TRANSFORMING SPEECH THERAPY FOR CHILDREN WITH AUTISM THROUGH A VIRTUAL APPLIED BEHAVIOR ANALYSIS APPROACH

MASERI, M.[1] – MAMAT, M.[1*] – YEW, H. T.[1] – ZAKARIA, Z.[2]

[1] *Faculty of Engineering, Universiti Malaysia Sabah, Sabah, Malaysia.*

[2] *Faculty of Social Sciences and Humanities, Universiti Malaysia Sabah, Sabah, Malaysia.*

**Corresponding author*
e-mail: mazlina[at]ums.edu.my

(Received 13th July 2024; revised 11th October 2024; accepted 19th October 2024)

Abstract. Speech therapy is essential for children with Autism Spectrum Disorder (ASD), many of whom face challenges in developing functional speech and language skills. Traditional therapy approaches require significant resources and may not be accessible to all individuals in need. This gap necessitates exploring alternative, technology-assisted methods for delivering speech therapy. Although technological tools for interventions have advanced, there is still a lack of tools specifically tailored to children with ASD that integrate evidence-based behavioral therapy principles. Hence, this study developed a Virtual Speech Therapy System (VSTS) that employs Applied Behavior Analysis (ABA) principles, particularly Discrete Trial Training (DTT), within a digital framework to provide speech therapy in the Malay language. The speech therapy is established through a text-to-speech unit, a speech-recognition unit based on the Hidden Markov Model Toolkit (HTK), and a performance counter unit to measure progress. The VSTS passed the acceptance testing, and the speech-recognition unit demonstrated a promising percentage of accuracy and a relatively low word error rate during development, with 93.75% and 6.25%, respectively. It also showed a higher recognition rate for words with distinct phonetic compositions. This study contributes to the educational development of children with ASD, particularly in verbal communication through assistive technology.
Keywords: *assistive technology, verbal communication, discrete trial training, speech recognition*

Introduction

Individuals with Autism Spectrum Disorder (ASD) often struggle with social communication skills such as starting conversations, responding to others' communicative bids, and engaging in reciprocal conversations (Girolamo et al., 2024; Qin et al., 2024; Hodges et al., 2020). Some have little or no functional speech or have not developed language by the time they are teens. Traditional speech therapy, though effective, is not always accessible due to geographic and financial constraints. Thus, technology-based interventions, such as mobile apps, could offer a cost-effective and accessible communication therapy for them. Several mobile apps have been developed to improve communication among children with ASD, like IMPUTE ADT-1 (Panda et al., 2024), Listening2Faces (Baron et al., 2024), and Aseel (Zibin et al., 2023), to name a few. While these apps use prompting and structured learning, none incorporate the evidenced-based Applied Behaviour Analysis (ABA) principle, which is known as a gold standard in ASD treatment (Adelson et al., 2024).

This paper introduces a new educational approach to enhance verbal communication in children with ASD. This study represents an advancement in assistive technology by integrating Applied Behavior Analysis (ABA) principles, specifically Discrete Trial Training (DTT), into a digital Malay language speech therapy system. The primary

eISSN: 2716-6481
https://doi.org/10.55197/qjssh.v5i6.471

objective is to demonstrate the feasibility and efficacy of digitizing traditional DTT to provide more accessible and scalable ABA-based speech therapy for children with ASD. This aligns with a study by Parsons et al. (2015), which agrees that technology can help to improve communication and foster greater inclusivity. Additionally, this paper aims to encourage researchers from diverse fields to investigate the potential and limitations of digitized DTT. We believe this exploration will significantly contribute to developing more practical and accessible assistive technology solutions, ultimately benefiting the community in need.

Literature review

ABA is the scientific application of principles and techniques derived from behavioral psychology to produce meaningful and socially significant changes in behavior. ABA can be administered at varying intensities and is categorized into comprehensive or focused treatment approaches (Maharjan et al., 2023; Wergeland et al., 2022). Generally, an ABA program has seven elements. First, the targeted or chosen behaviors must be applicable or have functional significance. Second, the behaviors must be observable so that their performance can be measured and recorded. Third, the program must involve analysis of results (i.e., behavioral improvements because of the program), which includes baseline measures (no intervention) and intervention measures. Fourth, the techniques must be documented so that another person can replicate the program. Fifth, the established principles must be included in the program. Sixth, the program should result in behavior changes and improve the person's quality of life. Seventh, changes in behavior produced during the program should be generalized to other situations and environments.

