CHILD'S AGE RANGE PREDICTION USING SINHALA SPEECH RECOGNITION SYSTEM

Ashimi Kathriarachchi
Faculty of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
it17171298@my.sliit.lk

Veerandi Kulasekara
Faculty of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
veerandi.k@sliit.lk

Vijani Piyawardana
Faculty of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
vijani.p@sliit.lk

Koliya Pulasinghe
Faculty of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lank
koliya.p@sliit.lk

***Abstract*— This research aims at determining the age range of a child who speaks to the computer by analyzing voice characteristics and acoustics features. To screen speech impairment of children between 6 months to 72 months, it is important to implement a system that can recognize their native language to predict the age range. The implemented system generates a highly accurate report for speech pathologists to use as their second opinion when diagnosing a child. In this research, Multilayer Perceptron based neural network is proposed to identify the age group of a child who speaks Sinhala as their native language. The age recognition system's performance is determined by the speech features that have been used by the child. Mel Frequency Cepstral Coefficients (MFCC) are chosen to capture the difference because it is a useful function for speech recognition. A blend of pitch and Mel Frequency Cepstral Coefficients (MFCC) features were used to increase recognition rates even further. The final method implemented shows a 77% accuracy rate of overall identification of age range for a child.**

[1] ***Keywords-Speech Recognition System (SRS), Mel Frequency Cepstral Coefficients (MFCC), Speech pathologist, Age range, Speech impairment***

I. INTRODUCTION

Speech is the most efficient mode of communication between humans and shares their feelings. Humans learn to speak the language according to what they have been heard and what they have seen in their early childhood. The majority of children in the early childhood stage prefer to speak their native language as it is the most hearing language that the parents tend to use [1]. Essentially, children learn to talk in an environment that is influenced by a variety of social and linguistic characteristics, including language, dialect, and social standing [2]. According to researchers when considering the world population, approximately 6% of children have speech and language difficulties of which the majority of children will not have any other significant developmental difficulties [3]. In Sri Lanka 3-8% of children population have communication and speech-related impairments [4]. To treat a person who is having a speech disorder or an impairment the recoverability rate is much higher if they can be identified at an early age. So, the key factor is the time that taking to identify the disorder [5][6]. Furthermore, children who receive early intervention have improved long-term prognoses [7]. Considering the above information this research is to implement a tool to detect speech impairments of children who are aged between 6 months to 6 years. If a child is suffering from a speech disorder healthcare teams are attending to that because it affects the child's lifetime social interaction and communication. According to the World Health Organization (WHO), the speech impairments rate is increasing [8]. And they pay considerable attention to preventing steps and early diagnosing steps.

When it comes to diagnosing speech impairments of children, it is important to identify the child's age. Because early speech and language skills are associated with success in developing reading, writing, and interpersonal skills [9]. If a speech pathologist identifies a speech or language defect in the early stages of a child and starts treatments, according to medical research the recovery possibility rate is also high [4]. Based on the age, speech pathologists try to examine whether the child has any speech or language impairments. Also, they monitor whether there is speech development that happened after they started the treatments. So, considering all the above information it is obvious that predicting age plays an important role in identifying speech impairments [10]. Predicting a child's age range based only on the vocal tract is difficult and also time-consuming for the medical team. Also, it is difficult for them to make a discission because sometimes the symptoms may not be externally visible. Currently, the speech pathologist manually does an examination routine to identify child speech defects based on their age range [11][12]. Summary World Report which was published in 2011 by WHO mentions that the human resource capacity is degrading or even many countries don't have enough speech pathologists to examine the disorders.

Based on the above information to identifying speech impairment using an Automatic Speech Recognition (ASR) system is the best possible way to help medical teams or even speech pathologists to diagnose the impairments or language differentiation and speech delays of children. ASR is a method of converting input speech signals to digital signals and can be used to analyze features of the voice [7]. As mentioned before, children usually tend to speak their native language in their early childhood stage.

During the past few years, many researchers work on developing SRS for their native language because many of the existing systems are developed to recognize the English language [13][14]. At present, there is a massive improvement in the development of SRS to recognize the native language using the latest technologies and algorithms [15]. Throughout the proposed research project, it is developed to recognize Sinhala language impairments, because Sinhala is an Indo-Aryan language which is the native language of Sri Lanka. Approximately 70% of Sri Lanka's population speak in Sinhala and 15 million Sinhalese use the language as their mother tongue while 19 million people around the world know the language[16][17][18].

