WEAVING THE DREAM: ASSISTIVE TEXT-TO-SPEECH TECHNOLOGY IN REALIZING TRUE INCLUSIVE EDUCATION

[1]Rahmania Ananda, [2]Hesty Widiastuty
Institut Agama Islam Negeri Palangka Raya

Keywords:
Text-To-Speech (TTS), Autism Spectrum Condition (ASC), XpressiveTalk.

***Correspondence Address:**
Email: rhmnaananda325@gmail.com

Abstract: Technology can have an auxiliary role in education, especially text-to-speech (TTS) systems, playing an important role in providing inclusive education for students with special needs. This shows that, as defined by UNESCO (2009), it is important for every student to have equal learning opportunities. This research emphasizes how TTS technology can help students with reading difficulties or visual impairments become closer and more engaged. The research shows that this technology can improve emotional understanding and social interaction among individuals with Autism Spectrum Conditions (ASC) by incorporating tools such as XpressiveTalk, which combines expressive visual avatars with emotional recognition capabilities. The results suggest that personalized learning experiences with assistive technology can significantly help bring about real inclusion in educational settings.

INTRODUCTION

One of the key issues in education today is inclusive education. This concept emphasizes how important it is for all students, including those with special needs, to have the opportunity to learn together in a supportive environment. In such a situation, assistive technology appears to be a promising tool to help students with special needs reach their full potential (Parmigiani et al., 2021). The main goal of this study is to investigate in depth how assistive technology supports student learning in order to achieve truly inclusive education. The process of guaranteeing that all children, including those with impairments, have access to shared learning is known as inclusive education, according to UNESCO (2009). According to Cook & Hussey (2002), Whatever assistive technology, or gadget that helps people with disabilities maintain, increase, or improve their functional abilities is considered assistive technology. Overall, assistive technology, such as text-to-speech (TTS) devices, is essential to achieving inclusive education.

These technologies help students with accessibility issues such as reading difficulties or visual impairments. According to Connolly et al. (2024), incorporating these technologies into the curriculum from the start can make the learning environment more welcoming and inclusive for all students, while maintaining students' special needs. In education, special attention is paid to the way TTS technology can help students achieve certain goals, such as improving their understanding of reading and increasing their engagement with special needs. TTS devices are shown in studies that they not only help students who have difficulty reading but also increase students' confidence in completing academic tasks (Silvestri et al., 2022).

This method emphasizes how important it is to adapt technology to the unique needs of students. The implementation of this technology is focused on tailoring learning tools and methods to the specific needs of students based on their physical and cognitive profiles (Troussas et al., 2020). For example, students with visual impairment or dyslexia can have a personalized reading experience through software such as Natural Reader. These customizations show how important it is to ensure every student has optimal access to educational materials (Wood et al., 2018). These concepts will be applied in this study to instruct special education pupils in inclusive schools. There are still certain knowledge gaps that need to be filled, even though the use of assistive technology in inclusive education has been the subject of numerous projectsh (Mishra & Koehler, 2006; Rao & Gravel, 2023).

One of the most recent discoveries in this research is the use of XpressiveTalk, a visually expressive text-to-speech (TTS) technology intended to assist individuals with autism spectrum conditions (ASC) in improving their emotion recognition and social interaction skills. Unlike conventional TTS tools that focus on reading comprehension, XpressiveTalk combines dynamic, near-real visual avatars with emotional expressions. This technique, as described by Cassidy et al. (2016), allows individuals with ASC to practice and improve emotional understanding in a more personalized and controlled environment. This study represents a novel approach that combines advanced facial animation techniques with educational and therapeutic interventions to promote inclusion, especially for those with social and emotional processing difficulties.

RESEARCH METHODS

In order to broaden the scope of analysis and deepen the understanding of the issue of primary focus, this study adopts the methods and results of previous research from expert researchers who have conducted excellent field experiments compiled by Cassidy et al. (2016).

By utilizing data and results from previous studies, we were able to integrate a broader perspective and make new contributions to understanding the phenomenon under study. This process not only enriches theoretical understanding but also strengthens the validity of the research.

