https://jurnaleeccis.ub.ac.id/
e-ISSN(Online): 2460-8122

Jurnal EECCIS Vol. 17, No. 1, April 2023
pp 1-7

Text to Speech Using Finite State Automata on Health Data

Indrianto Indrianto [1], Abdurrasyid Abdurrasyid [1], Meilia Nur Indah Susanti [1], Givari Fairus Ferdiansyah Deu[1], Arief Ramadhan[2]
[1] Institut Teknologi PLN, West Jakarta, Indonesia, [2] School of Computing, Telkom University, Bandung, Indonesia
Email: arasyid@itpln.ac.id, arieframadhan@telkomuniversity.ac.id

***Abstract*— Computer science and engineering has provided many benefits that can be applied in our lives, such as in the field of automata theory, one of the largest areas related to the efficiency of an algorithm in solving problems in computational models. Text to Speech is a technology that converts text into sound using a phonetization system, phonemes that are arranged to form a speech to make computers able to communicate and interact with everyday spoken language. Data that need to be interpret is the health data such as body temperature, heart rate per minute, and oxygen levels. Text to Speech is very useful to be applied to blind aids who need information in the form of sound because of their limitations. For this reason, it is necessary to make an application that can read text-based data that is stored into a voice that can be heard by the blind, the method used in this study is Finite State Automata (FSA) which is used to split Indonesian words into words according to its syllable patterns and facilitate the pronunciation process which is included in the blind aids so that it is expected to help the visually impaired to be able to find out their health condition. in this study, the test was carried out using the Confusion Matrix method, while the results obtained were 97% accurate.**

***Index Terms*— Confusion Matrix, Finite State Automata, Text-to-Speech, visually impaired aids.**

***Abstrak*–- Ilmu dan teknik komputer telah memberikan banyak manfaat yang dapat diaplikasikan dalam kehidupan kita, seperti dalam bidang teori automata, salah satu bidang terbesar yang berkaitan dengan efisiensi sebuah algoritma dalam menyelesaikan masalah dalam model komputasi. Text to Speech merupakan teknologi yang mengubah teks menjadi suara dengan menggunakan sistem fonetisasi, fonem-fonem yang disusun membentuk sebuah ucapan untuk membuat komputer dapat berkomunikasi dan berinteraksi dengan bahasa yang diucapkan sehari-hari. Data yang perlu diinterpretasikan adalah data kesehatan seperti suhu tubuh, detak jantung per menit, dan kadar oksigen. Text to Speech sangat berguna untuk diterapkan pada alat bantu tunanetra yang membutuhkan informasi dalam bentuk suara karena keterbatasannya. Untuk itu perlu dibuat suatu aplikasi yang dapat membaca data berbasis teks yang disimpan menjadi sebuah suara yang dapat didengar oleh tunanetra, metode yang digunakan pada penelitian ini adalah Finite State Automata (FSA) yang digunakan untuk memecah kata bahasa Indonesia menjadi kata-kata sesuai dengan pola suku katanya dan memudahkan proses pengucapannya yang dimasukkan ke dalam alat bantu tunanetra sehingga diharapkan dapat membantu para tunanetra untuk dapat mengetahui kondisi kesehatan mereka. pada penelitian ini dilakukan pengujian dengan metode Confusion Matrix, adapun hasil yang didapat adalah 97% akurat.**

***Kata Kunci*— Alat bantu tuna netra, Confusion Matrix, Finite State Automata, Text-to-Speech.**

I. INTRODUCTION

The development of knowledge and technology is making human work relatively easier, because the technology created comes from the background of problems and anxiety that arise from the users themselves. Computer science and engineering that exists today has provided many benefits that can be applied in our lives, one of which we can see is in the field of automata theory, one of the largest areas related to the efficiency of an algorithm in solving problems in computational models.

In recent years, many researchers have created and developed applications to help people who have visual impairments when they want to communicate with other people, one of which is technology from Text-to-Speech (TTS). The technology is converting text into sound using a phonetization system, that is, phonemes that are arranged to form an utterance. The purpose of this technology is to make computers able to communicate and interact with everyday spoken language. In the configuration process, the words are split into syllables using Indonesian language rules with the aim that the resulting sound can be heard as Indonesian language.

