JOURNAL OF MODERN SCIENCE
SPECIAL ISSUE
3/57/2024
www.jomswsge.com

DOI: doi.org/10.13166/jms/191136

KONRAD NIDERLA
WSEI University in Lublin, Poland
ORCID iD: orcid.org/0000-0003-1280-0622

WIESŁAW KOWALSKI
WSEI University in Lublin, Poland
ORCID iD: orcid.org/0000-0001-8402-2837

WIESŁAW POLESZAK
WSEI University in Lublin, Poland
ORCID iD: orcid.org/0000-0002-5253-0789

UTILIZING ROBOTS FOR VOICE AND SOUND ANALYSIS IN THERAPY: ENHANCING EMOTIONAL UNDERSTANDING IN CHILDREN WITH AUTISM SPECTRUM DISORDERS

WYKORZYSTANIE ROBOTÓW DO ANALIZY GŁOSU I DŹWIĘKU W TERAPII: POPRAWA ZROZUMIENIA EMOCJONALNEGO U DZIECI Z ZABURZENIAMI ZE SPEKTRUM AUTYZMU

Abstract

Exploring and understanding emotions when treating children with autism spectrum disorders (ASD) is of fundamental importance for both therapists and children themselves. People with ASD often use specific communication methods, which may make it difficult to express their feelings using conventional verbal and non-verbal techniques. Therefore, a deep analysis of the subtle aspects of their speech, intonation, rhythm, and other auditory forms of expression can provide valuable clues about their emotions, needs, and reactions to therapeutic activities. Children's emotional responses can manifest in many ways, including voice modulation, which can signal anger, sadness, or happiness. Therapists can better interpret the child's intentions and reactions to surrounding stimuli using voice analysis methods. For example, a flat tone of voice may indicate apathy or anxiety, while a high volume may indicate excitement or tension. Additionally, various unconventional sounds a child makes can provide valuable information about their emotional and mental state. To capture these signals, therapists can use advanced equipment for recording and analyzing sounds and robots equipped with image and sound recording functions, which help in therapy. These tools allow you to register and recognize nuances that an experienced observer may overlook. The use of sound analysis in the treatment of children with ASD may play a significant role in assessing the effectiveness of the therapy. Systematic sessions during which changes in the way of speaking and other sound expressions are assessed provide information on the treatment progress and the need for possible corrections of the techniques used. This makes it possible to tailor the therapy more precisely to each child's specific needs, which leads to improved effectiveness. Children with ASD may have difficulty using speech to express their emotions or fully understand other people's body language. In this context, a deep analysis of their speech can facilitate understanding their internal state. It is important to remember that emotions in children with ASD may be expressed in less typical ways than in peers without this disorder. Assistive technologies such as speech and audio analysis software can complement traditional observational methods by providing more objective data about children's emotional responses. Combining facial expression analysis with speech and sound analysis allows for a comprehensive approach to understanding the emotions and behaviors of children with ASD. This integration will enable therapists to adapt their work methods better, significantly supporting each child's emotional and social development and increasing the therapy's effectiveness. Automation of processes in the rehabilitation of children with developmental disorders, such as autism spectrum disorders (ASD), may be an essential element of modern therapy. In particular, automation plays a fundamental role in monitoring progress, providing suggestions, and implementing modifications to therapeutic processes.

