

ALAT BANTU PINTAR UNTUK KAUM TUNA NETRA DENGAN IMPLEMENTASI PENGOLAHAN SUARA DAN PELACAKAN LOKASI

SMART AID FOR THE VISUAL IMPAIRED WITH SPEECH PROCESSING AND LOCATION TRACKING IMPLEMENTATION

Giri Pridyasmoro Putra¹, Budhi Irawan, S.Si., M.T.², Anggunmeka Luhur P., S.Si., M.T.³
^{1,2,3}Prodi S1 Teknik Komputer, Fakultas Teknik Elektro, Universitas Telkom
¹gputraa@telkomuniversity.ac.id, ²budhiirawan@telkomuniversity.ac.id,
³anggunmeka@telkomuniversity.ac.id

Abstrak

Tunanetra merupakan orang yang memiliki gangguan penglihatan dimana yang menyebabkan masalah tersebut tidak bisa diperbaiki dengan cara biasa, seperti kacamata. Dengan berbagai rintangan berbeda seperti dibutuhkannya cara yang cepat dan mudah untuk menghubungi keluarga saat ada emergensi saat bekerja, maka dibuatlah alat bantu yang dapat membantu kaum tuna netra dan orang disekitarnya.

Alat bantu memiliki fitur pelacakan lokasi yang bisa diakses oleh keluarga untuk melihat lokasi orang tunanetra tersebut lewat Google Maps yang terintegrasi ke aplikasi android. Alat bantu memiliki modul GSM untuk mengirim SMS, dimana SMS tersebut memiliki data berupa koordinat GPS yang dapat dilihat di aplikasi handphone, selain itu koordinat GPS juga dapat diubah ke bentuk alamat dengan reverse geocoding. Modul GSM tersebut juga bekerja sama dengan perintah suara untuk menjalankan perintah-perintah yang sudah di konfigurasi seperti mengirim SMS ke anggota keluarga saat keadaan darurat. Alat bantu tidak ada tombol fisik dan pengguna hanya menggunakan perintah suara untuk berinteraksi dengan alat bantu.

Setelah penelitian ini dijalankan alat bantu dapat memudahkan keluarga untuk melakukan monitoring lokasi terhadap pengguna alat dengan tingkat akurasi yang cukup akurat hingga 10 meter, hal ini juga menyebabkan hasil reverse geocoding yang akurat sehingga alamat keberadaan tunanetra cukup akurat. Perintah suara alat bantu hanya mendapatkan tingkat akurasi sebesar 66.07% karena beberapa faktor seperti kualitas mikrofon, penyebutan perintah yang terlalu cepat dan tingkat kebisingan lokasi.

Kata Kunci: GPS, GSM, perintah suara, alat bantu

Abstract

The visual impaired is a person who has vision problems that cannot be fixed by normal means, such as eyeglasses. With all the different challenges that is faced by someone with visual impairment such as having a quick access to get in contact with a family member or guardian if there is any emergency while working, a device is made that can help a visual impaired person and his/her family.

The device has features implemented such as location tracking that can be accessed by a family member through google maps that is integrated within an android application. This device also has a GSM and GPS module. The GSM module is able to send SMS where it contains data such as the coordinates that can be viewed at the application. Other than that, the coordinate can also be converted into a full address that is read to the person with visual impairment with the help of a text-to-speech system so that they can know where they are currently at. The module also works with a speech recognition system implemented in the microcontroller, where it will be able to run pre configured commands such as sending his/her location details to the application. The device does not have any physical buttons and can only be interacted by the voice commands that is configured.

After the research has been done, the device is able to help the family members locate the person with visual impairment up to an accuracy within 10 meters, this also cause the result of reverse geocoding to be accurate. The speech recognition system is only able to get an accuracy of 66.07% due to a few factors such as the microphone quality, the speed of commands being said and the noise level of the surrounding areas.