Researchers have done a lot of research that makes the visually impaired as the object of their research, but the most research done is more on walking aids for the blind using computer vision [1], [2], GPS[3], [4], and ultrasonic[5], [6] however, there is only some research that discusses how the health condition of a blind person can be conveyed from text data into voice to blind people. This is considered very useful, as evidenced by research conducted that people with disabilities in reading are greatly helped by the presence of text-to-speech application[7].

Many methods are widely used for text processing, the one that widely used is the Finite state automata method. Researchers have used this method to detect syllables in bahasa Indonesia[8], to read the prefix of a word [9], read and translate bahasa Indonesia to Madurese language[10], and Latin to Sundanese[11]. The method is not only used to help translate text but can also be used to make predictions on time series data [12], verification test[13], and reading document similarity[14].

It is proven that this method has sufficient accuracy with a value of more than 80% [10], [11], [15], however, it still has limitations where it only processes text but not many converts it into speech, and the data used does not yet use health data, such as body temperature, heart rate, and oxygen levels. Some additional time information in

the form of date, month, year, hour and minute are also not added there. This will certainly be very useful if applied to a tool that is used by the blind people so as to facilitate their accessibility in knowing their health condition.

The rest of this paper is structured as follows. The literature review, which outlines the idea behind the study approach, is presented in Section 2. A research strategy is presented in Section 3. Section 4 presents and discusses the experiment's results. Section 5's conclusions and work-related suggestions round up the section.

II. LITERATURE REVIEW

A. Text - to - Speech

A Text-to-speech system (TTS) or commonly called Text-to-speech Synthesis is a computer-based system that can read text aloud automatically, whether the text is introduced by a computer input stream, or a scanned input which is sent to Optical Character Recognition (OCR) machine. Speech synthesizers can be implemented by both hardware and software and have made very rapid improvements over the decades and many high quality TTS systems are now available for commercial use.

Speech which is often based on natural speech sequences i.e., units taken from natural speech and put together to form words or sentences of simultaneous speech synthesis, the latter has become very popular in recent years due to its increased sensitivity to the context of the unit over its simpler predecessors. Rhythm is an important factor in making speech synthesized from the TTS system to be more natural and understandable, the prosodic structure which provides important information to produce a prosodic generation model that is the effect in the synthesized speech. Many TTS systems were developed based on the principle of corpus-based speech synthesis due to the natural sound output, high quality, and very popular.

B. Finite State Automata

As an abstract mathematical concept that describes the behavior of a logical machine that explains the workings of a physical machine, a program, an algorithm, or a problem-solving conception. In the context of language theory, the FSA engine can be applied to recognize a string that comes from a regular language that is generated from a regular grammar. Thus, there is a reciprocal relationship between a regular language and an FSA, that is, if it is owned by a regular language, a language can be constructed by an FSA machine, and then, if it is owned by an FSA, a language will be derived and can be recognized by the machine.

The Finite State Machine can be a machine that has no output. A finite state machine that does not issue this output is known as a FINITE STATE AUTOMATA (FSA). In FSA the machine is initially in state S0 and receives a series of inputs which can change it to the next states. In FSA, there is also a certain set of states known as the FINAL STATE. Changes from one state to the next follow certain rules that are formulated as a transition function.

FSA is an automatic machine of regular language. An FSA has a finite number of states and can move from one state to another. This state change is represented by a transition function. The FSA has no storage space, so the ability to 'remember' is limited, it can only remember the most recent state. Examples of FSAs include elevators, text editors, lexical analysis, network communication protocols, and parity checks.

Formally the FSA can be defined as TUPLE-5 : A collection of 5 sets, or annotated as :

$\boldsymbol{FSA\ is\ M} = \boldsymbol{S}, \boldsymbol{\Sigma}, \boldsymbol{\delta}, \boldsymbol{S0\ dan\ F}$ Where :

S = finite set of states

Σ = finite set of symbols on the machine

$\boldsymbol{\delta}$= Q x **Σ** is a transition function that governs the movement of the machine. Among them is a function that takes states and an input alphabet as arguments and returns a state.