Streszczenie

Badanie i zrozumienie emocji podczas terapii dzieci z zaburzeniami spektrum autyzmu (ASD) ma fundamentalne znaczenie zarówno dla terapeutów, jak i samych dzieci. Osoby z ASD często posługują się specyficznymi metodami komunikacji, co może utrudniać wyrażanie swoich uczuć przy użyciu konwencjonalnych technik werbalnych i niewerbalnych. Dlatego głęboka analiza subtelnych aspektów ich mowy, intonacji i rytmu, a także innych dźwiękowych form ekspresji może dostarczyć cennych wskazówek na temat ich emocji, potrzeb i reakcji na działania terapeutyczne. Reakcje emocjonalne dzieci mogą objawiać się na wiele sposobów, w tym poprzez modulację głosu, która może sygnalizować stany takie jak złość, smutek lub szczęście. Stosując metody analizy głosu, terapeuci mogą lepiej interpretować intencje dziecka i jego reakcje na otaczające bodźce. Na przykład płaski ton głosu może wskazywać na apatię lub niepokój, podczas gdy zwiększone natężenie głosu może być oznaką podniecenia lub napięcia. Dodatkowo różne niekonwencjonalne dźwięki wydawane przez dziecko, również mogą dostarczyć cennych informacji na temat jego stanu emocjonalnego i psychicznego. Do wychwytywania tych sygnałów terapeuci mogą wykorzystać zaawansowany sprzęt do rejestracji i analizy dźwięków, a także roboty wyposażone w funkcję rejestracji obrazu i dźwięku, które służą pomocą w terapii. Narzędzia te pozwalają na zarejestrowanie, a następnie rozpoznanie nawet tych niuansów, które nawet doświadczony obserwator może nie zauważyć. Zastosowanie analizy dźwiękowej w terapii dzieci z ASD może odegrać znaczącą rolę w ocenie skuteczności prowadzonych terapii. Systematyczne sesje, podczas których oceniane są zmiany w sposobie mówienia i innych wypowiedziach dźwiękowych, dostarczają informacji o postępie terapii, a także konieczności ewentualnej korekty stosowanych technik. Dzięki temu możliwe jest dokładniejsze dopasowanie terapii do konkretnych potrzeb każdego dziecka, co z kolei prowadzi do poprawy jej efektywności. Dzieci z ASD mogą mieć trudności z używaniem mowy do wyrażania swoich emocji lub pełnego zrozumienia mowy ciała innych osób. W tym kontekście głęboka analiza ich mowy może ułatwić zrozumienie ich stanu wewnętrznego. Należy pamiętać, że emocje u dzieci z ASD mogą wyrażać się w mniej typowy sposób niż u rówieśników bez tego zaburzenia. Technologie wspomagające, takie jak oprogramowanie do analizy mowy i dźwięku, mogą stanowić cenne uzupełnienie tradycyjnych metod obserwacyjnych, dostarczając bardziej obiektywnych danych na temat reakcji emocjonalnych dzieci. Połączenie analizy wyrazu twarzy z analizą mowy i dźwięków pozwala na kompleksowe podejście do zrozumienia emocji i zachowań dzieci z ASD. Integracja ta pozwala terapeutom na lepsze dostosowanie metod pracy, co znacząco wspiera rozwój emocjonalny i społeczny każdego dziecka, zwiększając efektywność terapii. Automatyzacja procesów w rehabilitacji dzieci z zaburzeniami rozwoju, takimi jak zaburzenia ze spektrum autyzmu (ASD), może być ważnym elementem współczesnej terapii. W szczególności automatyzacja odgrywa zasadniczą rolę w monitorowaniu postępów oraz dostarczaniu sugestii i wdrażaniu modyfikacji w procesach terapeutycznych.

Keywords: *Emotions, voice, sound analysis, robot, rehabilitation, ASD, therapeutic interventions, machine learning, artificial intelligence*

Słowa kluczowe: *Emocje, głos, analiza dźwięku, robot, rehabilitacja, ASD, interwencje terapeutyczne, uczenie maszynowe, sztuczna inteligencja*

Introduction

Exploring and understanding emotions, tone of voice, and other auditory aspects when treating children with autism spectrum disorder (ASD) is essential for both therapists and children themselves. People with ASD often use specific ways of communicating, which may make it difficult to express their feelings using conventional verbal and non-verbal techniques. This is why analyzing delicate aspects of their speech, intonation, rhythm, and other sonic expressions can provide clues about their emotions, needs, and responses to therapeutic interventions. Children's emotional responses can manifest themselves in many ways, including vocal inflections that signal anger, grief, or happiness. Therapists can better interpret the child's intentions and reactions to surrounding stimuli using voice analysis methods. For example, a steady tone of speech may indicate apathy or concern, while an increased voice intensity may indicate excitement or tension. Additionally, various unconventional sounds a child produces, such as purring or rustling, can provide valuable information about their emotional and mental state. To capture these signals, therapists can use advanced equipment for recording and analyzing sounds and robots equipped with image and sound recording functions that support therapy. These tools allow you to register and recognize even those nuances that may be unnoticeable to an experienced observer.