Keyword : GPS, GSM, Speech recognition, Smart Aid

1. Pendahuluan

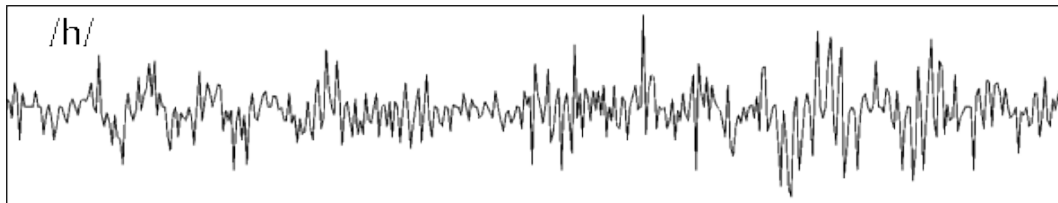
Tunanetra merupakan orang yang memiliki gangguan penglihatan dimana yang menyebabkan masalah tersebut tidak bisa diperbaiki dengan cara biasa, seperti kacamata. Tunanetra dibagi menjadi 3 tingkat yaitu moderate visual impairment, sever visual impairment dan total blindness. Orang dengan gangguan penglihatan mempunyai kesusahan dalam menjalankan aktivitas sehari-hari seperti, berjalan kaki, bersosialisasi dan membaca. Kondisi ini diperkirakan dimiliki oleh lebih dari 285 juta orang di seluruh dunia[10].

PSBN (Panti Sosial Bina Netra) Wyata Guna Bandung yang bertugas untuk mengrehabilitas dan melakukan perlindungan kepada penyandang disabilitas netra, memiliki aktivitas setiap hari layaknya seperti sekolah pada umumnya yang ditujukan agar bisa membangun keterampilan yang bisa di gunakan oleh penyandang disabilitas netra. Oleh karena itu, menjadikan aktivitas sehari-hari penyandang disabilitas tuna-netra saat setelah lulus dari PSBN Wyata Guna, sebagai dasar untuk merancang sistem alat bantu, dan juga membuat alat berbeda dari smart cane yang sudah banyak beredar, dibuatlah alat bantu pintar untuk keamanan serta memudahkan kaum tunanetra dan keluarganya. *Alat Bantu Pintar Untuk Kaum Tuna Netra Dengan Implementasi Pengolahan Suara dan Pelacakan Lokasi* di desain sesederhana mungkin sehingga hal yang harus dipelajari kaum tunanetra tersebut adalah hanya **perintah-perintah suara**, ditambah itu alat pintar akan berbentuk sebuah tas punggung sehingga selain bisa digunakan sebagai tempat untuk alat pintar, tas punggung juga dapat digunakan untuk menyimpan barang saat beraktivitas. Alat pintar diharapkan akan memudahkan keluarga untuk mencari tahu keberadaan pengguna alat dan juga dapat membantu Kaum Tuna Netra menjanjalkan aktivitas sehari-harinya.

2. Dasar Teori

2.1 Suara

Sound (suara), dalam ilmu fisika, adalah getaran suatu benda yang menghasilkan gelombang longitudinal. Getaran tersebut menghasilkan perbedaan tekanan di udara sekitarnya yang menghasilkan sebuah Pola Osilasi, pola tersebut dinamakan sebagai gelombang. Gelombang yang mempunyai pola sama berulang dengan interval tertentu bernama suara periodik seperti sebuah instrument musik. Suara non-periodik adalah suara yang tidak terjadi secara berlanjut seperti percikan ombak[8].



Gambar 2.1 Bentuk Gelombang voiceless speech

Suatu ucapan dapat dipresentasikan menggunakan model *waveform* atau bentuk gelombang karena dapat memperlihatkan perubahan intensitas suara tersebut pada selang waktu tertentu yang ingin diamati. Ada dua tipe sumber dari suatu ucapan suara[6]:

1. *Periodic vibrations* yang dihasilkan oleh pita suara sehingga menghasilkan *voiced speech*
2. *Aperiodic vibrations* yang dihasilkan oleh turbulensi pada saluran vokal sehingga menghasilkan suatu *voiceless speech*.

Suara juga dapat dibagi ke unit bunyi terkecil bernama fonem. Fonem adalah unit bunyi terkecil yang dapat dibedakan oleh manusia. Ucapan kata-kata atau kalimat secara dasar adalah kumpulan urutan fonem. Perkumpulan fonem pada setiap bahasa berbeda-beda dan direpresentasikan dengan simbol yang unik[8].

