

Personalized Model of an Interactive System for Improving the Pronunciation Skills in Autistic Children

Rohit Rathore, Priyamvada Singh, and R. H. Goudar, *Senior Member, IACSIT*

Abstract—The three major problems identified in the children suffering from Autism Spectrum Disorders are learning difficulties, communication problems and social interaction problems. The social interaction problems and communication problems further hinder the learning abilities of autistic children. In addition to the special methodologies and techniques that should be employed in designing the effective learning solutions for autistic children, a high degree of personalization is also required, as the problems and learning requirements vary greatly from child to child. This paper focuses on the pronunciation problem which is different for each autistic individual, hence a personalized model of an interactive system is proposed for improving the pronunciation of words in children suffering from Autism Spectrum Disorders, that takes into account the custom needs of the children.

Index Terms—Autism, autism spectrum disorders, virtual avatar, speech-to-text conversion.

I. INTRODUCTION

Autism is a developmental disorder that appears in the first 3 years of life, and affects the brain's normal development of social and communication skills. Autism Spectrum Disorders (ASDs) are a group of neurodevelopment disorders characterized by core deficits in three domains: social interaction, communication, and repetitive or stereotypic behavior [1].

Ability to form a word correctly while speaking (pronunciation problem) is a common problem found in autistic children between the age group of 3 to 10 years [2]. To train them using the traditional methods and techniques (i.e. being instructed by a teacher in a class room setting) has been found to be restrictive in more than one ways. Teaching the children how to pronounce a word correctly through the traditional method (via interaction with an instructor) becomes more difficult in case of autistic children because they are more likely to have speech difficulties than a normal child of the same age. The main reason behind this is that the children suffering from Autism Spectrum Disorders (ASDs) find it difficult to interact with the non family members including the teachers (Social Interaction Problems)[3],[4]; so taking help of a professional like a speech therapist turns out to be less effective than in case of normal children. A number of research literatures thus advocate the use of

computer-game based methods for teaching as they have been found to be more useful in imparting basic education and knowledge to the autistic children[5], [6].

In further sections of this paper we have first given a brief summary of the work already done in the areas that form the basis of our model. We introduce our model in the next section with the block diagrams and an example. The two modes of operation have been defined and explained. We have then tried to identify the future scope of the proposed system-model followed by the conclusion and references of the material we have used.

II. RELATED WORK

The effect of interactive video and attractive pictures or interface on children's learning has been analyzed and proven [7]. A large number of softwares are available on CD/DVDs which offer an animated environment for students to learn a new topic. We have taken a similar approach by presenting the user with the input in form of images and the written text simultaneously so that the child can use more than one preceptor sense at the same time to identify a word.

Another approach seeks to encourage children with ASD to "play," where playing is mediated through technology. By creating technological methods of interaction (visual displays), play and comfortable interactions can be garnered from children with autism. There is a feeling of "safety" by having the main form of interaction occur with non-humans. Further, these devices allow the child, rather than a third party, to be in control of the interactions [1].

Recent researches have shown that the use of virtual agents have much to offer in teaching the essential social skills effectively, as their ability to model realistic facial expressions and dialogue gives individuals with social skills deficits a non judgmental environment to practice their skills in prior to facing complex and often confusing real-world situations.[8], [9]. This motivated us to introduce the use of virtual agents which we are referring to as "avatars" as a means to provide the interaction between the proposed system and the user.

Another paper describes an interactive computer game for the autistic children to improve the fluency in their speech. This game produced encouraging results with a number of participants during three months of observation [10].

III. PROPOSED MODEL

We have proposed a model which is an interactive system that aids in improving pronunciation difficulties in the children suffering from ASDs through training and feedback

Manuscript received April 14, 2012; revised August 8, 2012.

Rohit Rathore and Priyamvada Singh are with Information Technology from Graphic Era University, Bell Road, Dehradun (248005), India (e-mail: rohitrathor@gmail.com, priyakip@gmail.com).