The use of sound analysis in the therapy of children with autism spectrum disorders (ASD) may play an essential role in assessing the effectiveness of the treatment. Systematic sessions during which changes in the way of speaking and other sound expressions are assessed provide information about progress in therapy and the need to make any corrections in the techniques used. This makes it possible to tailor the treatment more precisely to the specific needs of individual children, improving its effectiveness. In everyday clinical practice,

assessing children's sounds may be as important as diagnostic and observational skills. Training in identifying characteristic aspects of speech and sounds should, therefore, be an integral part of the education of therapists working with children with ASD and similar disorders. Children with ASD may experience difficulty using speech to express their emotions or fully understanding other people's body language. In this context, an in-depth analysis of their speech can facilitate understanding their internal state. It is important to remember that emotions in children with ASD may be expressed in less typical ways than in their peers without this disorder. Assistive technologies, such as speech and audio analysis software, can serve as a valuable complement to traditional observational methods by providing more objective data about children's emotional responses. Combining the analysis of facial expressions with the analysis of speech and other sounds allows for a comprehensive approach to understanding the emotions and behaviors of children with ASD. Such integration will enable therapists to adapt their work methods better, significantly supporting each child's emotional and social development and increasing the therapy's effectiveness.

Automation of processes in the rehabilitation of children with developmental disorders, such as autism spectrum disorders (ASD), may be an essential element of modern therapy. In particular, automation plays a fundamental role in monitoring progress, providing suggestions, and implementing modifications to therapeutic processes. Emotion analysis is of invaluable importance in this context, as it enables the collection and processing of objective data about the emotional state of patients, which directly translates into therapeutic effectiveness. Modern technologies allow objective and precise tracking of subtle changes in children's behavior and emotions, which may be invisible to the human eye.

Advantages of automation in the analysis of emotions in the context of rehabilitation:

1. Automation allows for continuous monitoring and documentation of even small changes in the child's emotional expression, which is crucial for adequate assessment of therapy's effectiveness. Technologies such as facial and voice analysis allow therapists to access regular reports that inform about the need for intervention or confirm the efficacy of the methods used.

2. therapists can adapt treatment methods to the child's needs and reactions by constantly providing up-to-date information. For example, increasing a child's stress level may result in introducing relaxation techniques or other strategies to reduce tension.
3. Automation also makes it easier for parents to understand their children's progress and needs. Detailed reports and analyses provided by the systems can help parents better support their children's development in the home environment and participate more consciously in the therapeutic process.
4. Automating diagnostic and therapeutic processes allows for more effective use of therapy time. Therapists can focus on more complex aspects of therapy, having already had access to the data analyzed by the system.
5. Advanced technologies such as machine learning and artificial intelligence are becoming the foundation in automating emotion analysis. These systems can be trained to recognize specific emotional and behavioral patterns in children, which allows for monitoring progress and quick detection of possible development problems. This, in turn, allows for immediate reaction and adjustment of the therapeutic plan in response to the observed changes.

As a result, automating emotion analysis in the rehabilitation of children with developmental disorders brings numerous benefits, improving not only therapeutic effectiveness but also supporting parents in the everyday therapy process. Thanks to modern technologies, therapists can more precisely adapt therapy methods to each child's individual needs.

The use of rehabilitation robots in the therapy of children with developmental disorders, including ASD, represents an innovative and groundbreaking approach that has the potential to enrich the therapeutic process profoundly. These interactive machines engage children in play and provide precise and objective methodologies for assessing and supporting their development. These robots can be programmed to conduct a variety of games and activities that are particularly attractive to young users, which significantly increases their motivation and involvement in therapy. This way of conducting sessions allows children to participate more actively in classes and

derive greater joy and satisfaction from them. Moreover, these robots can be adapted to the individual reactions of each child, which allows the therapy to be tailored to their specific needs and reactions. Advanced algorithms enable these robots to analyze a child's gestures, facial expressions, and voice intonation, allowing them to adjust their interactions to maximize their therapeutic impact on their development. Robots equipped with advanced sensors and cameras ensure continuous monitoring and documentation of therapeutic progress, providing therapists with invaluable data on the effectiveness of their activities. This information can then be used to adjust further and optimize treatment plans. Additionally, these robots can teach social skills such as emotion recognition and verbal and non-verbal communication. Children have the opportunity to acquire these critical skills in a controlled but dynamic environment, which often brings better results than traditional educational methods. Robots in the therapy of children with developmental disorders accelerate their emotional and social development and contribute to a more effective and involved therapeutic process. As a result, therapy using rehabilitation robots opens new perspectives for a better future for children with ASD, providing them with more comprehensive and tailored support (Schiavo et al., 2024).