Pengenalan ucapan atau suara (*speech recognition*) adalah suatu teknik yang memungkinkan komputer untuk menerima input berupa suara *periodic* yang diucapkan dan mengubahnya menjadi *raw data* yang bisa ditafsirkan oleh sistem dalam bentuk teks.

2.2 Markov Model

Markov Model adalah sebuah model stokastik yang didasarkan kepada teori probabilitas yang dapat digunakan untuk memodelkan sebuah rangkaian kejadian. Pada Markov Model, probabilitas sebuah *state* (kondisi) bergantung pada kondisi sebelumnya. *Finite automaton* adalah kumpulan kondisi dimana transisi antar state dilakukan berdasarkan masukan observasi. Pada Markov Model setiap transisi antar kondisi berisi probabilitas yang mengindikasikan kemungkinan jalur yang akan diambil[5].

Namun kondisi sebelumnya tidak akan selalu dapat diketahui, maka hadirlah Hidden Markov Model. Perbedaan HMM dengan Markov Model adalah pada HMM suatu sistem yang dimodelkan memiliki kondisi yang tidak terobservasi. Setiap kondisi awal memiliki distribusi kemungkinan ke seluruh kondisi output yang ada[2].

Secara umum, HMM terdiri dari elemen-elemen sebagai berikut[2]:

1. Himpunan *state* atau kondisi, N , adalah jumlah *state* atau kondisi pada sebuah Hidden Markov Model, dengan ruang *state* dan *state* pada waktu, t , dinyatakan dengan Q_t .

$$S = \{s_1, s_2, \dots, s_n\} \quad (1)$$

2. Himpunan nilai observasi, M , yaitu jumlah pengamatan (observasi) tiap *state*, dengan ruang observasi

$$V = \{v_1, v_2, \dots, v_m\}. \quad (2)$$

3. Himpunan probabilitas transisi antar *state*. Dimana *state* berikutnya tergantung pada *state* saat ini.

Probabilitas transisi dapat dinyatakan dengan matriks:

$$A = [a_{ij}] \tag{3}$$

dimana a_{ij} adalah probabilitas transisi dari state i ke state j .

4. $B=[b_{jm}]$, adalah matriks peluang bersyarat observasi v_m jika proses berada pada kondisi j , dimana:

$$b_{jm} = b_j(O_t) = P(O_t = v_m | Q_t = s_j), 1 \leq j \leq N \text{ dan } 1 \leq m \leq M \tag{4}$$

5. Himpunan kondisi awal adalah π_i , dimana i menjadi state awal pada urutan HMM.

Sehingga *Hidden Markov Model* dapat dituliskan dalam notasi:

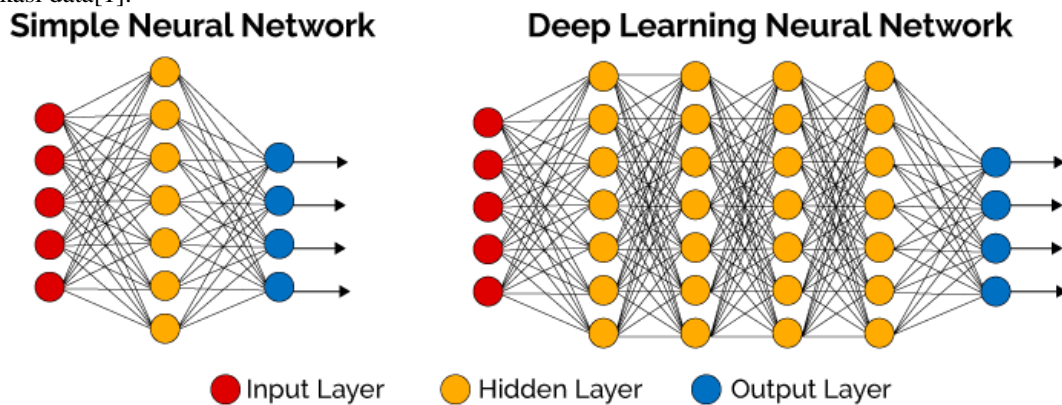
$$\lambda = (A, B, \pi) \tag{5}$$

Jika parameter tersebut diberikan maka HMM dapat digunakan sebagai pembangkit barisan observasi:

$$O = O_1, O_2, O_3, \dots, O_T \tag{6}$$

2.3 Deep Neural Network

Neural Network pada dasarnya mengadopsi kemampuan otak manusia yang mampu menjanjalkan proses dan memberika suatu output. Setiap neuron pada otak manusia saling terhubung dan mengalirkan informasi setiap neuron tersebut. Neural Network bisa diaplikasikan untuk banyak hal tertentu seperti pengenalan pola atau klasifikasi data[1].