Building a system that identifies Sinhala speech impairments would be a great help for Sri Lankan healthcare teams to detect speech impairments between 6 months to 6 years at an early age. The system proposed is based on speaker-driven speech recognition where the system recognizes human inputs or training speaker voice accordingly in real-time and generates a result based on comparing the input data with the trained data as a support to the speech pathologist, the system records and monitors the improvements child shows with the age.

The rest of this paper is organized as follows: Section II illustrates the related work for age range recognition and speech recognition. Section III describes the methodologies used to build the model and the system artificial neural networking to recognize the age range of the child who speaks the Sinhala language. Section IV illustrates the results that the system generates and the accuracy rate of the system. Section V includes the conclusion of the paper and finally, Section VI is about the future work.

II. RELATED WORK

There have been several studies conducted in this domain. In [19], the system implemented for detecting Autism Spectrum Disorder (ASD) using ASR System for Children between ages 6 years to 9 years. They have used Domain Focused Deep Neural Network Transfer Techniques to identify the speech disorders and age range of the speaker. They have collected audio data which includes short vocal tracts, a variety of non-verbal vocalizations, and the wide phonological variance to implement the system. CELF-4 is an individually administered language assessment, published by the Psychological Corporation and used by speech-language pathologists to determine if a child has a language disorder or delay. The best performance accuracy was 26.21% for the above-mentioned system.

Research [20] has implemented a SRS for predicting age groups by recognizing specific Acoustic Models (AMs). To implement the system, they have used Portuguese and European people since that is focused to identify age range by specific acoustic features. Technology-wise they used a Support Vector Machine (SVM) age group classifier to predict the age range of the speaker. This system is implemented to identify young and elderly people using the SRS. The human voice is used to identify human biometrics which can be used to identify gender, age range, and many more. In this research study, one component is to identify the speaker's age group.

Paper [21] illustrate research conduct to recognizing age from voice, the research was tested for younger adults (under age 35) from older aged (over age 65) and the research accuracy rate was 78%. They collected samples using a script that has the same sentence to the speaker to identify the age by analyzing the fundamental frequency range of the speaker.

The research [22] it predicting age using a combination of Long Short-Term Memory (LSTM) and Recurrent Neural Networks (RNN). The above system can deal with short utterances ranging from 3 to 10 seconds. And they were able to estimate the age up to 28% accuracy. They were collected audio data from 20 to 70 years aged people from United State to implement the system.

Research paper [23] illustrate age and gender classification from speech using decision level fusion and ensemble-based techniques. They used SVM classifiers, Naive Bayes classifiers, simple cascading specialists approaches to predict the age and gender of the speaker. The age accuracy rate was 54.9% for this system.

In [24] it uses Artificial Neural Networks to go through the Continuous Speech Recognition approach and Mel Frequency Cepstral Coefficients (MFCC) and Dynamic Time Warping (DTW) as feature matching and feature extraction techniques under isolated word recognition. This Recognizer can identify the Sinhala language as it was trained using the Sinhala data set. The system identifies isolated words of the speaker.

In recognition of Sinhala speech [25] paper able to recognize Sinhala language using Hidden Markov Toolkit for this study they have collected 10M Sinhala corpus. Using those data, they had able to achieve a 75% accuracy rate for identifying the Sinhala language through the system.

Analyzing the above information, we can conclude that predicting a speaker's age range gets important attention in this research area. SRSs that are currently available in the market do not work accurately for recognizing speech used by Sinhala society and even more, the existing systems do not focus on recognizing speeches made by children. Available systems are mainly focused on recognizing the speech of the young and middle-aged community. Many of the developed

SRSs are built to recognize the English Language [20][21][22][23]. So, there is a lack of research in the field of recognizing Sinhala speech. Many Sinhala SRSs [24][25] are built to identify adult speech not many of them are focused on recognizing children's speech. Very few [19] systems implement to recognize the child's utterance and it is a huge gap in this area that need to be fulfilled.

TABLE I. COMPARISON OF SPECIFIC FEATURES ON FORMER RESEARCH.

