Nada : Bahwa dari data diatas tiap manusia memiliki nilai kunci atau nilai ciri, seperti contoh Widiyanto mendapatkan nilai kunci C3 pada intonasi pertama, berbeda dengan Hafidz yang mendapatkan nilai Gb2 di intonasi pertama, hal ini membuktikan bahwa jenis suara seperti tinggi rendahnya suara yang diucapkan tiap seseorang berbeda – beda.
3. Nilai CT (*Compressor Threshold*) : Nilai CT ini adalah nilai ambang batas tingkat yang perlu dinaikkan sinyal agar kompresor dapat mulai bekerja, jika sinyalnya terlalu rendah atau tidak melewati ambang batas, kompresor hanya akan membiarkan sinyal lewat mengenai data latih tanpa berubah. Contoh : Nilai CT Mutia di intonasi pertama -5 CT dengan Kunci Nada

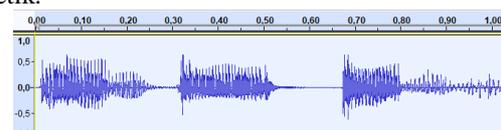
Db3, hal ini membuktikan bahwa nilai suara optimal Mutia yaitu di -5 ct dengan nilai ambang batas yang menyentuh sinyal, dengan demikian suara berikutnya akan berubah. Karena nilai optimal kunci nada di Oct, ketika mendapatkan nilai +ct yang besar , sinyal akan melewati di atas dan berada di ambang batas suara yang normal. Ketika nilai ct berada di tengah – tengah maka nilai nada suara akan mengenai sinyal dan suara akan berubah



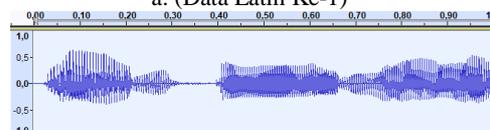
Gambar 13 Perbandingan Nilai Suara Tiap Intonasi dan Ketukan

Pada gambar 13 merupakan perbandingan nilai suara tiap intonasi dan ketukan pada saat melakukan perintah suara, gambar diatas juga dijelaskan bahwa tiap intonasi atau ketukan memiliki nilai karakteristik yang berbeda – beda, contohnya ketika kita melakukan perintah suara dengan nada rendah, itu memiliki nilai ciri atau *pitch* contohnya C3, sebaliknya kita melakukan perintah suara dengan nada tinggi itu memiliki nilai ciri DB4, jadi dari semua nilai perintah suara di atas, setiap manusia dapat diuji dengan mengambil sampel dari nilai frekuensi nada, nilai ct (*compressor threshold*) dan nilai ciri kunci nada. Nada dari total 3 sampe 4 nada ketukan. Sample diatas memiliki :

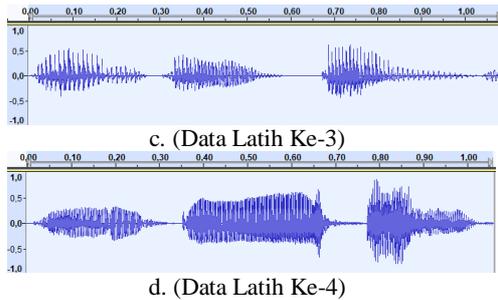
Gambar 14 (a) : memiliki 4 *pitch* ketukan dengan *time* : 1.50 detik.
 Gambar 14 (b) : memiliki 3 *pitch* ketukan dengan *time* : 1.34 detik.
 Gambar 14 (c) : memiliki 4 *pitch* ketukan dengan *time* : 1.60 detik.
 Gambar 14 (d) : memiliki 4 *pitch* ketukan dengan *time* : 1.61 detik.



a. (Data Latih Ke-1)



b. (Data Latih Ke-2)



Gambar 14 Data Sampel Suara Manusia (Audacity)

Pada gambar 13 merupakan pengujian gambar sample orang pertama dengan contoh perintah rekaman “lampu on”, pengolahan data perintah suara ini dilakukan dengan menggunakan perangkat lunak (*software*) Audacity 2.4.2 dan dikombinasikan menggunakan Melodyne Std 4. Sample pengujian direkam dan dipotong dengan aplikasi tersebut, untuk mengetahui dan memperoleh data perintah ucapan yang diinginkan, pengambilan sample pada gambar 14 merupakan pengambilan sample pada kondisi stabil dan ideal.