Interactions with rehabilitation robots may prove to be less emotionally burdensome for children with autism spectrum disorders (ASD), who sometimes experience fear in contact with other people. These robots ensure consistency and predictability of actions, which is crucial for the emotional comfort of children with ASD. Robot models such as NAO, Pepper, and PARO – a seal-like robot collected and reviewed in the 2021 scientific papers of Szymona and colleagues and Puglisi and colleagues from 2022, are examples of the use of technology to reduce stress and improve the ability to interact socially. These robots are used primarily in sensory therapy, and other robots created for specific research, such as the robot described by Niderla and Maciejewski in 2021, demonstrate how technology can support rehabilitation and promote the development of social skills. Robots in therapeutic sessions create new opportunities to support children with developmental disorders, offering innovative and effective therapy methods. These interactive tools help children cope with everyday emotional and social challenges and enable therapists to take a more individualized approach to each child, thereby increasing therapeutic effectiveness. Robots in

therapy open the door to more advanced forms of rehabilitation, enabling children to acquire the necessary skills in a safe and controlled environment, which is crucial for their further development and social adaptation.

Rehabilitation robot

The rehabilitation system used for data acquisition functions is based on the cooperation of three main components, which create an integrated support mechanism for patients with various therapeutic needs. The first element of the system is a device dedicated to the patient, which in the discussed case is a rehabilitation robot. It is an interactive agent who cooperates directly with the patient during rehabilitation sessions. The second component is the rehabilitator's device – a computer equipped with specialized software. This software supervises and controls the robot's operation, managing activities based on the entered data and current interaction with the patient. This computer also serves as an interface for the therapist, enabling him to monitor the course of therapy and make necessary modifications to the robot's reactions and interactions.

The third key element of the system is the machine learning module. Its primary function is to process data collected during the patient's interaction with the robot, analyze this data, and formulate proposals for therapeutic activities that best respond to the patient's current needs. Using artificial intelligence algorithms, this module constantly learns from new data, enabling optimization and personalization of the therapeutic process. The interactions between these three components are designed to ensure smooth communication and synergistic action, which translates into greater effectiveness of therapy. A system diagram illustrating the dependencies and information flow between components is presented in Figure 1.

The rehabilitation robot has been specially designed to engage children in interactions inspired by forms found in nature. Its design resembles a rabbit, which makes it especially attractive to younger users. The friendly, zoomorphic appearance of the robot, which combines humanoid features with rabbit-like elements, makes it charming and attracts children's attention. The robot is covered with a soft, gray material resembling fur, which increases its visual

and tactile attractiveness, adding sensory impressions for children. The robot consists of several critical structural modules, such as the chassis, arms, neck, and head, each made with precision using 3D printing and laser cutting techniques. Thanks to these modern technologies, it is possible to precisely and flexibly refine each element, ensuring the device's functionality and durability. The robot chassis has an advanced drive system that allows smooth movement in different directions, achieved using DC motors with gears. The robot also has optical sensors in its chassis that will enable it to avoid obstacles and prevent accidental falls. The robot arms, made using 3D printing, have servo motors that allow complex movements such as lifting, tilting, or bending at the elbows. This allows the robot to grasp and manipulate light objects, significantly enriching interactions with children during therapy sessions.

Figure 1. *Diagram of the interaction of components of the rehabilitation robot system*