Dr. Rayan Goudar is with Graphic Era University Dehradun (248005), India as Associate Professor, Dept. of Computer Science and Engineering (email: rhgoudar@gmail.com)

(self learning), using the tools for speech processing and conversion. The interaction with the kids in the proposed system also involves the use of “avatars” which will be created by using the picture of the child. This, not only introduces an element of fun in the learning process, but also helps in enhancing the degree of learning in the children suffering from ASDs [8], [11].

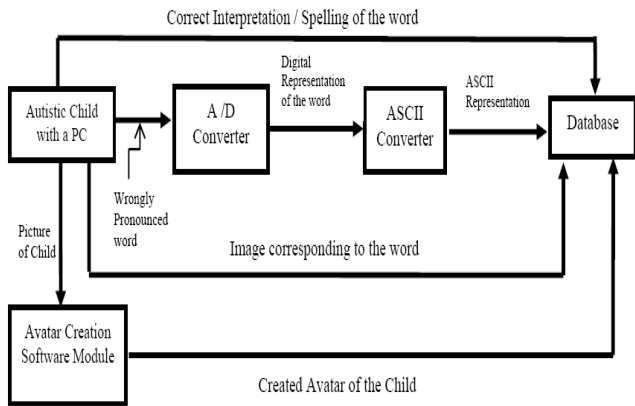



Fig. 1. Block diagram for setup mode

Every autistic child has his very own set of learning requirements, so any software application targeting to teach them should involve a high degree of customization and personalization. Proposed system identifies the words that are wrongly pronounced by a particular kid and uniquely creates an avatar for him. In this way, a separate, individual database section is created for each different user of the system which caters to his very specific personal learning requirements. The proposed application has two modes- a *setup mode* (for creating the database) and an *operational mode* (for training).

A. Setup Mode

1) The parents of the autistic children are required to work in close coordination with their teacher to identify the words that are wrongly pronounced by their child. These words and their corresponding correct interpretations are then fed into the system during the setup mode. The kid will speak the word through the microphone and the correct interpretation of that word could be entered by the parent or teacher manually as text. The word spoken by the child goes through a number of speech processing steps before being stored in the database (Refer to Fig. 1).

TABLE I: IF A CHILD PRONOUNCES THE WORD “MORNING” AS “MOANING”, THE DATABASE ENTRY WILL LOOK LIKE

Wrongly Pronounced Word (ASCII representation)	Correct Interpretation	Image(depicting the word or the concept)
Moaning	Morning	

The child interacts with the system with the help of a module called *supervising system* (which could be

These steps are:

- The word spoken through the microphone is converted from analog to digital form through the sound card/ sound device in the personal computer (PC) using a modulation technique such as PCM.
- The digital form of the word is then converted to the corresponding ASCII representation. This ASCII representation of the word is stored in the database along with the correct interpretation (that is to be fed directly as text by the teacher or the parent)
- An image corresponding to the word is also stored in the database. The image should either depict the word in the visual form directly or should depict the concept being meant by the word. E.g. image of an apple can be used for the word “apple”. And the image of the rising sun can be used to depict the concept in the word “morning”. Table I shows an example entry for the word “morning”.

- 2) To create an “avatar”, a picture of child can be taken using a web camera or an already existing digital image can be used. There is a number of open source software available for “animated avatar creation” through the supplied image files. One of them can be integrated as a module in implementing an application using the proposed model.
- 3) Also, a speech-to-text conversion software module is integrated in the application to handle the words that are not present in the database. Such software takes the word spoken by the child (during the operational mode) as an input and converts it to the nearest text representation found in the dictionary of the language being used. A number of open source tools which does speech to text conversion are available.

In this way a database of the words can be created and maintained by the system; and parents and/or the teacher can keep adding to it as and when they identify a new word.