	Age range recognition.	**Sinhala recognition.**	**Children speech recognition.**
Research A [19]	×	×	√
Research B [20]	√	×	×
Research C [21]	√	×	×
Research D [22]	√	×	×
Research E [23]	√	×	×
Research F [24]	×	√	×
Research G [25]	×	√	×

The proposed system in this research has better approaches that can fulfil the above-mentioned research gap.

There are several toolkits and systems that are available in the present market for building ASR systems such as Hidden Markov Toolkit (HTK) [26], Kaldi [27], CMUSphinx [28], and Espresso [29]. All of the above mention toolkits and systems are open source. HTK is a toolkit that supports Linux platforms and is based on recurrent neural network language [26]. Kaldi toolkit original version is based on C++, but there is a Kaldi version with a python wrapper named as PyKaldi which many researchers used to implement SRSs. Kaldi is a Linux based system such as HTK, CMUSphinux, is a Java-based cross-platform system that supports many languages like C, C++, Python [28]. But this research has been used the Librosa framework to build the Sinhala SRS to predict a child's age.

The methodology that has been used to implementing the system shows high accuracy for identifying the age range of a child. The methodology section illustrates the methods that have been used in the system more clearly.

III. METHODOLOGY

Speech is a continuous audio stream where rather stable states mix with dynamically changed states [28]. In this sequence of states, one can define more or less similar classes of sounds, or phones. Words are built using a collection of phonemes, but this is always not true. The acoustic properties of a waveform corresponding to a phoneme can vary greatly depending on many other factors - phone context, speaker, style and so on. The popular coarticulation makes phones sound drastically different from their "canonical" representation [30].

To develop a system to recognize speech an acoustic model is used. It is used to establish a relationship between acoustic information and language construct in speech recognition [31]. So, the present developed SRS has typically been trained with audio recording collected from children of age 6 months to 6 years. This research uses pitch and frequency range to predict the child's age range because this particular method is most suited to identifying a child's voice. It has drastically changed frequency ranges comparing to different age ranges. When diagnosing speech impairments, the proposed system compares a typical child's frequency range and pitch range with the atypical child's features. Based on this result, the system produces a report which can be used by a speech pathologist to identify whether the child has an abnormal behaviour or a speech delay comparing to a healthy child

Neural Networking technology has been used to develop the proposed system. Librosa 0.8.0 package is used to extract features from audio and analyze them. Librosa is a python package used for audio analysis used in this research. It provides many features that can be easily used to build a SRS and feature extraction. Like previously mentioned it has been used the library to build the Sinhala ASR system.

Sckit- learn library was used which is a well-known library built for Machine Learning using python [30]. Sckit- learn library was used to build the Sinhala SRS as it used to predict the unknown future values using operations. This is also a well-known library built for Machine Learning using python [31]. It is used for logistic regression because it is extremely straightforward to implement logistic regression models.

When implementing the proposed system, we have been used several other modules. Numpy was used to manipulating arrays when we trained the data for the system and it is a python package used for calculating statistical data. In the proposed system also, we have been used to statistically analyse each age range data when training the system. Numpy arrays were used to store the results after extracting features of the audio data which was used to predict the age range of the child and do a linear algebraic operation of the system. Numpy version 1.19.3 was used to build this system. It helps to get the standard deviation, variances, and percentiles of the data set after store

Multi-layer Perceptron Classifier is used as the classifiers in this proposed system. This is a supervised machine learning algorithm used for neural network models. Using this classifier, able to differentiate from logistic regress between the input layer and output layer. Also, used hidden layers to break down the voice data and improve the accuracy range [32].