Experimental methods were used in this study to evaluate how effective XpressiveTalk technology is as an emotion aid for individuals with autism spectrum conditions (ASC). The technique generates visual avatars that closely resemble real human expressions, allowing users to identify various emotions with customizable intensity levels. To improve the emotional representation of the avatar, a Hidden Markov Model (HMM) was used and the Cluster Adaptive Training (CAT) method was employed. The control group with ASC and the control group without ASC were the two groups used in this study. They were asked to identify emotions in two categories of videos: real face videos and synthetic videos created by XpressiveTalk. Five main emotional expressions-happy, sad, angry, fearful, and neutral-were used in each video with the same neutral sentence. The videos were randomly displayed through a secure online platform, which allowed for bias control. Participants provided responses in three dimensions to ensure the validity of the measures: emotion recognition accuracy, preference for the video, and assessment of the realism of the expression.

It seems that the results of these responses were examined using ANOVA analysis. The accuracy and preference of the ASC and control groups differed significantly, particularly when it came to specific emotions like feelings of sadness, which were more simple to recognize in the XpressiveTalk state as opposed to human faces. The first stage in creating XpressiveTalk was to use the actress's voice and video data to train the model to produce dynamic emotional expressions. Relevant facial movement modeling is made possible by this visual model's Active Appearance Model (AAM) technique. Complex expressions, with changeable intensity levels for each emotion were created when audio and visual parameters were combined. XpressiveTalk is used as a test medium in situations similar to real-world social interactions to evaluate its efficacy. By allowing passengers to interact with the created expressions, the system should offer a predictable and secure learning environment. According to the results, XpressiveTalk can be used as an intervention technique to help people with ASC become more considerate of social emotions, especially low-intensity expressions like sadness and neutrality.

XpressiveTalk technology, a visually expressive text-to-speech (TTS) system, used in a previous study has shown that people with autism spectrum conditions (ASCs) are linked to emotion processing. Using speech and video data, XpressiveTalk can create visual avatars with customizable dynamic emotional expressions that closely resemble the real world. After using the actress' video corpus to develop a face model, an HMM (Hidden Markov Model) based statistical model is trained using the Cluster Adaptive Training (CAT) method to change the intensity and complexity of the expression. The system generates videos that have different emotional intensities, including neutral, sad, happy, angry, fearful, and so on (fig. 1).

Fig. 1. A model of active appearance. (a) shows the look of the mesh. Synthesis outcomes for (b) neutral, (c) tender, (d) happiness (e) sadness, (f) anxious and (g) frustrated are shown. (Cassidy et al., 2016).

In this study, people who used ASC and people who did not use ASC were asked to identify emotions from videos of real and synthetic faces displaying neutral sentences in five different emotional tones. The videos were presented randomly on a secure online platform. For each of the videos (fig. 2), participants provided responses that included emotion recognition, preference judgment, and degree of realism. According to ANOVA analysis, individuals with ASC were overall less accurate in identifying emotions than the control group; however, individuals with ASC demonstrated the ability to identify specific emotions, such as the expression of sadness in the context of XpressiveTalk compared to real facial expressions. This demonstrates the ability of this technology to assist in interventions.

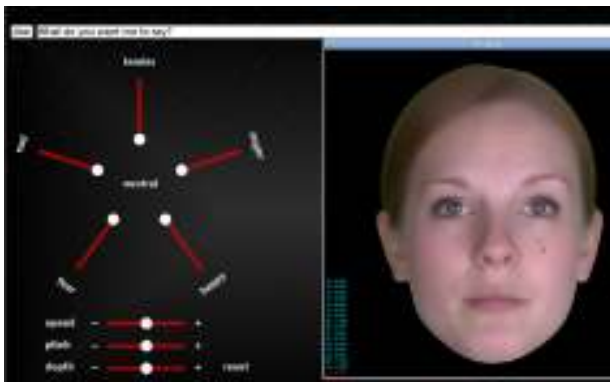

Fig. 2. Screenshot of the interface for synthesizing with XpressiveTalk. The interface allows for inputting text and setting the values of the expression parameters, which are used to create the animation of the talking avatar (Cassidy et al., 2016).

RESULTS AND DISCUSSION

An analysis of the data shown in Table 1 which shows the proportion of emotion inferences for XpressiveTalk and real faces among the typical group begins this part. Accurate emotion inference across the two mediums is revealed by the data, which is arranged into confusion matrices. This comparative analysis focuses on the patterns, distinctions, and possible consequences for recognizing emotional cues using AI-generated facial expressions as opposed to actual human emotions. The discussion that follows explores these findings' relevance to current theories and actual-life situations.