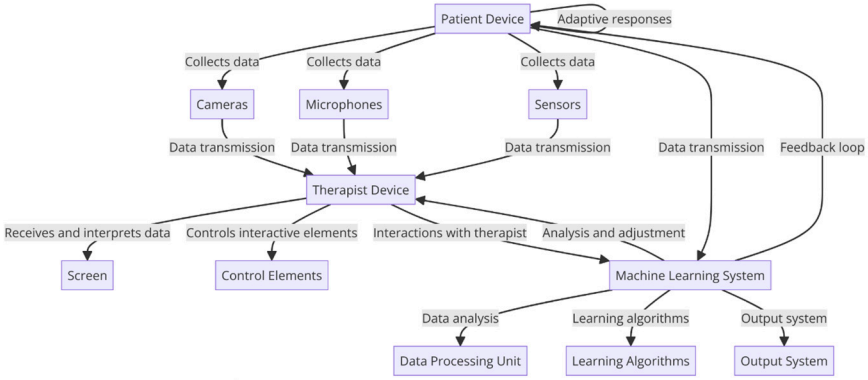


Figure 2. *Rehabilitation robots in the form of a controlled or autonomous toy*

The robot head plays a key role as a communication and interaction center, equipped with speakers, a microphone, a camera, and a touch screen, giving it considerable functionality. Using a microphone and camera to record sounds and images enables the analysis of the child's sounds, voice, and emotions. Thanks to this, the robot can respond appropriately and adapt to the user's current needs and moods, significantly enriching interactions. The robot was designed so that its head movements mimic human gestures such as nodding and tilting, allowing for more intuitive expressions of emotions such as interest or sadness. The robot's independently moving ears, and movable tail add a dimension of expression, making its presence more dynamic and natural. The robot is more than just a toy; it is also an educational tool, which, thanks to its advanced technological functions and well-thought-out design, is an essential element in the rehabilitation and education of children. It supports children's emotional and cognitive development through interactive play and learning, acting as a mediator between the therapist and the child, allowing for more effective and engaging therapy sessions.

The rehabilitator's device in the discussed system serves as a critical interface for receiving data from the patient's device, enabling effective user

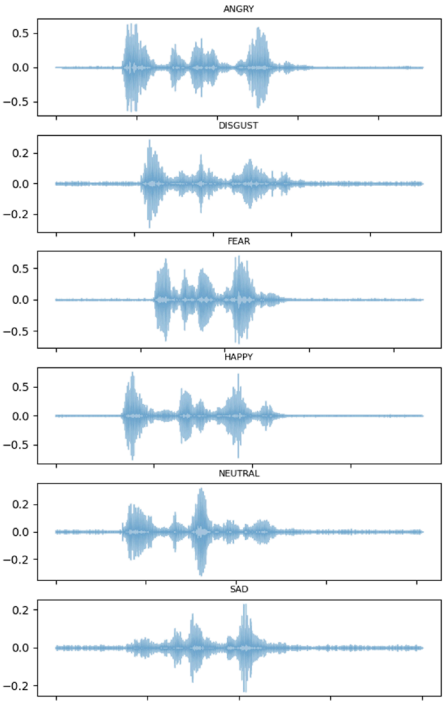
interaction. This device has a screen that displays real-time data and recommendations generated by a machine-learning system in a feedback loop. This information allows the therapist to continuously modify therapeutic activities or adapt the patient's device responses to the patient's current needs. An essential link in this advanced system is the machine learning module, which processes collected data using advanced algorithms to analyze and identify behavioral and emotional patterns. This module assesses the patient's emotional state and, on this basis, adapts the proposed therapeutic activities to support the rehabilitation process optimally. The data processor analyzes the information sent from the patient's device, and machine learning algorithms appropriately adapt the operation of the device in response to the identified needs and reactions of the patient. The system also includes a dynamic feedback loop, which allows for the continuous transfer of adjustments and recommendations between the machine learning module, the rehabilitator's device, and the patient's device. This solution provides for ongoing adaptation of the therapy to the changing conditions and needs of the patient. The system can also automatically adjust patient device responses to meet evolving therapeutic requirements. In this way, the integrated rehabilitation system uses modern technologies and advanced machine learning techniques to create a responsive and adaptive therapeutic environment that is tailored to the individual needs of patients, especially those with disorders such as autism. Through continuous data collection and analysis, this system not only enhances traditional therapy methods but also provides comprehensive data sets that can be used to develop further and improve treatment methods and therapies, contributing to significant progress in rehabilitation.