The ABA involves a systematic design, implementation, and evaluation of instructional and environmental modifications by a behavior analyst. The process includes functional assessment and analysis, empirically identifying the relationships between behavior and environmental factors. These ABA interventions utilize related factors, establishing operations, antecedent stimuli, positive reinforcement, and other consequences to help individuals develop improved behaviors to a meaningful degree (Yanchik et al., 2024). An ABA technique known as Discrete Trial Training (DTT) is a teaching method that employs repetition and sequenced instruction to help individuals with ASD develop a wide range of skills (Lovaas, 1987). In DTT, each skill being taught is broken down into several steps and built up using discrete trials that teach each step individually. Each trial has its own set of clearly defined and scripted steps and must always be followed. DTT-based intervention can be given to children as early as two years old and can last anywhere from two to six years, depending on the child's performance. DTT remains very useful for teaching a wide variety of skills to children with ASD, as its implementation has proven to maximize children’s success and minimize their failures during the intervention (Cardinal et al., 2017). DTT can boost a child's motivation and learning through its discrete and clear approach. Because each trial is brief, many teaching trials can be completed, providing numerous learning opportunities.

The DTT consists of five stages: Antecedent, Prompt, Response, Consequence, and Inter-trial interval (Smith, 2001), defined as follows: *Antecedent*. A discriminative stimulus is formally referred to as a Cue or Antecedent. The instructor gives a quick, specific instruction or question, such as "Do this" or "What is it?" in the classic DTT application. *Prompt*. The instructor helps the student respond appropriately to the cue

eISSN: 2716-6481
https://doi.org/10.55197/qjssh.v5i6.471

either concurrently with or after it. For instance, the instructor could hold the child's hand and instruct them on how to respond, or she could act out the response herself. In order to train the kid to respond to the cue alone, the instructor gradually reduces and then eliminates the prompt as the child advances (e.g., by coaching the student through less and less of the response). *Response*. The child gives a correct or an incorrect answer to the instructor's cue. *Consequence*. The consequences could be a Correction or a Reinforcement based on the child's response. For Correction, the instructor promptly encourages the child's response if it is correct by giving hugs, praise, nibbles of food, access to toys, or other enjoyable activities. For Reinforcement, the instructor signals that the response was inappropriate in some other way, such as by looking away, saying "No," taking away teaching materials, or removing the child from the classroom. *Inter-trial interval*. The instructor waits for a brief (1-5 second) period following the consequence before providing the signal for the following trial.

Materials and Methods

A digitized DTT was invented to be incorporated into VSTS. In the digitized DTT, the interactions between instructor and child were replaced with a Text-to-speech unit and a Speech-recognition unit. A Performance counter unit was also included to decide the Consequences. A Timer unit is created to measure the time for Inter-trial Intervals. *Table 1* defines the adaptation from the traditional to the Digitized DTT.

***Table 1**. The traditional and Digitized DTTs.*

Step	Action	Traditional	Digitized
Antecedent	Vocal output	Instructor	Text-to-speech unit
	Vocal input	Child	Child
Prompt	Vocal output	Instructor	Text-to-speech unit
	Vocal input	Child	Child
Response	Vocal output	Child	Child
	Vocal input	Instructor	Speech-recognition unit
Consequence	Interpret answer	Instructor	Performance-counter unit
	Vocal output	Instructor	Text-to-speech unit
	Vocal input	Child	Child
Inter-tial interval	Count time	Instructor	Timer-unit
	Result	Instructor	Performance-counter unit

Text-to-speech unit

The Text-to-speech unit represents the instructor's vocal instruction. The instruction texts are pre-defined in the learning module. *Figure 1* defines the operation flowchart of the Text-to-speech unit. Every input text will be analyzed and identified if it is similar to any reference words from the local database. If any match is found, the unit will use the pre-recorded speech of the reference word as the output speech. The speech rate is set to 0.8, slightly lower than the standard speech rate of 1.0. This is to allow children with ASD to hear and understand the instructions.