B. Operational Mode

The proposed system is an interactive one where all the interaction with the system is provided through the use of “virtual avatars” that have been personalized as per the user (the child). Thus, a suitable user interface that integrates the concept of personalized virtual avatars within can be designed and presented to the user.

administrator controlled, controlled by the teacher/instructor or could be automated).The roles of the *supervising system* include:

- 1) Presenting a new word on the screen along with the picture depicting the word or the concept related to the word: The word will be chosen from among the words stored in the database (created in the setup mode).
- 2) To decide whether the word needs to be re-spoken by the child or not: In case the child pronounce the displayed word/concept correctly, he/she can be given a new word otherwise he will be instructed to re-pronounce the word (After the word has been correctly pronounced by the avatar)
- 3) Keep the statistics regarding the performance of the child

The child is presented with a word along with the corresponding picture depicting the word. He is asked to identify the word through spelling or/and picture and should be instructed (by the avatar) to pronounce the word. The microphone will catch the speech (Fig. 2). Through the speech processing steps, the caught spoken word will be converted to the digital form and then to the corresponding ASCII representation.

There may be three cases depending on the ASCII value:

- 1) If the ASCII value of the word = interpretation of the word (actual spelling): The child has spoken the word correctly. A new word is to be presented to him by the system
- 2) If the ASCII value of the word matches the ASCII value of the wrongly pronounced word in the table entry: The child has pronounced the word wrongly. The correct pronunciation of the word shall be presented through the avatar (i.e. the avatar will speak back the word) and child is instructed to re-speak the word imitating the correct pronunciation he/she just heard.

- 3) If the ASCII value of the caught word is different from both the correct spelling (correct interpretation) and the stored wrong pronunciation present in the table: Then the new “wrong pronunciation” of the word has been identified and it is to be stored in the database (Table II). In this case also, the system gives the right pronunciation back through the avatar and instruct the child to try speaking the word again in the correct manner.

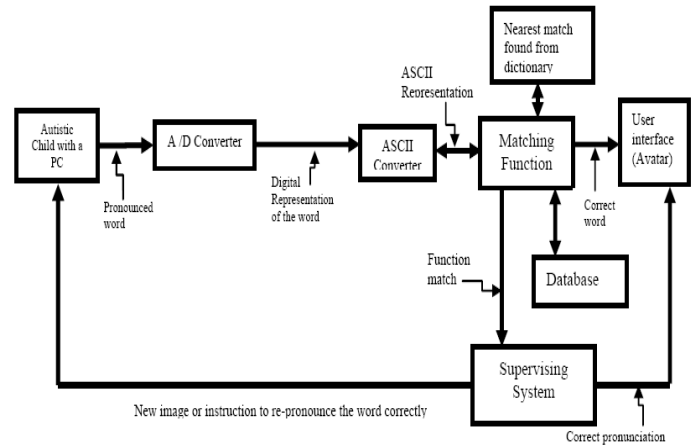



Fig. 2. Block diagram for operational mode

TABLE II: IF CHILD NOW PRONOUNCE “MORNING” AS “MOURNING”, WHILE THE PREVIOUSLY IDENTIFIED WRONG PRONUNCIATION FOR THE SAME WORD THAT WAS STORED IN THE TABLE WAS “MOANING”, THE NEW ENTRY LOOKS LIKE

Wrongly Pronounced Word (ASCII representation)	Correct Interpretation	Image (depicting the word or the concept)
Moaning	Morning	
Mourning	Morning	

Various types of statistical analysis can be provided at the supervising system node (parameters like degree of correctness of the pronunciation etc). Each user (child) has a different database and avatar associated with him/her which is tailor-made according to his/her requirements.

The feature 3) above helps to identify the different wrong pronunciations of the same word that a child uses.

IV. FUTURE SCOPE

We have presented a *conceptual model* of an interactive system for improving the pronunciation in autistic children. We believe that the proposed model could be implemented without much difficulty as many of the modules required for designing the full application are available as open source software. A few modules are needed to be developed and can be integrated with the existing open-source modules. After

providing the necessary customization the system as a whole can be implemented. Once the system is implemented, the actual results can be gathered and analyzed. A metrics for measuring the performance improvement in the pronunciation skills of the children can be defined and the system can be tested against the real subjects (i.e. autistic kids)

V. CONCLUSION

In this paper, we have identified “communication skill problems in autistic children” as an area which requires personalized solutions to make the learning effective. Through this paper we have targeted a specific area i.e. “pronunciation problem in children suffering from ASDs” and have proposed a conceptual model for improving it through the use of an interactive system that integrates the concepts like “computer based teaching”, “virtual avatars”, “customization according to the user needs” to provide an

effective solution for the same.