5. Kesimpulan

Alat sistem pengenalan suara manusia ini berhasil mengenali kata dan perintah ucapan dengan cukup baik, dari semua pengujian perintah suara, pada jarak 0 – 5 cm tingkat persentase keberhasilan yang paling besar mencapai 95 %. Kemudian pengujian persentase keberhasilan dengan menggunakan aplikasi (*software*) dengan perintah rekaman perintah ucapan tingkat persentase keberhasilan mencapai 96.95 % .

Perintah ucapan pada alat ini mampu memberikan respon suara dengan jangkauan *microphone* dari 30 cm – 120 cm. Jika volume perintah suara lebih besar, maka persentase tingkat keberhasilan suara akan semakin tinggi, ketika volume suara perintah kecil dan menurun persentase tingkat suara akan semakin sedikit. Hal ini sangat berpengaruh dan akan membuat kinerja alat sistem pengenalan suara bekerja dengan baik.

Pada aplikasi perangkat *software* pengujian perintah suara dengan rekaman berfungsi dengan sangat baik, dalam hal ini tingkat kepekaan pada alat sangat lah berpengaruh dengan rekaman suara tersebut, karena ketukan dan intonasi sesuai dengan apa yang di *record* pada rekaman perintah suara tersebut dari tiap manusia.

Pengujian nilai karakteristik suara berhasil dengan penjelasan bahwa setiap manusia memiliki nilai *pitch* atau nilai nada yang berbeda – beda. Dengan demikian *recognition rate* bisa ditampilkan melalui nilai *Pitch*, Time, Tempo, Ketukan dan intonasi suara yang ditampilkan pada Gambar 13 , 14 dan Tabel 4.

Pengujian perangkat pada sistem pengenalan suara manusia ini sebanyak 4 Manusia, 3 Laki – laki dan 1 Perempuan. Dengan masing – masing manusia memiliki output 7 perintah suara, dengan maksimal pengaturan perintah suara sebanyak 80 *command* suara .

Daftar Pustaka

- [1] A. Faroqi, M. S. WS, and R. Nugraha, “Perancangan Sistem Kontrol Otomatis Lampu Menggunakan Metode Pengenalan Suara Berbasis Arduino,” *TELKA - Telekomun. Elektron. Komputasi dan Kontrol*, vol. 2, no. 2, pp. 106–117, 2016, doi: 10.15575/telka.v2i2.31.
- [2] V. W. Rumopa and J. Luther Mappadang, “Kontrol Penerangan Ruang Menggunakan Sensor Suara (Speech Recognition) Berbasis Android,” *Tugas Akhir*, 2015.
- [3] S. N. Rohman, A. Hidayanto, and A. A. Zahra, “Aplikasi Pencirian Dengan Linear Predictive Coding Untuk Pembelajaran Pengucapan Nama Hewan Dalam Bahasa Inggris Menggunakan Jaringan Syaraf Tiruan Propagasi Balik,” *Transmisi*, vol. 14, no. 4, pp. 1–9, 2012.
- [4] N. R. Pradipta, T. Tasripan, and H. Kusuma, “Perancangan Perangkat Antarmuka Berbasis Pengenalan Suara pada Purwarupa Mesin Cetak Huruf Braille ITS,” *J. Tek. ITS*, vol. 7, no. 2, 2019, DOI: 10.12962/j23373539.v7i2.30923.
- [5] Vipul C. Rajyaguru, “Different Methods Used In Voice Recognition Techniques,” *Int. Res. J. Eng. Technol.*, pp. 699–703, 2016.
- [6] S. Aulia, L. Lifwarda, and Y. Yustini, “Pengenalan Bentuk Benda Berdasarkan Sinyal Suara dengan Transducer Mikrofon dan Teknologi Kinect,” *J. Nas. Tek. Elektro*, vol. 7, no. 3, p. 191, 2018, doi: 10.25077/jnte.v7n3.600.2018.
- [7] J. M. T. S., D. Puspitaningrum, and B. Susilo, “Penerapan Speech Recognition Pada Permainan Teka-Teki Silang Menggunakan Metode Hidden Markov Model (HMM) Berbasis Desktop,” *J. Rekrusif*, vol. 4, no. 1, pp. 119–129, 2016.
- [8] P. Saputra, “Smart Home dengan Speech Recognition Melalui Bluetooth Berbasis Android,” *Tugas Akhir*, vol. 6, pp. 5–9, 2017.