eISSN: 2716-6481
https://doi.org/10.55197/qjssh.v5i6.471

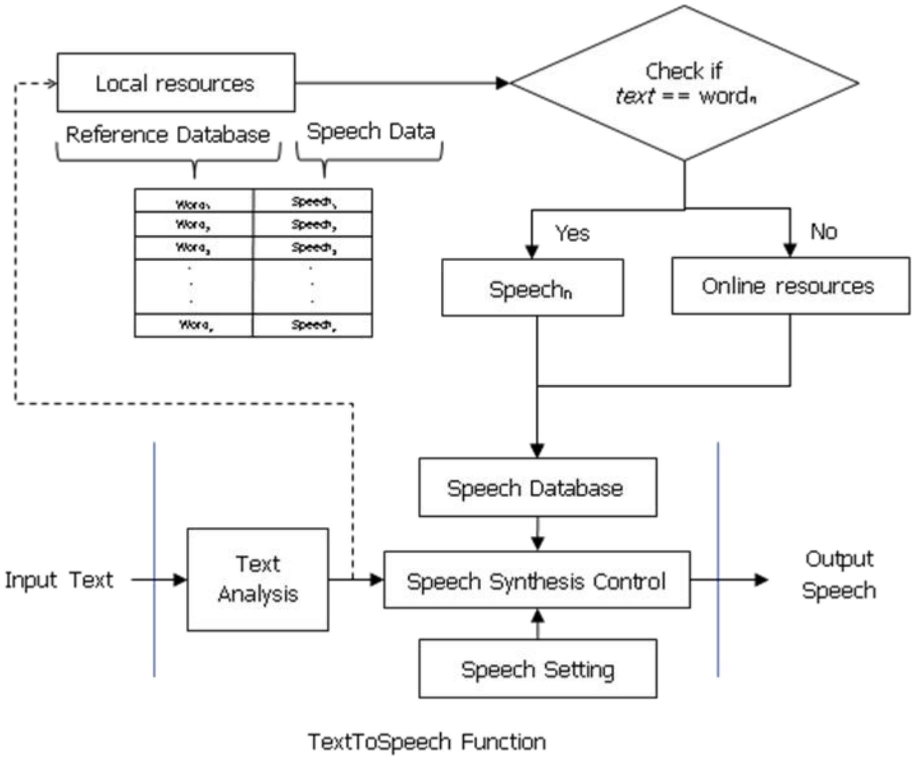


Figure 1. *Flowchart of the computer program for the Text-to-speech unit.*

Speech-recognition unit

A speech-recognition unit was created to interpret and identify the child's vocal input (input speech). The speech-recognition unit was developed using the Hidden Markov Model Toolkit (HTK), created by the Engineering Department at the University of Cambridge (CUED). HTK is specifically designed for building Hidden Markov Models (HMMs) for speech processing (Cambridge University Engineering Department, 2017). The development process starts with gathering and preprocessing audio data, followed by labeling and creating transcription files, preparing training data, and training and testing the speech recognizer. Each speech signal undergoes preprocessing steps, including end-point detection (EPD), pre-emphasis, and feature extraction (Maseri and Mamat, 2020; 2019).

Timer unit

The timer unit counts the time for pauses for the Inter-trial Interval. In this system, the countdown timer function is used, and the interval is 5 seconds. The operation of the timer is described in *Figure 2*.

eISSN: 2716-6481
https://doi.org/10.55197/qjssh.v5i6.471

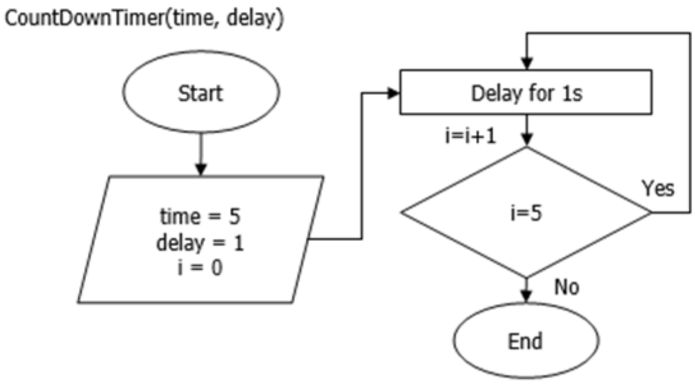


Figure 2. *Flowchart of the computer program for the Timer unit.*

Performanc e-counter unit

A performance counter unit was developed to identify the response and calculate the result at the end of a lesson. This unit will record the child's performance based on the flowchart in *Figure 3*. The count of "result" will increase by one (+1) for each accurate response. If the child responds incorrectly, the system will prompt a similar question up to three times. The system will ask the following question following the third trial. The total result is calculated by dividing the number of "result" by the total number of questions ("noQue").