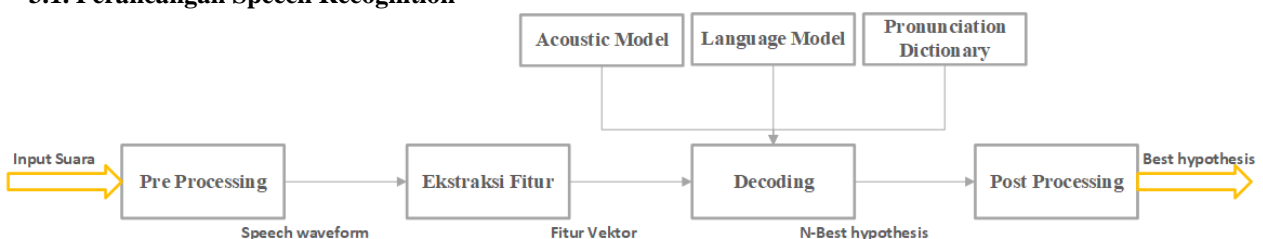


Gambar 2.2. Deep Neural Network Layers

DNN adalah suatu feedforward artificial neural network yang memiliki lebih dari satu lapisan tersembunyi antara bagian input dan outputnya. Kelebihan Deep Learning Methods untuk Speech Recognition adalah arsitektur jaringan lebih baik dan juga dapat mengoptimalkan banyaknya parameter untuk mendapatkan hasil yang lebih baik untuk Speech Recognition[3].

3. Pembahasan

3.1. Perancangan Speech Recognition



Gambar 3.1 Proses Umum Speech Recognition

Secara garis besar Speech Recognition memiliki alur kerja seperti pada gambar 3.1. Proses tersebut akan dijelaskan pada bagian berikut[4]:

3.1.1. Pre-Processing

Input suara yang merupakan sebuah gelombang suara analog dikonversi ke dalam bentuk digital sehingga menghasilkan suatu speech waveform.

Sistem pada komputer mengukur amplitude pada satuan waktu tertentu yang menghasilkan sejumlah angka. Tiap satuan pengukuran ini dinamakan "sample". ADC atau analog-to-digital converter mengubah sinyal analog ini ke sebuah data digital dengan mengubah amplitudo gelombang suara ke dalam waktu interval (sampling), sehingga menghasilkan representasi suara dalam bentuk digital.

Pada pre processing sinyal suara digital tersebut dapat dilakukan normalisasi, menghilangkan noise, dan melakukan proses filtering.

3.1.2. Ekstraksi Fitur

Tujuan ekstraksi fitur adalah untuk mengubah speech yang diterima menjadi representasi internal, dimana speech tersebut dibagi ke bagian-bagian dengan tingkat informasi yang jauh lebih rendah. Data yang dibagi tersebut di proses satu per satu berdasarkan urutannya, seperti pada Gambar 2.3 dimana hasilnya bisa disebut sebagai fitur vektor.

Ada beberapa metode yang digunakan untuk melakukan ekstraksi fitur, yaitu Mel Frequency Cepstral Coefficients (MFCC), Linear Prediction Coefficient (LPC), dan Perceptually Based Linear Predictive Analysis (PLP).

MFCC adalah metode ekstraksi ciri yang paling sering digunakan. MFCC merupakan metode ekstraksi fitur sinyal suara yang memiliki karakteristik paling mirip dengan pendengaran telinga manusia dibandingkan metode lainnya. 'Mel Scale' merupakan skala yang digunakan untuk menghitung pendengaran telinga manusia[7].