S0=Initial state
F = set of FINAL state

The behavior of Finite State Automata is expressed in the form of a transition table or in the form of a transition diagram.

Below is an example of syllable breaking from Finite State Automata itself. The transition function in the transition table is as follows:

TABLE 1.
TRANSITION TABLE

Transition Table		
	δ	
State	**Input**	
	0	**1**
S0	S0	S0
S1	S1	S1
S2	S2	S2
S3	S3	S3

where from the transition table above can be described the FSA transition diagram as figure 1 below.

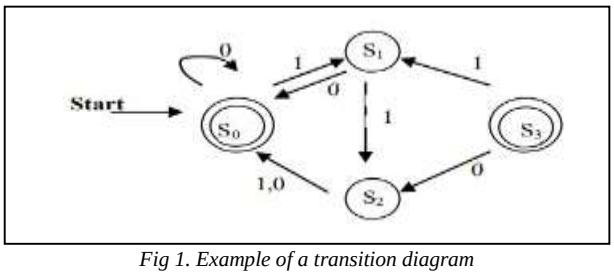

Fig 1. Example of a transition diagram

III. RESEARCH METHOD

A. Research Framework

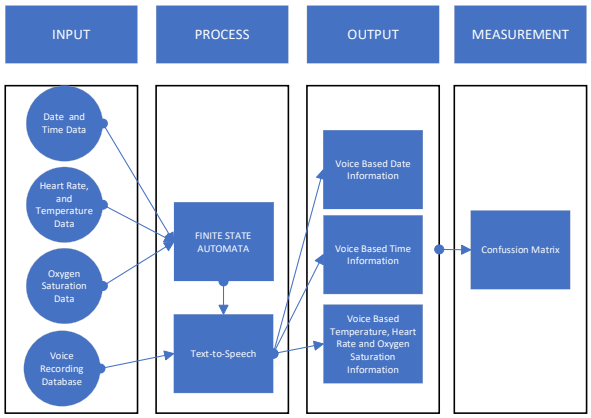

Fig 2. Research Framework

The framework of thought in the research is divided into four parts, the first is data input, the second part is the process, the third is the output produced, and the last is the measurement. For the first, there are five data inputs, first the date and time data, heart rate, and temperature data, and oxygen saturation data which will then be processed by the finite state automata method, next there is voice recording data that will be used after the finite state process automata is translated into sound.

The third part is the output generated in the form of voice date and time, as well as the sound of heart rate, temperature, and oxygen saturation data. The last part is the testing section where testing is carried out using a confusion matrix to find out how accurate the finite state automata method is in recognizing and processing input. It is necessary to know the data obtained can come from sensor input stored in the database

B. Application of Finite State Automata

The Finite State Automata method is used to separate the words that are inputted and will be sent to Text-to-Speech (TTS) to be processed and outputted with sound results. This method consists of 4 stages that need to be known before processing with FSA, which are explained as follows.

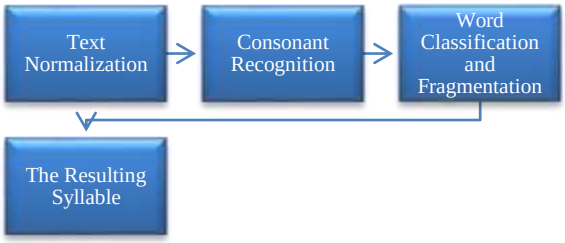

Fig 3. FSA stage.

Text Normalization, in this step every sentence text containing numbers, currency units, symbols, time, date, temperature, units and abbreviations will be carried out in the text normalization process first, which in the table below is explained as follows.