ACKNOWLEDGMENT

The authors would like to thank Dr. Ankush Mittal for introducing them to this area of research and for his support and motivation throughout the work.

REFERENCES

- [1] Md. M. Rahman, S. M. Ferdous, and S. I. Ahmed, "Increasing Intelligibility in the Speech of the Autistic Children by an Interactive Computer Game," *2010 IEEE International Symposium on Multimedia*, 2011.
- [2] *Center for Disease Control and Prevention*, CDC. Autism Information Center, DD, NCBDDD, CDC, Atlanta, April 25, 2007
- [3] C. B. Baskett, *The Effect of Live Interactive Video on the Communicative Behavior in Children with Autism*, University of North Carolina at Chapel Hill, Chapel Hill, 1996.
- [4] L. Kanner, *Autistic Disturbances of Affective Contact*, Kanner, L. ed. *Nervous Child* 2, V.H. Winston, pp.217-250, 1943.
- [5] M. E. Hoque, "Analysis of Speech Properties of Neurotypicals and Individuals Diagnosed with Autism and Down Syndrome," in *Proc. of 10th ACM conference on Computers and Accessibility (ASSETS)*, Halifax: Nova Scotia, October, 2008.
- [6] M. E. Hoque, J. K. Lane, R. elKaliouby, M. Goodwin, and R. W. Picard, "Exploring Speech Therapy Games with Children on the Autism Spectrum," in *Proc. of InterSpeech, Brighton, UK*, September, 2009.
- [7] C. B. Baskett, *The effect of live interactive video on the communicative behavior in children with autism*, Chapel Hill: University of North Carolina at Chapel Hill, 1996.
- [8] M. Milne and M. H. Luerssen, Trent W. Lewis, Richard E. Leibbrandt, and David M. W. Powers, "Development of a Virtual Agent Based Social Tutor for Children with Autism Spectrum Disorders," *IEEE Standard*, 2010.
- [9] M. Milne and M. H. Luerssen, Trent W. Lewis, Richard E. Leibbrandt, and David M. W. Power, "Development of a Virtual Agent Based Social Tutor for Children with Autism Spectrum Disorders," *IEEE Standard*, 2010.

- [10] A. Anwar, Md. M. Rahman, S. M. Ferdous, S. A. Anik, and S. I. Ahmed, "A Computer Game based Approach for Increasing Fluency in the Speech of the Autistic Children," in *Proc. of 2011 11th IEEE International Conference on Advanced Learning Technologies*, 2011.
- [11] C. Putnam, and L. Chong, "Software and technologies designed for people with autism: what do users want?" in *Proc. of 10th International ACM SIGACCESS conference on Computers and accessibility*, Halifax, Nova Scotia, Canada, 2008.



Rohit Rathore is currently pursuing M.Tech. in Computer Science from Graphic Era University, Dehradun, India. He earlier completed his Bachelors in Computer Engineering from DIT, Dehradun, India. He has industry experience in Microsoft Technologies and Software Testing fields. His research areas include Semantic Web, Ontology based Information Retrieval Systems, E-Learning among others.



Priyamvada Singh is currently pursuing M.Tech from Graphic Era University, Dehradun, India. She has completed her B.E (Computer Science) from Bharati Vidyapeeth College of Engineering, Pune, India. Her areas of interest include Semantic Web, E-learning.



Dr. R. H. Goudar is currently working as an Associate Professor, Dept. of CSE, Graphic Era University, Dehradun, India. He also worked as Faculty at International Institute of Information Technology, Pune for 4 years and Indian National Satellite Master Control Facility, Hassan, India. His subjects of interest include Semantic Web, Network Security and Wireless