3.1.3. Decoding

Tahap paling penting dalam proses speech recognition. Decoding dilakukan untuk mendapatkan hasil yang paling cocok untuk fitur vektor yang diterima dari tahap sebelumnya. Decoder adalah tahap dimana speech yang sudah di proses di tentukan hasilnya dengan cara menggabungkan dan mengoptimalkan informasi yang diterima dari model akustik dan language model.

Ada tiga bagian dari decoding yaitu:

1. Acoustic Modelling - Model akustik digunakan untuk merepresentasikan hubungan antara bentuk gelombang dan fonem atau unit lainnya yang menghasilkan suatu ucapan. Salah satu metode yang digunakan adalah Hidden Markov Model.

2. Language Modelling

Digunakan untuk pencarian urutan kata yang benar dari speech dan memprediksi kemungkinan kata ke N-1, atau kata sebelumnya. Model Bahasa dapat dibagikan ke dua kategori, *Uniform Model* dan *Stochastic Model*.

3. Pronunciation Modelling

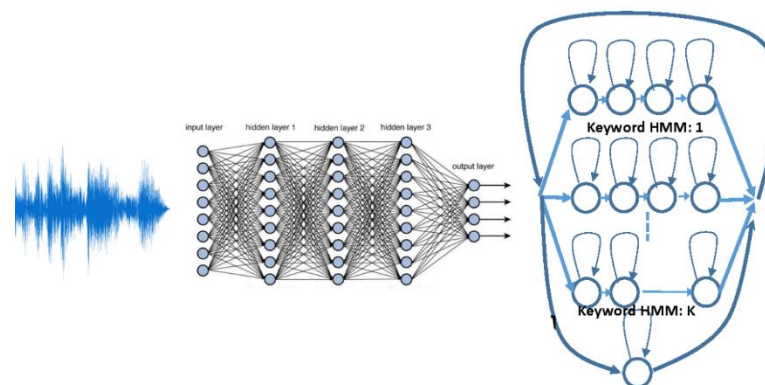
Urutan yang dihasilkan oleh model akustik HMM dibandingkan dengan serangkaian kata pada suatu database kosakata. Urutan kata-kata yang dihasilkan memiliki informasi tentang kata yang dapat di deteksi dan juga bagaimana kata tersebut diucapkan, atau urutan fonem dari kata tersebut. Decoder kemudian digunakan untuk mengenali pesan tersebut dengan menggabungkan dan mengoptimalkan informasi yang diterima dari model akustik dan model bahasa.

Ukuran database kosakata mempengaruhi kompleksitas dan akurasi pada hasil yang didapatkan. Dengan database kosakata yang lebih besar, memungkinkan sebuah perangkat speech recognition untuk mengenali ucapan yang lebih panjang dan kompleks.

3.1.4. Post Processing

Setelah menjalankan proses dari decoder maka didapatkan hasil dari speech yang menjadi input. Proses ini menentukan hasil akhir dari speech yang diterima dan memberi output berupa raw data berbentuk plain teks.

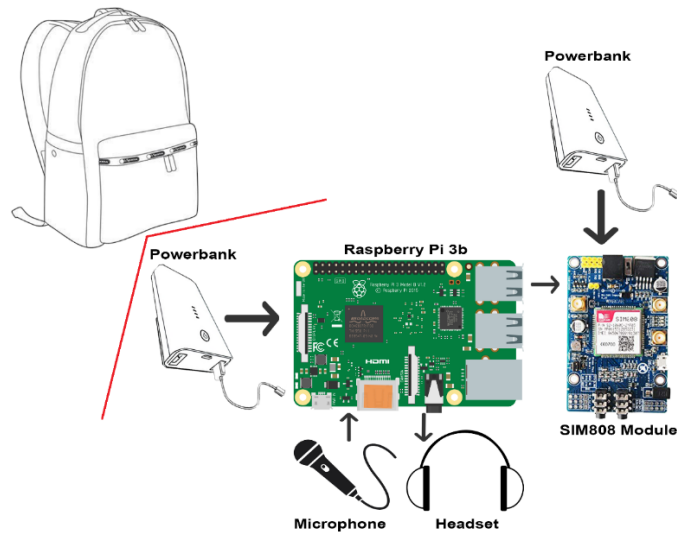
3.2. Speech Recognition menggunakan DNN-HMM



Gambar 3.2 Implementasi Speech Recognition berbasis DNN-HMM

Salah satu metode terbaik untuk melakukan speech recognition adalah dengan gabungan antara Deep Neural Network dan Hidden Markov Model. Output dari DNN dijadikan sebagai input untuk proses HMM sehingga meningkatkan keakuratan probabilitas speech[9].