TABLE 2.
TEXT NORMALIZATION

Text	Normalization Result
0	Nol
1	Satu
2	Dua
3	Tiga
4	Empat
5	Lima
6	Enam
7	Tujuh
8	Delapan
9	Sembilan
10	Sepuluh
N > 19 & < 100	Puluh
11	Sebelas
N > 99 & < 1000	Ratus
100	Seratus
N > 999 & < 10000	Ribu
1000	Seribu
℃	Derajat celsius
%	Persen

Consonant recognition, the next step the writer performs the process to recognize letters after there is text input given from the tool. Besides letter recognition, there will also be an introduction to space punctuation marks. The letters B, C, D, F, G, H, J, K, L, M, N, P, Q, R, S, T, V, W, X, Y, and Z will be recognized as "K" or consonant. The letters A, I, U, E, O will be recognized as "V" or vowel. As for the letters N, Y, G will be recognized as the letter itself, those are: N as "N", Y as "Y" and "G" as "G". This arrangements will later aims to facilitate the classification of syllables if later in the reading of the text there are consecutive consonants presents.

Word classification and fragmentation, in this step the writer performs the classification and fragmentation of words described in the form of a Transition Diagram designed in three levels. At the first level that is recognized is the pattern: V, K or KV. The results of the first level itself will be a continuation to the next level of FSA.

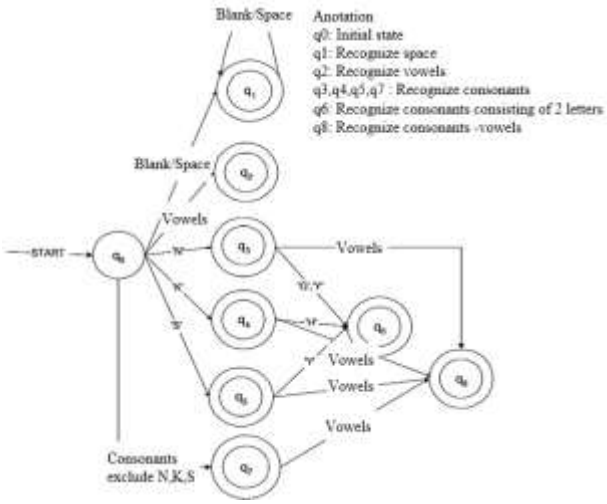

Fig 4. First level FSA transition diagram

Next is the level of the FSA transition diagram process, which is the second level which will recognize syllables with the pattern V, VK, VKK, KV, KVK, KKV, KKVK, KKKV, KKKVK for all consonants other than n, k, s.

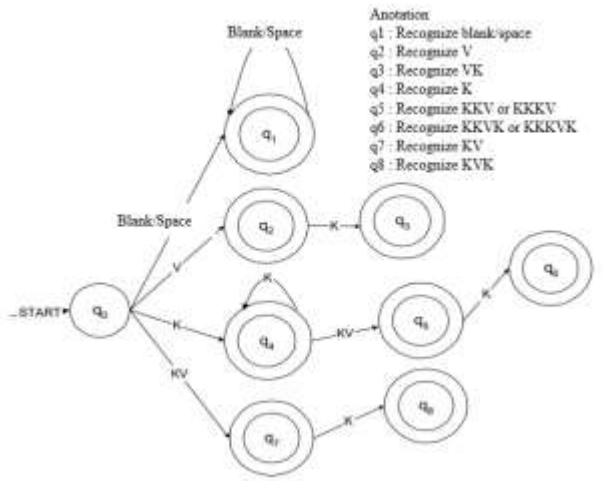

Fig 5. Second level FSA transition diagram.

Then, at the next level, the third level, it was explained that the syllable pattern of VKK, KVKK and KKVKK could not be recognized at the previous level. Therefore, the third-tier FSA can recognize these syllables.