eISSN: 2716-6481
https://doi.org/10.55197/qjssh.v5i6.471

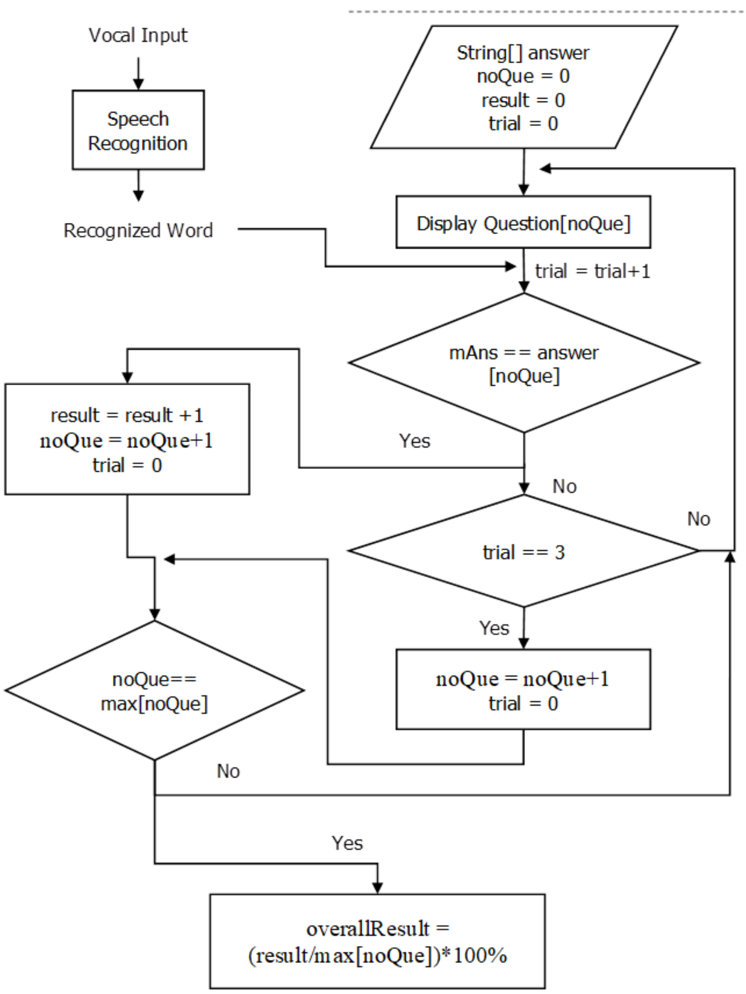


Figure 3. *Flowchart of the computer program for the Performance-counter unit.*

Lesson framework

The lesson incorporates ten Malay numerals: kosong (zero), satu (one), dua (two), tiga (three), empat (four), lima (five), enam (six), tujuh (seven), lapan (eight), and sembilan (nine), along with six body parts: telinga (ear), mata (eye), tangan (hand), kaki (foot), mulut (mouth), and hidung (nose) (Maseri and Mamat, 2019). *Table 2* displays the structure and phoneme sequence of each word. The words in this corpus are composed of syllables containing vowels (V) and consonants (C). The identified syllable structures include: (i) V-2, (ii) CV-17, (iii) VC-2, (iv) CVC-8, (v) CCV-1, (vi) CCVC-1, and (vii) CVCC-1. The corpus consists of 33 syllables: {/sa/, /tu/(2), /du/, /a/, /ti/, /ga/, /em/, /pat/, /li/(2), /ma/, /e/, /nam/, /juh/, /la/, /pan/, /sem/, /bil/, /an/, /ma/, /ta/(2), /hi/, /dung/, /mu/, /lut/, /ngan/, /ka/, /ki/, /te/, /li/, and /nga/}. The lesson is delivered by following the digitized DTT approach as depicted in *Figure 4*. First, in the "Antecedent" stage, the Text-to-speech unit will produce an audio output, "What is it?" and display an image. For example, for the word "One", the picture of the number one will be shown. The second stage is "Prompt," whereby the audio output "One" will be produced. The prompt encourages the User (Child) to respond with the correct answer (speech input). Third, the Speech-recognition unit is activated to get the "Response".

eISSN: 2716-6481
https://doi.org/10.55197/qjssh.v5i6.471

After five seconds, the speech-recognition unit will begin recognizing the speech input, which activates the fourth step, "Consequence". The "Consequence" step will be one of the following: (1) If the child has responded correctly, an audio output "Good job" is given to praise the child; (2) If the child has given an incorrect or non-response, an audio output stating the correction for the answer (e.g., "One, it is Number One") is provided, followed by a neutral statement, "Let us try again". Finally, the "Inter-trial interval" will be activated, whereby the system will pause for 5 seconds (counted by a timer unit) before proceeding with the subsequent trial or lesson. A performance counter unit will count the correct or incorrect responses and update the learning performance.

***Table 2**. Speech lexicon for the selected Malay words.*

No	Word	Phoneme	Structure	Syllables
1	Kosong	/ko song/	CV+CVCC	2
2	Satu	/sa tu/	CV+CV	2
3	Dua	/du a/	CV+V	1
4	Tiga	/ti ga/	CV+CV	2
5	Empat	/əm pat/	VC+CVC	2
6	Lima	/li ma/	CV+CV	2
7	Enam	/ə nam/	V+CVC	2
8	Tujuh	/tu juh/	CV+CVC	2
9	Lapan	/la pan/	CV+CVC	2
10	Sembilan	/sem bi lan/	CVC+CVC+VC	3
11	Telinga	/te li nga/	CV+CV+CCV	3
12	Mata	/ma ta/	CV+CV	2
13	Tangan	/ta ngan/	CV+CCVC	2
14	Kaki	/ka ki/	CV+CV	2
15	Mulut	/mu lut/	CV+CVC	2
16	Hidung	/hi dung/	CV+CVCC	2