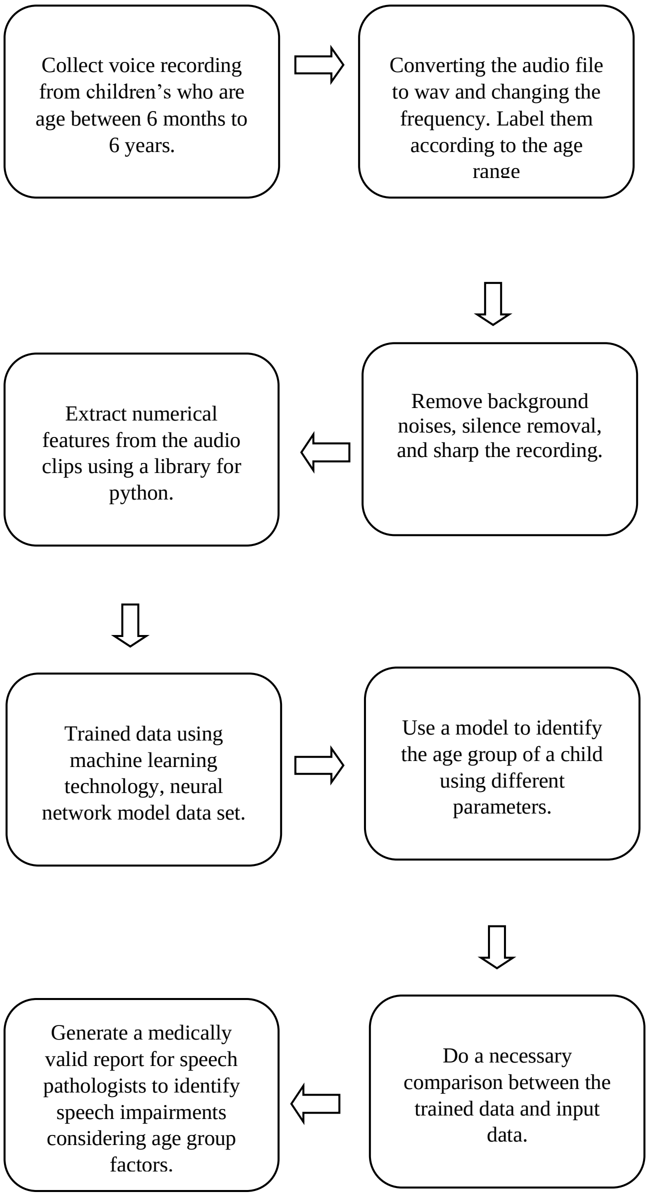

Fig. 1. The flow of the age range recognition system.

Fig.1 illustrates the system diagram that is proposed in the research paper. First loaded the audio dataset which was collected from children between 6 months to 6 years, extracted audio features from the data collection which requires to predict age range, split into training and testing sets. Then initialized the ML model, MPL classifier and train the data. At last, calculate the accuracy rate for predicting the age range of the child. The following mentioned steps are the specific phases and the step which took to build the Sinhala speech recognition's age prediction component. The steps are clearly illustrating the methods behind each phase or step.

A. Data Collection

For this research as raw data, audio data from children aged between 6 months to 6 years were collected and create the data collection. To develop a system to identify the age of children, the input audio utterances are compared with the trained data. When implementing a system that needs to interact with children it is difficult to get more accuracy because children are using a limited number of words and their vocabulary also has a certain limitation. so cannot force them to say what we wanted to collect. Hence, a library with children's short utterances was created, where it can be used to identify the age and other features such as emotions, pronunciation and intonation in speech. When it comes to building ASR systems for children, further in this study better to have a certain script and collect data according to that and build a high accuracy system in which children use many emotions, intonation and phonetics.

To collect data to the trained model, collected many voice samples from various age ranges and in various situations like when the child is in anger, happy and calm. To collect data from children we got help from parents and used smartphone devices which were able to collect many voice samples. Collecting data was a time-consuming task because children are usually don't speak with strangers and also, they are speaking the same word using different pronunciations and different accents.

After collecting the data, it is converted from mp3 recording into 16 kHz, mono streaming. wav files. The six age categories along with specific acoustic features that are used to identify the age ranges are listed below.

- 6 months to 1 year age range: 5 female voice samples and 3 male voice samples,
- 1 to 2 years age range:14 female voice samples and 11 male voice samples,
- 2 to 3 years age range: 20 sample recordings; 14 female voice samples and 6 male voice samples,
- 3 to 4 years age range: 20 sample recordings; 16 female voice samples and 4 male voice samples,
- 4 to 5 years age range: 26 female voice samples and 14 male voice samples,
- 5 to 6 years age range: 18 female voice samples and 12 male voice samples,

were collected to build the data set for identifying the age range of a speaker. Mainly collected audios from different age ranges, in different emotional areas, different intonations and different kinds of phonemes to build a vast Sinhala data set. This was taken to identify the spectro-temporal features to recognize the particular age range of a child using the Sinhala speech.