Table 1

The percentage of emotion inferences for XpressiveTalk and real faces in the typical group is displayed through confusion matrices.

		Real face					XpressiveTalk				
		Correct emotion					Correct emotion				
		Happy	Sad	Angry	Afraid	Neutral	Happy	Sad	Angry	Afraid	Neutral
Emotion response	Happy	87.2	0.0	0.0	0.0	1.9	66.0	0.0	1.3	0.0	1.9
	Sad	0.0	74.4	0.0	5.8	3.2	0.0	85.9	0.6	10.9	0.0
	Angry	1.3	0.0	94.9	2.6	1.9	1.9	0.0	64.7	1.9	3.2
	Afraid	0.6	22.4	1.9	89.1	1.9	15.4	12.2	15.4	85.9	0.0
	Neutral	10.9	3.2	3.2	2.6	91.0	16.7	1.9	17.9	1.3	94.9

Table 2

Confusion matrices that display the ASC group's percentage of emotion inferences for both XpressiveTalk and real faces.

		Real face					XpressiveTalk				
		Correct emotion					Correct emotion				
		Happy	Sad	Angry	Afraid	Neutral	Happy	Sad	Angry	Afraid	Neutral
Emotion response	Happy	77.5	0.0	1.9	0.0	2.5	43.8	0.0	2.5	0.0	6.9
	Sad	0.0	60.0	0.0	13.8	4.4	5.0	79.4	2.5	11.3	3.8
	Angry	4.4	1.3	86.3	5.6	2.5	1.3	0.0	53.1	6.3	5.0
	Afraid	2.5	20.6	2.5	68.8	3.1	14.4	13.8	19.4	60.0	0.6
	Neutral	15.6	18.1	9.4	11.9	87.5	35.6	6.9	22.5	22.5	83.8

In both situations, tables 1 and 2 (Cassidy et al., 2016) show that participants overall identified emotions more correctly than not in both real and XpressiveTalk situations. However, compared to the control group, participants with ASC showed greater difficulty identifying "happy" and "angry" in the XpressiveTalk condition compared to the real situation (Cassidy et al., 2016).

DISCUSSION

1. XpressiveTalk Technology Development

XpressiveTalk is intended to create visual avatars with dynamic emotional expressions that more closely resemble reality. This method generates a face model from a video corpus of an actress. This face model is then trained using a Hidden Markov Model (HMM) through Cluster Adaptive Training (CAT). The system allows customization of the intensity and complexity of emotional expressions, such as neutral, happy, sad, angry, and fearful, resulting in a flexible and relevant interactive experience for the needs of individuals with Autism Spectrum conditions (ASC).

2. Evaluation Method

In this study, two groups of participants, one group with ASC and one control group, were asked to identify their feelings from videos of synthetic and real faces. Five different emotional tones were used in these videos, which were presented randomly through a secure online platform. Each video was asked to be identified through emotion, preference judgment, and level of realism. With this method, an evaluation of the effectiveness between XpressiveTalk synthetic videos and real faces can be made.

3. Main results of the study

from the data analysis was that individuals with ASC were overall less accurate in recognizing emotions compared to the control group. However, they were better at recognizing expressions of sadness in the context of XpressiveTalk than real faces. This demonstrates the ability of XpressiveTalk to support the recognition of certain emotions in individuals with ASC, especially in a customizable environment.

4. Implications and Potential Interventions

The results showed that XpressiveTalk can assist individuals with ASC in emotion processing and social attention. With the ability to change the intensity of emotions, this technology offers a safe and flexible approach to support technology-based interventions. The research also provides important insights into ways in which technology can be used to enhance learning and social experiences for groups with special needs.

CONCLUSIONS AND RECOMMENDATION

The results of this study demonstrate the transformative potential that assistive technology can have in promoting inclusive education. The analysis shows that tools such as XpressiveTalk improve reading comprehension and emotional recognition and social skills of students with ASC. Although individuals with ASC may find it more difficult to identify some emotions, the use of TTS technology in a controlled environment may be more effective than conventional approaches to identifying emotions. This suggests that educational tools should be customized to meet the specific needs of each student and ensure each student has optimal access to educational resources. In addition, the study found that there is not enough literature on the use of assistive technology in inclusive education, which suggests that further research is needed. Overall, incorporating assistive technology into the curriculum is essential to create a supportive and inclusive learning environment that allows all students to reach their full potential.