3.3. Rangkaian Perangkat Keras Alat Pintar



Gambar 3.3 Gambaran Umum Sistem

Pada Gambar 3.3 merupakan gambaran umum sistem yang penulis kerjakan. Sistem alat bantu pintar ini dirancang menggunakan Raspberry Pi 3 sebagai mikrokontroler dengan bantuan modul SIM808 dan sebuah microphone. Software yang digunakan adalah Python 2.7 dengan library Cloud Speech-to-Text (CSTT) dari Google untuk mengkonversi sebuah suara ke text (speech recognition), sehingga bisa diolah untuk menjalankan speech processing, yaitu pengolahan suara untuk menjalankan perintah tertentu. Modul SIM808 digunakan untuk mendapatkan koordinat GPS dan juga untuk mengirim SMS kepada aplikasi handphone.

4. Pengujian dan Hasil

Performa dari alat pintar untuk mengenali perintah suara menggunakan metode DNN-HMM dapat dinilai berdasarkan banyaknya kata yang sukses dikenali dalam sekian banyak percobaan.

$$Akurasi = \frac{\text{banyak kata terdeteksi}}{\text{total kata diuji}} \times 100\% \tag{7}$$

4.1 Pengujian Jarak Microphone

Sebelum menguji tingkat akurasi pendeteksian suara maka dicari dahulu jarak optimum microphone. Kata “STATUS” akan disebut sebanyak 5 kali dalam setiap jarak, dan diuji tingkat akurasinya. Pengujian ini di ulang sebanyak 5 kali dan dilakukan dalam dua skenario berbeda yang memiliki 2 tingkat decibel.

Tabel 4.1 Tingkat Akurasi Jarak Microphone 35 dB

Jarak	Akurasi rata-rata(%)
5 cm	76
10 cm	68
20 cm	48
Akurasi keseluruhan(%)	64

Tabel 4.2 Tingkat Akurasi Jarak Microphone 55 dB

Jarak	Akurasi rata-rata(%)
5 cm	60
10 cm	56
20 cm	48
Akurasi keseluruhan(%)	54.67

4.2 Pengujian Deteksi Perintah Suara

Pada pengujian ini, ada 7 perintah suara pada alat pintar yang akan di uji coba akurasinya dengan metode pengujian yang sama pada pengujian jarak microphone. Uji coba dilakukan oleh 10 responden berbeda dimana responden akan menyebut setiap perintah secara berurut dan diulang sebanyak 5 kali.

Tabel 4.3 Tingkat Akurasi Deteksi Perintah Suara 35 dB

Perintah Suara	Akurasi rata-rata (%)
Hai Bola	54.00
Lokasi Saya	84.00
Kirim Lokasi	72.00
Waktu	74.00
Status	58.00
Kembali	70.00
Akurasi total (%)	68.67

Tabel 4.4 Tingkat Akurasi Deteksi Perintah Suara 55 dB

Perintah Suara	Akurasi rata-rata (%)
Hai Bola	54.0
Lokasi Saya	70.0
Kirim Lokasi	62.0
Waktu	66.0
Status	58.0
Kembali	74.0
Akurasi total (%)	64.0

Dari pengujian yang sudah dilaksanakan maka dapat diketahui bahwa perintah suara dengan suku kata lebih banyak dapat terdeteksi oleh alat pintar jika dibandingkan dengan perintah yang memiliki hanya satu atau dua suku kata. Tingkat kebisingan ruangan pengujian juga dapat mempengaruhi proses speech recognition dikarenakan noise berlebihan membuat alat untuk gagal mendeteksi perintah.