B. Feature Extraction

During the last two decades, short-time spectral representations have been one of the most common speech feature types in automatic speaker recognition research[2].

Mel Frequency Cepstral Coefficients (MFCC), Pitch (Using Chroma) was used as features selection as those are efficient methods. Feature extraction is a pattern recognition technique that involves extracting meaningful low-dimensional representations from higher-dimensional data to identify the age range of a child. The MFCC method is one of the most widely used feature extraction methods in ASR approaches. It identifies the audio and discards other stuff like noises, which is a major task for SRS. Pitch was extracted using Chroma since it is used to extract meaningfully characterized pitches of the voice which are used to predict the age range of the child.

C. *Training Data and Classifier*

The implemented system is based on Neural Networking Multilayer Perceptrons for training and testing the audio data collected from children. Other several models are used to build SRS such as Connectionist Temporal Classification (CTC), Recurrent Neural Network (RNN), Convolutional Neural Networks (CNN) and attention-based models [32]. This research, used Multilayer Perceptrons (MLPs) to frame classification with shuffled frame data. It uses a mini-batch size of 256 frames and 50 Hidden Layers. The frames are organized by sentences when loading data into training, and all frames in a sentence are passed in their natural order. The age range prediction model is trained, and perform backpropagation through the time algorithm and the cost associated with each frame is accumulated over time. For both networks, the model is trained until it begins to overfit on the training set and the dev accuracy begins to fall. These are the technologies that have been used to implement the training data for the Sinhala SRS.

This study concludes to predict a child's age. It was requested to detect the age of the speakers in six separate classes 6 months to 1-year, 1 - 2 years, 2 - 3 years, 3 - 4 years, 4 – 5 years, and 5 – 6 years. The training and collected data from the corpus are labelled according to these age groups.

To evaluate the performance of the classifier the repeated s-fold, cross-validation method is used. According to this method if s=20, the utterances in the data collection are divided into a training set containing 75% of available data and a disjoint test set containing the remaining 25% of the data. The output of the speech age classifier is predicting the value (label) of the actual speaker's age. Scikit- learn library is used for cross-validation. The procedure is repeated 20 times. The training and the test sets are selected randomly for the cross-validation. Then the system using MLP classifiers able to predict the age range of a child based on voice fundamental frequency, pitch, and considering the vocal tract. As a classifier, it used MLP Classifier because that gave the highest accuracy rate comparing to other classifiers and technologies.

Then able to build a system to generate the predicted output value in a human-readable format. According to the results.

IV. RESULTS AND DISCUSSION

From the system implemented, it can detect a child's age range using machine learning. The system implemented has an age prediction accuracy rate of up to 77%. And also, able to get the perdition accuracy for each age range separately.

TABLE II. ACCURACY COMPARISON FOR AGE GROUPS.

Age range	Training accuracy percentage (%)	Testing accuracy percentage (%)
6 months – 1 year	62	48
1 -2 years	75	58
2-3 years	88	70
3-4 years	86	65
4-5 years	90	78
5-6 years	83	63

TABLE II. illustrate the results obtained from the evaluation phase showing the effectiveness of the model compared to the baselines and the state of the art on the Sinhala custom dataset in training the model and when testing the model. In particular, TABLE II. shows each age range with particular accuracy. TABLE II., it shows that the 4 - 5 years age range accuracy is high because the number of trained data is higher when it compares to other age categories. For that particular age range, there were 40 sample data and from that, the model was able to train 36 samples correctly. The same logic applies to getting the lowest accuracy for 6 months - 1 year age range because one reason is that it has a smaller number of raw data to train the model which was only 8 samples. Using those 8 samples 6 samples were identified correctly in the training phase of the system. According to [33] there is no huge difference in the rhythms of the voice used by 6 months to 1 year old child with adjacent age range groups.

The above table information clearly illustrates that there is a difference between training and testing. Testing results are always lesser than training data. Because testing data background noise and the audio quality is lesser than the sample data.

When considering the age range, however, there is a slight difference between adjacent age ranges. Because even it cannot see a drastic vocabulary change in adjacent ages.

Using the proposed system, a speech pathologist can predict the user's age range and then comparing with the actual age range to conclude whether the child is having any speech impairments.