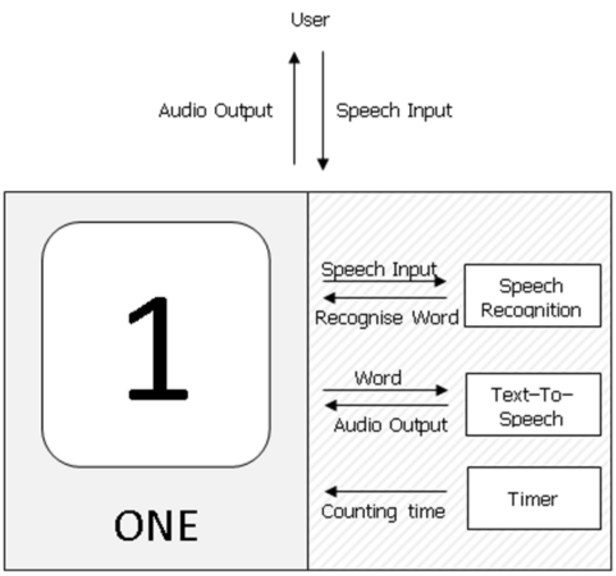


***Figure 4.** Lesson infused with the digitized DTT.*

eISSN: 2716-6481
https://doi.org/10.55197/qjssh.v5i6.471

Results and Discussion

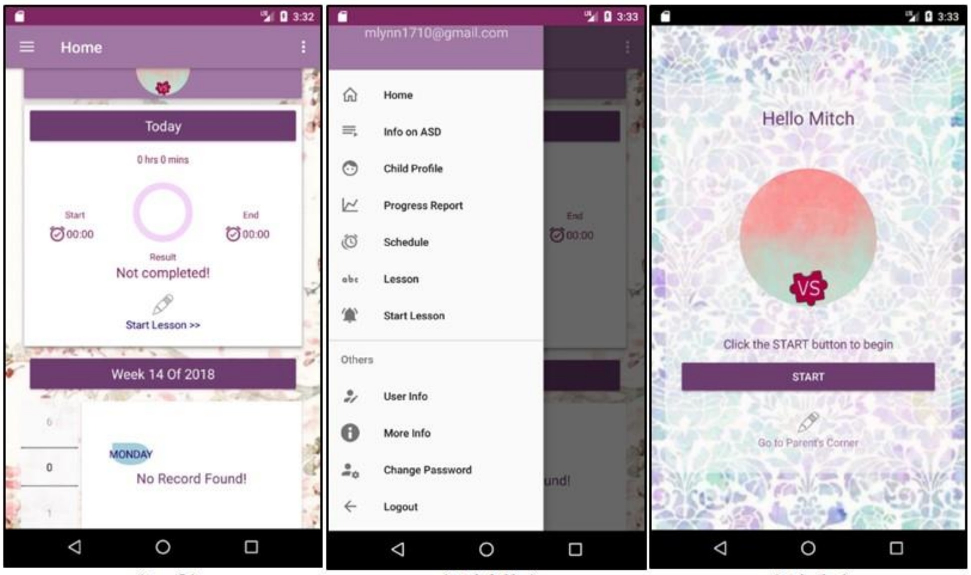
Virtual speech therapy system

The Virtual Speech Therapy System (VSTS) has been developed using AndroidStudio. Users must install the VSTS on their mobile phone or tablet. After installing the VSTS, users must register the user account using their valid email address. The user will receive an email notification to validate their email during registration. This procedure ensures that user activity, lessons, and other details are linked to one unique account. After registration, the user may use their newly registered account to sign in. Once logged in, the user will be prompted to go to the main menu interface (*Figure 5*). The main menu displays the recently completed lesson and the child's schedule. The menu includes Home, Info on ASD, Child Profile, Progress Report, Schedule, Lesson, Start Lesson, User info, More Info, and Change Password functions. The More Info section was added to provide parents with general knowledge regarding ASD and the guidelines for using the VSTS. Parents could start the lesson by clicking the "Start Lesson" button. Once activated, the system will navigate to the lesson page. The lesson starts by clicking the "START" button. Once the activity is completed, the results will be displayed on the main menu, including the outcome, start and end times, and duration. This allows parents to monitor their children's performance and adjust their lessons as needed. The VSTS has undergone acceptance testing to evaluate the functionalities of each component. This is to prove that the system could be made available for operational usage once testing findings satisfy the acceptance criteria (Santos et al., 2018; Leung and Wong, 1997). A mobile phone with an Android operating system is used to access the system for testing. The results of the acceptance testing are tabulated in *Table 3*. From the results, it can be concluded that most components function correctly except the speech-recognition unit.