4.3 Pengujian Akurasi GPS

Pada alat pintar selain speech recognition, tingkat akurasi GPS sangat dibutuhkan oleh alat pintar untuk melakukan *reverse geocoding* koordinat. Sehingga akurasi GPS yang tinggi menghasilkan alamat yang lebih sesuai. Oleh karena itu di uji coba tingkat akurasi GPS saat mendapatkan *lock* terhadap *Satelite Vechile*. Hasil koordinat yang didapat dibandingkan ke posisi pengujian berada dan dibandingkan berapa meter jauhnya.

Tabel 4.5 Parameter Tingkat Akurasi GPS

Parameter Pengujian	Sangat Akurat	Akurat	Cukup Akurat	Tidak Akurat
	Jangkauan koordinat berada dalam 11 meter.	Jangkauan koordinat berada dalam 21 meter.	Jangkauan koordinat berada dalam 51 meter.	Jangkauan koordinat melebihi 50 meter.

Dari 10 pengujian maka didapatkan hasil sebagai berikut:

Tabel 4.6 Tingkat Akurasi Koordinat GPS (*Outdoor*)

Lokasi 1 Pengujian	-6.990263, 107.646611		
	Koordinat Didapat	Jarak (meter)	Hasil
1	-6.990187,107.646572	9.48	Sangat Akurat
2	-6.990272,107.646545	7.35	Sangat Akurat
3	-6.990313,107.646528	10.72	Akurat
4	-6.990278,107.646523	9.85	Sangat Akurat
5	-6.990295,107.646535	9.11	Sangat Akurat
6	-6.990273,107.646545	7.37	Sangat Akurat
7	-6.990292,107.646518	10.76	Akurat
8	-6.990258,107.646510	11.16	Akurat
9	-6.990295,107.646535	9.11	Sangat Akurat
10	-6.990310,107.646530	10.36	Akurat

Tabel 4.7 Tingkat Akurasi Koordinat GPS (*Outdoor Restricted*)

Lokasi 2 Pengujian	-6.911037,107.630460		
	Koordinat Didapat	Jarak (meter)	Hasil
1	-6.910997,107.630517	7.71	Sangat Akurat
2	-6.911027,107.630613	16.93	Akurat
3	-6.911013,107.630512	6.33	Sangat Akurat
4	-6.911063,107.630535	8.77	Sangat Akurat
5	-6.910968,107.630395	10.5	Akurat
6	-6.911063,107.630350	12.48	Akurat
7	-6.911068,107.630457	3.46	Sangat Akurat
8	-6.911015,107.630383	8.84	Sangat Akurat
9	-6.911073,107.630420	5.96	Sangat Akurat
10	-6.911028,107.630582	13.5	Akurat

Tabel 4.8 Tingkat Akurasi Koordinat GPS (*Semi indoor*)

Lokasi 3 Pengujian	-6.910494, 107.626004		
	Koordinat Didapat	Jarak (meter)	Hasil
1	-6.910390,107.625977	11.94	Akurat
2	-6.910265,107.626202	33.56	Cukup Akurat
3	-6.910388,107.626077	14.28	Akurat
4	-6.910343,107.626068	18.22	Akurat
5	-6.910270,107.625910	26.98	Cukup Akurat
6	-6.910330,107.625950	19.19	Akurat
7	-6.910155,107.625835	42.06	Cukup Akurat
8	-6.910350,107.626043	16.58	Akurat
9	-6.910160,107.626088	38.28	Cukup Akurat
10	-6.910205,107.626152	36.05	Cukup Akurat

Jika hasil pengujian pertama dan kedua dibandingkan maka dapat diketahui bahwa kedua lokasi pengujian tersebut menghasilkan tingkat akurasi "Sangat Akurat" sebanyak 6 kali dari 10 percobaan, tetapi pada pengujian kedua (Tabel 4.7) hasil yang didapatkan memiliki fluktuasi yang cukup signifikan dibandingkan dengan pengujian pertama (Tabel 4.6) yang memberikan hasil lebih konsisten. Hal ini membuktikan bahwa walaupun hasil yang didapat pada kedua skenario tersebut tetap cenderung sangat akurat, adanya halangan dari alat bantu untuk terhubung ke satelit mempengaruhi tingkat akurasi koordinat yang didapat.