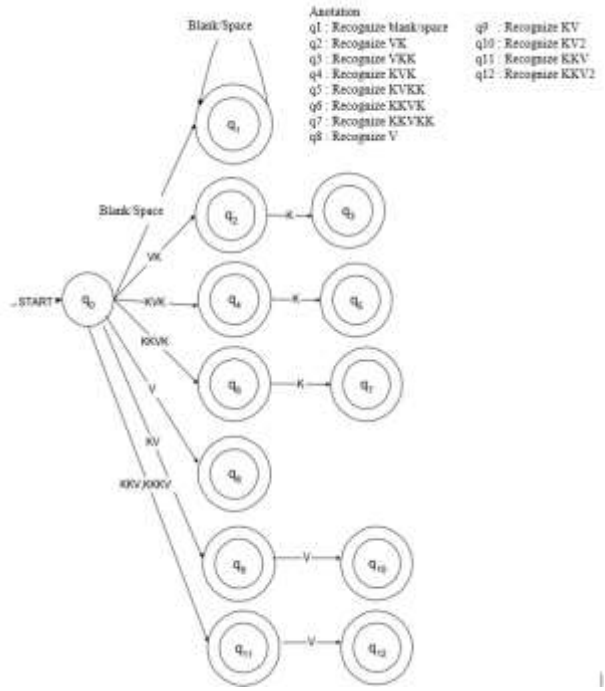

Fig 6. Third level FSA transition diagram.

The process of the three transition diagrams above explains how the process of producing 12 syllable/phoneme classifications from words that have been cut off according to Indonesian sentences. The goal is to recognize syllables in Indonesian sentences, by recognizing syllables in spoken language, it can be implemented into TTS.

C. Text-to-Speech Conversion

So that text can be converted into sound, voice recording and diphone concatenation are carried out.

a. At the voice recording stage, the writer performs a manual recording process by recording the syllables that will be used in TTS. There are recordings that are absolute and not absolute, recordings that are absolute are syllables that have been determined by the writer and not absolute, such as numbers and months, all stored in mp3 format.
b. In the diphone concatenation technique, this stage works by combining sound segments that have been previously recorded. Each segment is a diphone (a combination of two phonemes). The formation of the final speech of this technique is by arranging the appropriate diphone so that it gets the expected results. Below is an example of the formation of words or speech such as "TEMP" which is composed of diphone-diphone.

IV. RESULTS AND DISCUSSION

A. FSA Result

The results of the finite state automata in this study are divided into three parts, the first part produces the sound of the date, the second produces the sound of time, and the third produces the sound of temperature, oxygen saturation, and heart rate.

1. Date voice
   Known below :
   FSA is M = (S, Σ, δ , S0 and F) where
   S = { S0, S1, S2, S3, S4, S5}
   Σ = {Blank/Space, KV, V, KVK, KVKK}
   δ = (S0, Blank/Space) = S1, (S0, KV) = S2, (S0, V) = S3, (S0, KVK) = S4, (S4, K) = S5
   S0 = Initial State
   F = {S0, S5}

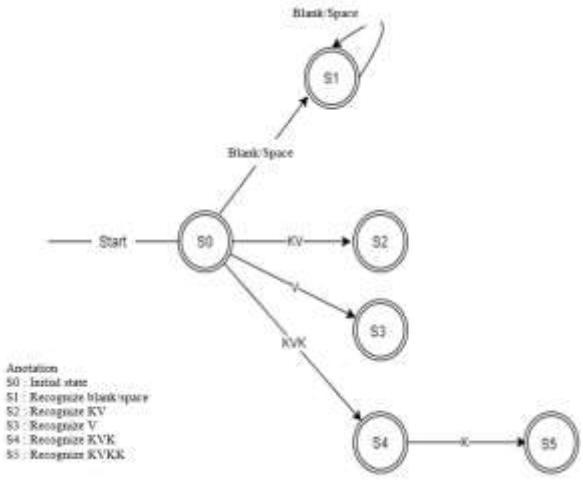

Fig 7. Date FSA transition diagram

Figure 7 above is written in table 3 below

TABLE 3.
DATE FSA TABLE TRANSITION

State	Input					
	Blank/Space	**KV**	**V**	**KVK**	**KVKK**	**K**
S0	S1	S2	S3	S4	S5	-
S1	S1	-	-	-	-	-
S2	-	-	-	-	-	-
S3	-	-	-	-	-	-
S4	-	-	-	-	-	S5
S5	-	-	-	-	-	-

The sounds that are processed in Figure 7 and Table 3 above, an example of the output that will come out is "HARI INI TANGGAL TIGA BULAN AGUSTUS TAHUN DUA RIBU DUA PULUH SATU". This sentence has been normalized in the form of numbers according to the previous step. If the word fragments described in the above process are described in "HA-RI", "I-NI", "DATE-GAL", "TI-GA", "BU-LAN", "A-GUS-TUS", "TA- HUN", "DU-A", "RI-BU", "DU-A", "PU-LUH", "SA-TU", in which each word will be separated according to the Indonesian language rules in the previous step.