Table 3. *Results of the acceptance testing.*

Id	Test case	Status
FR1	Able to log into the application	Pass
FR2	Able to register for the application	Pass
FR3	Able to log out from the application	Pass
FR4	Able to change password	Pass
FR5	Able to retrieve password	Pass
FR6	Able to view the Help Section	Pass
FR7	Able to change language (from Malay to English)	Pass
FR7	Able to change language (from English to Malay)	Pass
FR8	Able to modify or edit User Info	Pass
FR9	Able to modify or edit Child Info	Pass
FR10	Able to add resources, including new words, images, and audio	Pass
FR11	Able to edit resources	Pass
FR12	Able to delete resources	Pass
FR13	Able to add lessons	Pass
FR14	Able to edit lessons (resources or time)	Pass
FR15	Able to delete a lesson	Pass
FR16	Display accurate information	Pass
FR17	Able to generate audio output	Pass
FR18	Able to count time	Pass
FR19	Able to measure and record performance	Pass
FR20	Able to recognize input speech	Limited

eISSN: 2716-6481
https://doi.org/10.55197/qjssh.v5i6.471



(Left) (Middle) (Right)

Figure 5. *VSTS interface, (Left) main, (Middle) other functions, and (Right) lesson page.*

The most critical part of the VSTS is the speech-recognition unit because it directly influences the system's effectiveness in providing accurate and meaningful lessons. This unit must recognize input speech with high accuracy to ensure that the learning sessions are productive and that the responses from children with ASD are correctly understood and appropriately responded to. Accurate speech recognition is essential for tracking progress, providing feedback, and maintaining the integrity of the intervention process. Without reliable speech recognition, the system's ability to deliver effective and consistent therapy would be compromised. The performance of the speech-recognition unit is measured using two metrics: the Percentage of Accuracy (% acc) and the Word Error Rate (WER), given by Eq. (1) and Eq. (2), respectively. These metrics compare the predicted speech against a reference transcription (the correct, human-verified transcription) to quantify how well the speech-recognition unit performs. In the equations, N is the total number of words in the reference transcription of the test set. H is the number of correct labels (hits), indicating the words the speech-recognition unit has accurately recognized and transcribed as they are in the reference.

$$\text{Percentage of Accuracy } (\%acc) = \frac{H}{N} \times 100\% \qquad \text{Eq. (1)}$$

$$Word\ Error\ Rate\ (WER) = 100 - \%acc \qquad \text{Eq. (2)}$$

A higher Percentage of Accuracy indicates a more accurate system. Word Error Rate measures the percentage of errors the speech-recognition unit makes relative to the total number of words in the reference transcription. A lower WER signifies a better-performing system, indicating fewer mistakes were made during transcription. High accuracy and low Word Error Rate indicate a speech-recognition unit's effectiveness in producing transcriptions that closely match the human-generated reference, which is crucial for applications ranging from voice-activated assistants to automated transcription services. Seven hundred four (704) speeches comprising sixteen Malay

eISSN: 2716-6481
https://doi.org/10.55197/qjssh.v5i6.471

words were involved in the training and testing. Six hundred and forty data (640) were used in training, while the other sixty-four (64) were used in testing. The training and testing process for the HMM involves a series of systematic steps (Maseri and Mamat, 2020; 2019). Initially, a prototype HMM template file is created to define the model structure, including states and Gaussian components. The global mean and variance are computed from the training data to populate the initial HMM parameters, which are then duplicated to form copies for each monophone HMM. The model is expanded and iteratively re-estimated through several stages, resulting in the final optimized HMMs. Testing involves compiling the task grammar to create a network and recognizing test utterances.

Table 4 and *Table 5* show the confusion matrices for the training and testing, respectively. The Word Error Rate (WER) is 5.63% during training and 6.25% during testing, indicating that the speech-recognition unit demonstrates promising performance. “Sembilan” and “Hidung” are perfectly recognized in training and testing. These words have distinctive pronunciations with the other words. Both “Sembilan” (CVC+CVC+VC, /sem/ /bi/ /lan/) and "Hidung" (CV+CVCC, /hi/ /dung/) are made up of consonant-vowel-consonant (CVC), where the V is a monophthong, and the final C may be an approximant, reflecting the unique property in acoustic modeling that distinguishes them from other words. Furthermore, both words have a longer duration, more vowels, and distinct formant structures. This unique behavior makes the words easily recognizable from other phonemes. The most challenging word to recognize was “Mata,” with accuracies of 87.50% and 75.00% for training and testing, respectively, with a total substitution of five for training and one for testing. The substitution of the “Enam” occurs when pronouncing the word “Mata”. The low accuracy is due to the similar speech characteristics of the words. “Mata” has a phoneme sequence of /ma/ /ta/ and a structure of CV+CV, while “Enam” has a phoneme sequence of /ə/ /nam/ and a structure of V+CVC.