Pada pengujian ketiga yang dapat dilihat pada Tabel 4.8, jika alat bantu sukses terhubung ke SV di lokasi yang memiliki pandangan tertutup ke langit seperti adanya genteng, maka alat bantu tetap mendapatkan koordinat yang memasuki kategori cukup akurat. Tetapi jika pandangan alat ke langit terlalu padat, misalnya di dalam rumah maka alat bantu tidak akan dapat menghasilkan koordinat apapun.

Setelah melakukan pengujian tersebut, maka dapat diketahui bahwa tingkat akurasi koordinat cukup dipengaruhi oleh lokasi dan pandangan alat bantu ke langit, namun jika alat bantu sukses terhubung ke satelit tingkat akurasi yang di dapat cenderung akurat secara keseluruhan.

5. Kesimpulan dan Saran

5.1 Kesimpulan

Dari hasil pengujian dan analisa yang telah dilakukan, maka dapat disimpulkan sebagai berikut:

1. Alat bantu dapat memudahkan keluarga untuk melakukan monitoring lokasi terhadap pengguna tuna netra dengan tingkat akurasi yang cukup akurat dalam jangkauan 10 meter.
2. Alat bantu hanya dapat mendeteksi 66.33% dari perintah suara yang di ucapkan oleh responden. Sehingga mengakibatkan lebih dari 1/4 perintah suara untuk tidak jalan dengan baik, hal ini disebabkan oleh beberapa faktor seperti tingkat kebisingan lokasi pengujian dan juga kualitas mikrofon serta tingkat kecepatan penyebutan perintah suara.
3. Alat bantu berhasil menjalankan perintah jika terdeteksi dengan tingkat kesuksesan sebesar 100%. Hal ini membuktikan ke stabilan alat bantu saat berjalan tanpa mengalami error.

5.2 Saran

Hasil dari tugas akhir ini membutuhkan lebih banyak pengembangan, diantaranya:

1. Implementasi mikrofon (diubah ke bluetooth) dan Google Cloud Speech API agar bisa lebih konsisten sehingga dapat memudahkan pengguna alat bantu untuk menjalankan perintah.
2. Perubahan penggunaan SMS sebagai perantara alat bantu dengan aplikasi handphone, seperti penggunaan aplikasi messenger berbasis internet seperti Telegram.
3. Perangkat keras di *streamline*, dengan meminimalisir kabel serta menyatukan 4 perangkat keras tersebut menjadi dalam 1 kontainer.

Daftar Pustaka:

- [1] Christos Stregiou, Dimitrios Siganos, "Neural Networks". (Unknown). [Retrieved April 25, 2019], from https://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html
- [2] Firdaniza, Nurul Gusriani, Akmal, "Hidden Markov Model", *FMIPA Universitas Padjajaran*, 2006.
- [3] Li Deng, Geoffrey Hinton, and Brian Kingsbury, "New Type of Deep Neural Network Learning for Speech Recognition and Related Applications: An Overview", IEEE International Conference, 2013.
- [4] Magre, Smita & Deshmukh, Ratnadeep & P Shrishrimal, Pukhraj. (2013). A Comparative Study on Feature Extraction Techniques in Speech Recognition.
- [5] M. Fikri Shinwani, "Rancang Bangun Aplikasi Voice Translator Berbasis Android Menggunakan Hidden Markov Model", Skripsi, Malang: Universitas Islam Negeri Maulana Malik Ibrahim, 2016
- [6] N. Tia Ati Yanti, "Fonem Bahasa Indonesia". (2015). [Retrieved April 25, 2019], from https://www.academia.edu/12136237/FONEM_BAHASA_INDONESIA
- [7] Preeti Saini, Parneet Kaur, "Automatic Speech Recogniton: A Review", International Journal of Engineering Trends and Technology, 2013.
- [8] Ron Kurtus, "Overview of Wave Motion". (12 August 2014). [Retrieved 5 January 2019], from https://www.school-for-champions.com/science/wave_motion.htm
- [9] Tyagi, Vivek. "Hybrid context dependent CD-DNN-HMM Keyword Spotting (KWS) in speech conversations." 2016 *Machine Learning for Signal Processing (MLSP)* (2016): 1-6.
- [10] World Health Organization, "Visual impairment and blindness". (2014, August). [Retrieved September 18, 2017], from <http://www.who.int/mediacentre/factsheets/fs282/en/>