2. Time voice
   Known below :
   FSA is M = (S, Σ, δ , S0 and F) where :
   S { S0, S1, S2, S3, S4, S5}
   Σ = {Space/Blank, KV, KVK, KVKK, V}
   δ = (S0, Blank/Space) = S1, (S0, KV) = S2, (S0, KVK) = S3, (S3, K) = S4, (S0, V) = S5
   S0 = Initial State
   F = {S0, S5}

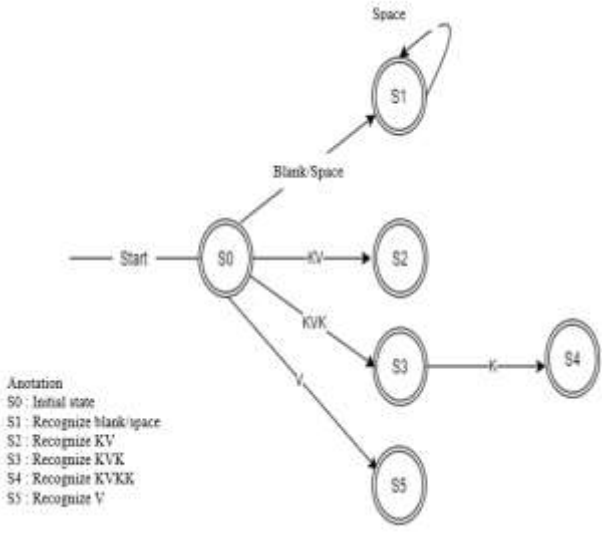

Fig 8. Time FSA transition diagram

Figure 8 above is written in table 4 below

TABLE 4.
TIME FSA TABLE TRANSITION

State	Input					
	Blank/Space	**KV**	**KVK**	**K**	**V**	**KVKK**
S0	S1	S2	S3	S4	S5	S4
S1	S1	-	-	-	-	-
S2	-	-	-	-	-	-
S3	-	-	-	-	-	-
S4	-	-	-	-	-	S5
S5	-	-	-	-	-	-

The sounds that are processed in Figure 8, and Table 4 above, an example of the output that will come out is "SEKARANG JAM DUA BELAS NOL NOL". This sentence has been normalized in the form of numbers according to the previous step. If the decapitation is described in the above process, "SE-KARANG", "JAM", "DU-A", "BE-LAS", "NOL", "NOL". In which each word will be separated according to the Indonesian language rules in the previous step.

3. Health data voice
   Known below :

   FSA is M = (S, Σ, δ , S0 and F) where :
   S = {S0,IS1, S2, S3, S4, S5, S6,}
   Σ = {Space/Blank, KV, VK, V, KVK, KVKK}
   δ = (S0, Blank/Space) = S1, (S0, KV) = S2, (S0, KVK) = S3, (S3, V) =
   S4, (S4, K) = S5, (S3, K) = S6
   S0 = Initial State
   F = {S0, S6}