REFERENCES

Cassidy, S. A., Stenger, B., Van Dongen, L., Yanagisawa, K., Anderson, R., Wan, V., ... & Cipolla, R. (2016). Expressive visual text-to-speech as an assistive technology for individuals with autism spectrum conditions. *Computer Vision and Image Understanding*, *148*, 193-200.

Cook, A. M., & Hussey, S. M. (2002). Assistive technologies. *Principles and practice*.

Connolly, A., Bailey, S., Lamont, R., & Tu, A. (2024). Factors associated with assistive technology prescription and acceptance in motor neurone disease. *Disability and Rehabilitation: Assistive Technology*, *19*(6), 2229-2238.

Halkiopoulos, C., & Gkintoni, E. (2024). Leveraging AI in E-Learning: Personalized Learning and Adaptive Assessment through Cognitive Neuropsychology-A Systematic Analysis. *Electronics*, *13*(18), 3762. https://doi.org/10.3390/electronics13183762.

Izzo, M., Yurick, A., & McArrell, B. (2009). Supported eText: Effects of Text-to-Speech on access and Achievement for High School Students with Disabilities. *Journal of Special Education Technology*, 24, 20-9. https://doi.org/10.1177/016264340902400302.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers college record*, *108*(6), 1017-1054.

Nieradka, F., & Kossewska, J. (2023). Camouflage, self-esteem, and gratitude in young adults with autism spectrum conditions in Poland. *International Journal of Special Education*, *38*(2), 188-198.

Parmigiani, D., Benigno, V., Giusto, M., Silvaggio, C., & Sperandio, S. (2021). E-inclusion: online special education in Italy during the Covid-19 pandemic. *Technology, pedagogy and education*, *30*(1), 111-124.

Rao, K., Gravel, J. W., Rose, D. H., & Tucker-Smith, T. N. (2023). Universal Design for Learning in its 3rd decade: A focus on equity, inclusion, and design. *International encyclopedia of education*, *6*, 712-720.

Silvestri, R., Holmes, A., & Rahemtulla, R. (2022). The interaction of cognitive profiles and text-to-speech software on reading comprehension of adolescents with reading challenges. *Journal of Special Education Technology*, *37*(4), 498-509.

Shaywitz, S. E., Shaywitz, J. E., & Shaywitz, B. A. (2021). Dyslexia in the 21st century. *Current opinion in psychiatry*, *34*(2), 80-86.

Srivastava, S., Varshney, A., Katyal, S., Kaur, R., & Gaur, V. (2021). A smart learning assistance tool for inclusive education. *Journal of Intelligent & Fuzzy Systems*, *40*(6), 11981-11994.

Troussas, C., Krouska, A., & Virvou, M. (2020). Using a multi module model for learning analytics to predict learners' cognitive states and provide tailored learning pathways and assessment. *Machine learning paradigms: Advances in learning analytics*, 9-22.

Vaughn, S., Miciak, J., Clemens, N., & Fletcher, J. M. (2024). The critical role of instructional response in defining and identifying students with dyslexia: A case for updating existing definitions. *Annals of Dyslexia*, *74*(3), 325-336.

Wattel, L. L., Walsh, R. J., & Krabbendam, L. (2024). Theories on the link between autism spectrum conditions and trans gender modality: A systematic review. *Review Journal of Autism and Developmental Disorders*, *11*(2), 275-295.

Wood, S. G., Moxley, J. H., Tighe, E. L., & Wagner, R. K. (2018). Does the use of text-to-speech and related read-aloud tools improve reading comprehension for students with reading disabilities? A meta-analysis. *Journal of learning disabilities*, *51*(1), 73-84.

Wolf, M., Gotlieb, R. J., Kim, S. A., Pedroza, V., Rhinehart, L. V., Tempini, M. L. G., & Sears, S. (2024). Towards a dynamic, comprehensive conceptualization of dyslexia. *Annals of Dyslexia*, 1-22.

Zwarych, F. (2023). Inclusive Education with Assistive Technology. *Technology and the Curriculum: Summer 2023*.