eISSN: 2716-6481
https://doi.org/10.55197/qjssh.v5i6.471

***Table 4**. Confucian matrix for training.*

Speech	0	1	2	3	4	5	6	7	8	9	Telinga	Mata	Tangan	Kaki	Mulut	Hidung	%acc	WER
0	39												1				97.50	2.50
1		36				1					1		2				90.00	10.00
2			39				1										97.50	2.50
3	1			37		1							1				92.50	7.50
4					37		1					2					92.50	7.50
5				1		36		1			1		1				90.00	10.00
6	1	1					37		1								92.50	7.50
7								39		1							97.50	2.50
8									39				1				97.50	2.50
9										40							100.00	0.00
Telinga			1			2				1	36						90.00	10.00
Mata		1					1	1		1		35	1				87.50	12.50
Tangan							1		1			1	36		1		90.00	10.00
Kaki														40			100.00	0.00
Mulut													2		38		95.00	5.00
Hidung																40	100.00	0.00
Average																	94.38	5.63

***Table 5**. Confucian matrix for testing.*

Speech	0	1	2	3	4	5	6	7	8	9	Telinga	Mata	Tangan	Kaki	Mulut	Hidung	%acc	WER
0	4																100.00	0.00
1		4															100.00	0.00
2			4														100.00	0.00
3				4													100.00	0.00
4					4												100.00	0.00
5						3											75.00	25.00
6							4										100.00	0.00
7								4									100.00	0.00
8									4								100.00	0.00
9										4							100.00	0.00
Telinga											3						75.00	25.00
Mata												3					75.00	25.00
Tangan													3				75.00	25.00
Kaki														4			100.00	0.00
Mulut															4		100.00	0.00
Hidung																4	100.00	0.00
Average																	93.80	6.30

eISSN: 2716-6481
https://doi.org/10.55197/qjssh.v5i6.471

The speech-recognition unit within the VSTS shows promising accuracy in identifying spoken words during model development. However, preliminary testing reveals notable limitations in speech-recognition accuracy when the system is deployed in real-world settings. These limitations are likely due to the constraints of the Hidden Markov Model (HMM) used (Wang et al., 2019). The model might suffer from limited training data, which leads to insufficient exposure to the variety and nuances of real-world speech. Also, a noisy environment may overpower the speech signal, leading to reduced clarity and increased error rates in speech recognition. Implementing noise reduction algorithms or improving the signal-to-noise ratio is essential for reliable performance. Employing a new speech recognition model, such as a hybrid HMM-DL (Deep Learning), may improve recognition accuracy (Santos et al., 2024), even in small datasets (Ghaffarzadegan et al., 2017). However, these two studies did not test the hybrid models in real-world settings, which remains a critical next step for validation. The audio sensor quality and silence duration are also important. Different devices come with different audio sensors. The quality and configuration of the audio sensor directly impact the fidelity of the captured speech data. Sensor sensitivity and placement variations can introduce inconsistencies in the audio recordings, potentially affecting recognition accuracy. Excessive silence within the speech samples can skew the recognition process. When silence constitutes a significant portion of the recording, it may be misinterpreted as speech or lead to truncation of spoken words, resulting in inaccuracies.

Conclusion

This study shows that traditional evidence-based intervention techniques such as DTT could be implemented in software applications. Still, adequate research must be conducted to ensure the transformation does not deviate from the conventional procedure and the final product is beneficial to the needed community. Further research should focus on improving the accuracy of the speech-recognition unit. Adaptive filtering methods that dynamically adjust to the audio sensor’s configuration and speaker's environment could provide a more robust speech recognition experience in various settings. Future studies can explore additional acoustic features to improve discrimination between similar words. Training models on a larger speech corpus containing mispronounced words can enhance performance accuracy. The developed VSTS influences advancements in speech therapy with digital DTT-infused lessons. The VSTS offers various tools and features that enable parents to deliver personalized and practical speech therapy sessions to children with ASD. The VSTS has the potential to help children with speech disorders improve their speech skills and communication abilities. However, for the system to be truly effective, it must be used as a supplement to traditional speech therapy and under the supervision of caregivers. Caregiver involvement ensures the therapy is applied correctly, monitors the child's progress, and provides additional support and encouragement. This combination maximizes the benefits of the VSTS while ensuring that the child's needs are adequately met.

Acknowledgement

This work was funded by Universiti Malaysia Sabah, through a Research Grant SBK0293.

Conflict of interest

The authors confirm that no conflict of interest is involved with any parties in this research.

REFERENCES

[1] Adelson, R.P., Ciobanu, M., Garikipati, A., Castell, N.J., Singh, N.P., Barnes, G., Rumph, J.K., Mao, Q., Roane, H.S., Vaish, A., Das, R. (2024): Family-centric applied behavior analysis facilitates improved treatment utilization and outcomes. – Journal of Clinical Medicine 13(8): 17p.

[2] Baron, A., Harwood, V., Woodard, C., Anderson, K., Fernandes, B., Sullivan, J., Irwin, J. (2024): Using the Listening2Faces App with Three Young Adults with Autism: A Feasibility Study. – Advances in Neurodevelopmental Disorders 13p.

[3] Cambridge University Engineering Department (2017): HTK Speech Recognition Toolkit. – Cambridge University 4p.