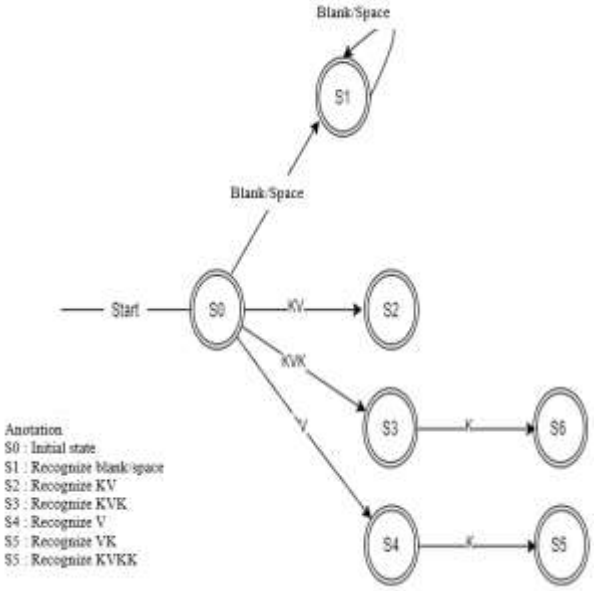

Fig 9. Time FSA transition diagram

Figure 9 above is written on table 5 below

TABLE 5.
TIME FSA TABLE TRANSITION

State	Input						
	Blank/Space	**KV**	**KVK**	**V**	**VK**	**K**	**KVKK**
S0	S1	S2	S3	S4	S5	-	S6
S1	S1	-	-	-	-	-	-
S2	-	-	-	-	-	-	-
S3	-	-	-	-	-	S6	-
S4	-	-	-	-	-	S5	-
S5	-	-	-	-	-	-	-
S6	-	-	-	-	-	-	-

The sound that is processed in Figure 9, and Table 5 above, an example of the output that will come out is "SUHU BADAN TIGA PULUH ENAM DERAJAT CELSIUS DETAK JANTUNG DELAPAN PULUH ENAM BIT PER MENIT KADAR OKSIGEN SEMBILAN PULUH PERSEN". This sentence has been normalized in the form of the appropriate number previous step. If the decapitation is described in the above process "SU-HU, "BA-DAN", "TI-GA", "PU-LUH", "E-NAM", "DERA-JAT", "CEL-SI-US" , "DE-TAK", "JAN-TUNG", "DE-LA-PAN", "PU-LUH", "E-NAM", "BIT", "PER", "ME-NIT", "KA-DAR", "OK-SI-GEN", "SEM-BI-LAN", "PU-LUH", "PER-SEN", in which each word will be separated according to the Indonesian language rules in the previous step.

B. Measurement

The test results from the Confusion Matrix with the writer's reference to find the accuracy value were tested with FSA as many as 15 trials for each word tested in the form of percentage accuracy from 0 to 100%, can be seen in table 6 below

TABLE 6.
CONFUSION MATRIX RESULT

Actual/Prediction Tested Voice	TP	TN	FP	FN
Date voice	174	0	6	0
Time voice	90	0	0	0
Health data voice	310	0	23	7

Table 6 above shows that out of a total of 630 tests carried out on 3 types of tests, there were 36 failures so that the results of the tests can be seen in figure 10 below.

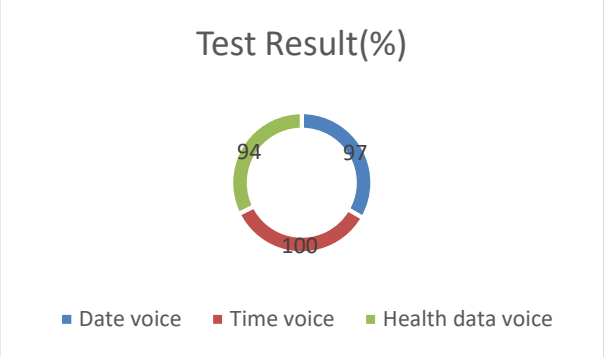

Fig 10. Time FSA transition diagram.

Figure 10 above shows from 15 trials, the average of the accuracy obtained reached 97%, which means that all can be detected properly, and the sound produced is appropriate.

V. CONCLUSION

The use of the Finite State Automata (FSA) method applied to Text-to-Speech (TTS) in terms of processing to recognizing or capturing and cutting words into syllable patterns according to Indonesian rules can be an alternative for the pronunciation process on the Blind Vision Wrist tool. Furthermore, the FSA can read any normalized symbol and input text given from the tool or web application.

The application of Finite State Automata (FSA) on the Blind Vision Wrist tool in terms of cutting words into syllables has been successful and tested using the Confusion Matrix to get appropriate results both from absolute and not absolute words, by normalizing text changing numbers and symbols into form of text, introduction of vowel consonants in words, classifying and splitting words and getting syllable results based on the processed FSA Transition Diagram.