V. CONCLUSION

This research study attempted to build a Sinhala ASR system for children aged 6 months to 6 years to predict age range, which helps speech pathologists to diagnose speech impairments of children. The test results show that the system achieves 77% age range recognition accuracy.

For some age ranges considering individually gives a high prediction accuracy rate because we were able to find more data in particular age ranges. When implementing the system, it used MLP as the class of Neural Network. But as mentioned before tried using CNN and RNN for This particular research but didn't gain competitive results comparing to MPL classifier. RNNs and LSTMs have been tested, but the results have been disappointing, to say the least. In those methods based on autoregression, even linear approaches, frequently outperform linear methods. Simple MLPs often outperform LSTMs when applied to the same data [34]. But in this search, it chose MLP Classifier and features a blend of pitch and MFCC to increase recognition rates even further.

Also, this system is a web-based system it can be used on any platform such and desktop, mobile, or in any operating system. Since the system is a web, it also can be accessed at any time in any place. The usefulness of the system is much higher because of the techniques that have been used to build the system.

ACKNOWLEDGEMENT

This research was supported by the Accelerating Higher Education Expansion and Development (AHEAD) Operation of the Ministry of Higher Education of Sri Lanka funded by the World Bank [35].

REFERENCES

[1] A. Clare, "Bilingualism in the early years," *Commun. Interact. Early Years*, vol. 7, no. 1, pp. 87–100, 2017, doi: 10.4135/9781473919631.n6.

[2] H. Melin, *Automatic speaker verification on site and by telephone : methods , applications and assessment*. 2006.

[3] J. Law, J. A. Dennis, and J. J. V. Charlton, "Speech and language therapy interventions for children with primary speech and/or language disorders," *Cochrane Database Syst. Rev.*, vol. 2017, no. 1, Jan. 2017, doi: 10.1002/14651858.CD012490.

[4] W. D. Y. N. Walpita and S. Ginige, "Timeliness of care received by children with speech and language disorders attending a speech therapy clinic at a tertiary care hospital," *Sri Lanka Journalof Child Heal.*, vol. 43, no. 3, pp. 147–153, 2014, doi: 10.4038/sljch.v43i3.7374.

[5] P. A. Prelock, T. Hutchins, and F. P. Glascoe, "Speech-Language Impairment: How to Identify the Most Common and Least Diagnosed Disability of Childhood," *Medscape J. Med.*, vol. 10, no. 6, p. 136, 2008, Accessed: May 28, 2021. [Online]. Available: /pmc/articles/PMC2491683/.

[6] *Speech and Language Disorders in Children: Implications for the Social Security Administration's Supplemental Security Income Program*. National Academies Press, 2016.

[7] D. L. Robins, D. Fein, M. L. Barton, and J. A. Green, "The Modified Checklist for Autism in Toddlers: An Initial Study Investigating the Early Detection of Autism and Pervasive Developmental Disorders," *J. Autism Dev. Disord.*, vol. 31, no. 2, pp. 131–144, 2001, doi: 10.1023/A:1010738829569.

[8] World Health Organization, "Summary World Report On Disability," *World Health*, pp. 1–24, 2011, [Online]. Available: www.who.int/about/licensing/copyright_form/en/index.html%0Ahttp://www.larchetoronto.org/wordpress/wp-content/uploads/2012/01/launch-of-World-Report-on-Disability-Jan-27-121.pdf.

[9] Dan Brennan, "Baby Talk: Communication With Your Baby," *July 2019* , Jul. 2019. .

[10] S. M. Mirhassani, A. Zourmand, and H. N. Ting, "Age estimation based on children's voice: A fuzzy-based decision fusion strategy," *Sci. World J.*, vol. 2014, 2014, doi: 10.1155/2014/534064.

[11] "Speech Disorders: Causes, Signs, and Diagnosis." https://www.healthline.com/health/speech-disorders#diagnosis (accessed May 31, 2021).

[12] L. Fabiano-Smith, "Standardized Tests and the Diagnosis of Speech Sound Disorders," *Perspect. ASHA Spec. Interes. Groups*, vol. 4, no. 1, pp. 58–66, Feb. 2019, doi: 10.1044/2018_PERS-SIG1-2018-0018.

[13] J.-P. Hosom, "Speech Recognition," 2003.

[14] National Center for Technology Innovation, "Speech Recognition for Learning," Aug. 2010. .