[4] Cardinal, J.R., Gabrielsen, T.P., Young, E.L., Hansen, B.D., Kellems, R., Hoch, H., Nicksic-Springer, T., Knorr, J. (2017): Discrete trial teaching interventions for students with autism: Web-based video modeling for paraprofessionals. – Journal of Special Education Technology 32(3): 138-148.

[5] Ghaffarzadegan, S., Bořil, H., Hansen, J.H.L. (2017): Deep neural network training for whispered speech recognition using small databases and generative model sampling. – International Journal of Speech Technology 20: 1063-1075.

[6] Girolamo, T., Shen, L., Monroe Gulick, A., Rice, M.L., Eigsti, I.M. (2024): Studies assessing domains pertaining to structural language in autism vary in reporting practices and approaches to assessment: A systematic review. – Autism 28(7): 1602-1621.

[7] Hodges, H., Fealko, C., Soares, N. (2020): Autism spectrum disorder: definition, epidemiology, causes, and clinical evaluation. – Translational Pediatrics 9(Suppl 1): S55-S65.

[8] Leung, H.K.N., Wong, P.W.L. (1997): A study of user acceptance tests. – Software Quality Journal 6(2): 137-149.

[9] Lovaas, O.I. (1987): Behavioral treatment and normal educational and intellectual functioning in young autistic children. – Journal of Consulting and Clinical Psychology 55(1): 3-9.

[10] Maharjan, J., Garikipati, A., Dinenno, F.A., Ciobanu, M., Barnes, G., Browning, E., DeCurzio, J., Mao, Q., Das, R. (2023): Machine learning determination of applied behavioral analysis treatment plan type. – Brain Informatics 10(1): 19p.

[11] Maseri, M., Mamat, M. (2020): Performance analysis of implemented MFCC and HMM-based speech recognition system. – In 2020 IEEE 2nd International Conference on Artificial Intelligence in Engineering and Technology (IICAIET), IEEE 5p.

[12] Maseri, M., Mamat, M. (2019): Malay language speech recognition for preschool children using hidden Markov model (HMM) system training. – In Computational Science and Technology: 5th ICCST 2018, Kota Kinabalu, Malaysia, Springer Singapore 10p.

[13] Panda, P.K., Elwadhi, A., Gupta, D., Palayullakandi, A., Tomar, A., Singh, M., Vyas, A., Kumar, D., Sharawat, I.K. (2024): Effectiveness of IMPUTE ADT-1 mobile application

in children with autism spectrum disorder: An interim analysis of an ongoing randomized controlled trial. – Iranian Journal of Materials Science and Engineering 15(2): 262-269.

[14] Parsons, S., Yuill, N., Brosnan, M., Good, J. (2015): Innovative technologies for autism: critical reflections on digital bubbles. – Journal of Assistive Technologies 9(2): 116-121.

[15] Qin, L., Wang, H., Ning, W., Cui, M., Wang, Q. (2024): New advances in the diagnosis and treatment of autism spectrum disorders. – European Journal of Medical Research 29(1): 11p.

[16] Santos, E.C.D., Vilain, P., Longo, D.H. (2018): A systematic literature review to support the selection of user acceptance testing techniques. – In Proceedings of the 40th International Conference on Software Engineering: Companion Proceeedings 2p.

[17] Santos, L., de Araújo Moreira, N., Sampaio, R., Lima, R., Oliveira, F.C.M.B. (2023): Speech Recognition Using HMM-CNN. – In World Conference on Information Systems and Technologies, Cham: Springer Nature Switzerland 10p.

[18] Smith, T. (2001): Discrete trial training in the treatment of autism. – Focus on Autism and Other Developmental Disabilities 16(2): 86-92.

[19] Wang, D., Wang, X., Lv, S. (2019): An overview of end-to-end automatic speech recognition. – Symmetry 11(8): 26p.

[20] Wergeland, G.J.H., Posserud, M.B., Fjermestad, K., Njardvik, U., Öst, L.G. (2022): Early behavioral interventions for children and adolescents with autism spectrum disorder in routine clinical care: A systematic review and meta-analysis. – Clinical Psychology: Science and Practice 29(4): 400-414.

[21] Yanchik, A., Vietze, P., Lax, L.E. (2024): The effects of discrete trial and natural environment teaching on adaptive behavior in toddlers with autism spectrum disorder. – American Journal on Intellectual and Developmental Disabilities 129(4): 263-278.

[22] Zibin, A., Altakhaineh, A.R.M., Suleiman, D., Al Abdallat, B. (2023): The effect of using an Arabic assistive application on improving the ability of children with autism spectrum disorder to comprehend and answer content questions. – Journal of Psycholinguistic Research 52(6): 2743-2762.

eISSN: 2716-6481
https://doi.org/10.55197/qjssh.v5i6.471