Voice and sound analysis methods

Emotion analysis using sounds from microphones uses advanced technologies in pattern recognition, artificial intelligence, and speech processing, which opens new possibilities for understanding human emotions. These tools allow you to precisely identify and analyze subtle signals that may indicate a person's various emotional states. Voice pitch analysis is another method where microphones collect audio data, which is then processed to analyze various acoustic features of the voice, such as pitch, tempo, loudness, and modulation (Landowska et al., 2022; Al-Talabani, 2015). These acoustic features of speech can reveal the speaker's emotions, enabling the identification of states such as stress, excitement, or calm. Speech processing, which combines the analysis of words and their context, also plays a crucial role in emotion recognition. Using natural language processing techniques, systems can analyze the content of statements and identify keywords or phrases related to emotions, which provides insight into what a person is saying and what emotions may accompany their statements. Combining visual and audio data allows for even more accurate and comprehensive analyses. By integrating information from various sources, systems can better interpret complex emotional signals, which is particularly important in therapies, social interactions, and commercial applications, such as analyzing customer reactions. Each of these methods uses advanced algorithms and specialized software for data processing and analysis, and their effectiveness may depend on many factors, including the quality of the equipment, environmental conditions, and the specific nature of the application. Despite these challenges, emotion analysis technologies using cameras and microphones open new possibilities in understanding and interacting with the emotional world of patients. In the case of children with autism, who often encounter difficulties in verbal and non-verbal communication, modern technologies for analyzing emotions using cameras and microphones can significantly contribute to their therapy and rehabilitation. The rehabilitation robot, as a complex system for interacting with children, is equipped with a camera and a microphone that record facial expressions and gestures, as well as the tone and modulation of the voice. The collected data is analyzed to detect emotions such as satisfaction, frustration, or interest. In

contrast, the microphones can analyze changes in the tone of voice, providing important information about the child's emotional state. With this data, we create a dataset linking the rehabilitation process with changes in patient moods to better support each child depending on their individual reactions and emotional needs. For example, if the system recognizes that a child is stressed, the system may suggest a calmer activity. A model may also encourage the child to interact when positive emotions are noticed, which helps build trust and emotional security. Speech analysis, which uses microphones to recognize intonation and voice inflection, can also be used to understand better the intentions and needs of children who may have difficulty expressing their thoughts and feelings clearly. This understanding is critical to personalizing therapy and tailoring interactions and subsequent rehabilitation steps, which can significantly improve the effectiveness of therapy. Integrating emotion analysis into therapy offers new opportunities for children with autism, helping them express and understand better and supporting the development of their social and communication skills development in a more controlled, dynamic, and interactive environment.

Figure 3. *Examples of sound intensity amplitudes as a function of time for statements with various emotional tones*



During the implementation of the rehabilitation robot project, advanced technologies and algorithms were used to conduct a detailed analysis of data recorded during the device's interaction with children, enabling precise understanding and response to their emotional and communication needs. Let us take as an example the analysis performed on the CREMA-D dataset (Cao et al., 2014). Crowd-sourced Emotional Multimodal Actors Dataset

is a dataset created for research on emotion recognition. This dataset contains audio and visual recordings of actors portraying various emotions. This collection offers audio and video recordings, allowing for the analysis of emotions from an acoustic and visual perspective. Actors portray Emotions such as anger, joy, sadness, fear, and neutrality. The dataset includes performances from many actors of different genders and ethnicities, which increases its diversity and possible applications. A graph of sample amplitudes as a function of time for utterances with different emotional tones is shown in Figure 3. These diagrams show graphs for the same actor's utterance in English: *Don't forget the jacket.* As we can see, even though the sentence is the same, the visual interpretation of the waveform is entirely different. We can observe the same in the frequency spectrogram (Figure 4). This graph contains three variables: frequency (y-axis), time (x-axis), and sound intensity represented by color. Here, too, the differences between individual statements are apparent. The intensity plot in the frequency domain also indicates differences (Figure 5).

Figure 4. *Spectrograms of statements with various emotional tones*

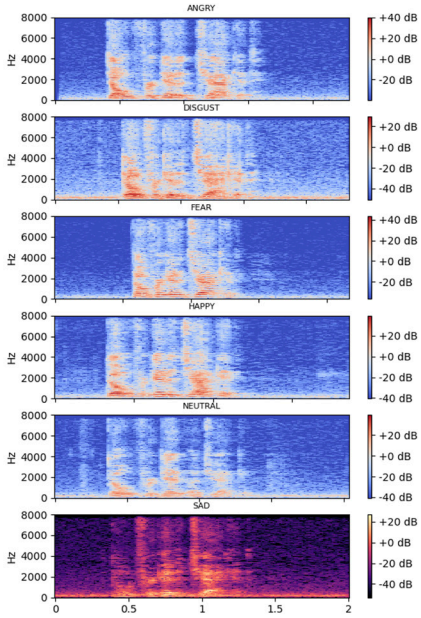
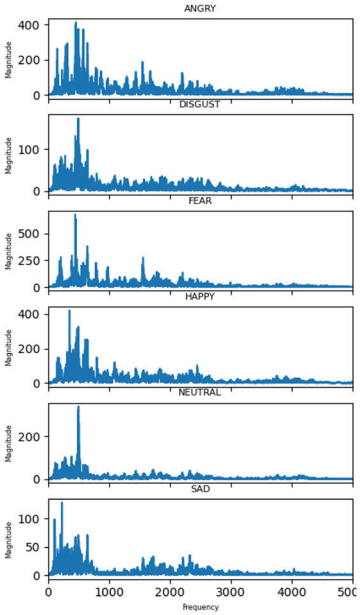