[15] Naomi Van der Velde, "A Complete Guide to Speech Recognition Technology," Jan. 2021. .

[16] T. Nadungodage and R. Weerasinghe, "Continuous Sinhala Speech Recognizer," 2011.

[17] S. Gallege, "Analysis of Sinhala Using Natural Language Processing Techniques." [Online]. Available: www.defence.lk/.

[18] "Official languages in Sri Lanka | Mai Globe Travels." https://www.maiglobetravels.com/language-in-sri-lanka# . (accessed May 23, 2021).

[19] R. Gale, L. Chen, J. Dolata, J. Van Santen, and M. Asgari, "Improving ASR Systems for Children with Autism and Language Impairment Using Domain-Focused DNN Transfer Techniques."

[20] A. Hämäläinen, H. Meinedo, M. Tjalve, T. Pellegrini, I. Trancoso, and M. S. Dias, "Improving Speech Recognition through Automatic Selection of Age Group-Specific Acoustic Models," 2014.

[21] P. H. Ptacek and E. K. Sander, "Age recognition from voice.," *J. Speech Hear. Res.*, vol. 9, no. 2, pp. 273–277, 1966, doi: 10.1044/jshr.0902.273.

[22] R. Zazo, P. Sankar Nidadavolu, N. Chen, J. Gonzalez-Rodriguez, and N. Dehak, "Age Estimation in Short Speech Utterances Based on LSTM Recurrent Neural Networks," *IEEE Access*, vol. 6, pp. 22524–22530, Mar. 2018, doi: 10.1109/ACCESS.2018.2816163.

[23] F. Lingenfelser, J. Wagner, T. Vogt, J. Kim, and E. André, "Age and Gender Classification from Speech using Decision Level Fusion and Ensemble Based Techniques." [Online]. Available: http://www.csie.ntu.edu.tw/.

[24] N. A. C. Sandasarani, "Sinhala Speech Recognition." [Online]. Available: www.ijert.org.

[25] T. Nadungodage and R. Weerasinghe, "Continuous Sinhala Speech Recognizer," *Conf. Hum. Lang. …*, no. May, pp. 2–5, 2011, [Online]. Available: http://hltd.org/pdf/HLTD201123.pdf.

[26] "HTK Speech Recognition Toolkit." https://htk.eng.cam.ac.uk/ (accessed May 29, 2021).

[27] "GitHub - kaldi-asr/kaldi: kaldi-asr/kaldi is the official location of the Kaldi project." https://github.com/kaldi-asr/kaldi (accessed May 24, 2021).

[28] "Basic concepts of speech recognition – CMUSphinx Open Source Speech Recognition." https://cmusphinx.github.io/wiki/tutorialconcepts/ (accessed May 23, 2021).

[29] Y. Wang *et al.*, "Espresso: A Fast End-To-End Neural Speech Recognition Toolkit," in *2019 IEEE Automatic Speech Recognition and Understanding Workshop, ASRU 2019 - Proceedings*, Dec. 2019, pp. 136–143, doi: 10.1109/ASRU46091.2019.9003968.

[30] G. S. M. B, "IDENTIFICATION OF AGE USING VOICE RECOGNITION," *Int. J. Adv. Electron. Comput. Sci.*, vol. 4, no.

ISSN:2393-2835, 2017.

[31] S. Bhatt, A. Jain, and A. Dev, “Acoustic Modeling in Speech Recognition: A Systematic Review,” 2020. [Online]. Available: www.ijacsa.thesai.org.

[32] “How Deep Neural Networks can improve Speech Recognition and generation | Packt Hub.” https://hub.packtpub.com/how-deep-neural-networks-can-improve-speech-recognition-and-generation/ (accessed May 28, 2021).

[33] F. Dick, S. Krishnan, R. Leech, and S. Curtin, “Language Development,” *Neurobiol. Lang.*, pp. 373–388, 2015, doi: 10.1016/B978-0-12-407794-2.00031-6.

[34] “When to Use MLP, CNN, and RNN Neural Networks.” https://machinelearningmastery.com/when-to-use-mlp-cnn-and-rnn-neural-networks/ (accessed May 29, 2021).

[35] “Result Area 3 -.” https://ahead.lk/result-area-3/ (accessed Jun. 05, 2021).