By testing the Finite State Automata (FSA) method using the Confusion Matrix to find the results of the accuracy values in the method, the accuracy rate is 97%.

ACKNOWLEDGMENT

We appreciate the support of the KEMDIKBUDRISTEK for funds provided for applied research schemes with no contract 155/ES/PG.02.00/PT/2022 so that this research can be carried out properly.

REFERENCES

[1] Abdurrasyid, Indrianto, and M. N. I. Susanti, "Face detection and global positioning system on a walking aid for blind people," *Bull. Electr. Eng. Informatics*, vol. 11, no. 3, pp. 1558–1567, 2022.

[2] A. Abdurrasyid, I. Indrianto, and R. Arianto, "Detection of immovable objects on visually impaired people walking aids," *TELKOMNIKA (Telecommunication Comput. Electron. Control.*, vol. 17, no. 2, p. 580, 2018.

[3] B. Kuriakose, R. Shrestha, and F. E. Sandnes, "Tools and Technologies for Blind and Visually Impaired Navigation Support : A Review Tools and Technologies for Blind and Visually Impaired Navigation Support :," *IETE Tech. Rev.*, 2022.

[4] J. Tohap and M. Nababan, "Development of Training Aids (Remote Control and Headset) for Tunanetra Sprint Athletes," in *Unimed International Conference on Sport Science*, 2020, vol. 23, no. UnICoSS 2019, pp. 167–170.

[5] F. Shaikh, M. A. Meghani, V. Kuvar, and P. S. Pappu, "Wearable navigation and assistive system for visually impaired," in *International Conference on Trends in Electronics and Informatics (ICOEI)*, 2018, no. Icoei, pp. 747–751.

[6] Abdurrasyid, R. Arianto, I. Indrianto, and B. A. Nugroho, "The Obstacles Detector with Tahani Fuzzy Logic as The Tool for Blind People," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 9, no. 2, p. 72, 2018.

[7] S. G. Wood, J. H. Moxley, E. L. Tighe, and R. K. Wagner, "Does Use of Text-to-Speech and Related Read-Aloud Tools Improve Reading Comprehension for Students With Reading Disabilities ? A Meta-Analysis," *J. Learn. Disabil.*, pp. 1–12, 2017.

[8] H. Haryanto and Aripin, "A Finite State Machine Model to Determine Syllables of Indonesian Text," *2019 1st Int. Conf.*

Cybern. Intell. Syst. ICORIS 2019, no. August, pp. 238–241, 2019.

[9] R. Singh and D. G. Goyal, "Algorithm Design for Deterministic Finite Automata for a Given Regular Language with Prefix Strings," *J. Sci. Res.*, vol. 66, no. 2, pp. 16–21, 2022.

[10] F. H. Rachman, Qudsiyah, and F. Solihin, "Finite State Automata Approach for Text to Speech Translation System in Indonesian-Madurese Language," *J. Phys. Conf. Ser.*, vol. 1569, no. 2, pp. 0–7, 2020.

[11] C. Slamet, Y. A. Gerhana, D. S. Maylawati, M. A. Ramdhani, and N. Z. Silmi, "Latin to Sundanese script conversion using Finite State automata algorithm," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 434, no. 1, pp. 0–10, 2018.

[12] U. Pavlova and A. Rakitskiy, "Development and Research of the Time Series Prediction Method Based on Finite State Automaton," in *Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology*, 2022, pp. 2021–2023.

[13] J. Barcelos, C. Basilio, J. Raphael, J. Barcelos, and C. Basilio, "New predictability verification tests for discrete-event systems modeled by finite state automata," *Int. Fed. Autom. Control*, vol. 53, no. 4, pp. 243–249, 2020.

[14] M. Abusafiya, "Measuring Documents Similarity using Finite State Automata," no. 2, pp. 2020–2023, 2020.

[15] Z. Li, H. Derksen, J. Gryak, C. Jiang, Z. Gao, and W. Zhang, "Biomedical Signal Processing and Control Prediction of cardiac arrhythmia using deterministic probabilistic finite-state automata," *Biomed. Signal Process. Control*, vol. 63, p. 102200, 2021.