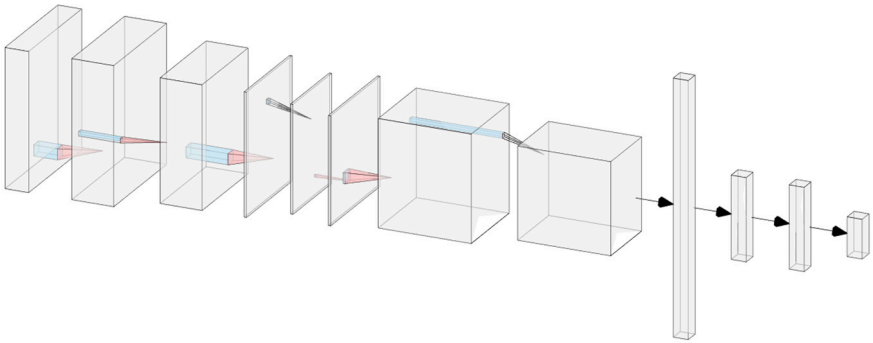
Figure 5. *Fourier transform of statements with different emotional overtones*



Although the analyzed signals are difficult to interpret based on the above graphs, they can be processed and classified using convolutional neural networks. Detecting emotions using the speech spectrogram and Mel-Frequency Cepstral Coefficients (MFCC) using Convolutional Neural Networks is an essential and dynamically developing field in the field of speech processing and emotion analysis (Zielonka et al., 2022). Originally designed to process visual data, these networks have found applications in recognizing emotional patterns in audio signals, offering exciting opportunities in human-machine interactions, psychological support systems, and security and monitoring applications. The primary step in analyzing emotions from speech is the appropriate processing of the sound signal. It usually starts with extracting features such as mel-cepstral coefficients (MFCC), pitch, energy, and tempo, which provide rich information about the signal characteristics. Traditional methods involve manually selecting these features, but you can automate this process with CNN. Convolutional networks can independently extract essential features directly from the raw audio signal or its spectrogram, a two-dimensional representation of the signal, where the vertical axis represents frequency, and the horizontal axis represents time. Convolutional networks consist of several layers, including convolutional layers that perform filtering operations on the input data. These filters adapt during the training process to detect specific data patterns relevant to a given task, such as emotion recognition. Convolutional layers are often followed by pooling layers that reduce the dimensionality of the data, which facilitates and speeds up the training process. At the network's end, one or more fully connected layers integrate information from the previously processed feature map into the final emotion class predictions. Training CNNs on audio data requires large datasets with labeled emotions. Training the network involves adjusting the network's weights to minimize the error between the model's predictions and the actual emotion labels. Using techniques such as cross-validation and regularization can increase the model's generalization on new, unknown data.

The visualization of the neural network model used to classify spectrograms is shown in Figure 6. From the left, the input layer, and further, convolutional, maxpooling, convolutional, maxpooling, dropout, convolutional, maxpooling, dropout, flatten, dense, dropout, dense.

Figure 6. *Visualization of the neural network model*



It is worth noting that models based on attention mechanisms have recently gained popularity and are now widely used in various sequential processing tasks, especially where mapping between two data formats is necessary. These models use previously predicted sequences and learn to map new sequences using an encoder-decoder approach, allowing for more efficient processing and better context understanding. The Listen-Attend-Spell (LAS) approach was one of the first to integrate methods based on LSTM and attention mechanisms. In this model, the encoder learns to extract features using bidirectional LSTMs. Then, the decoder, an attention-based unit, returns the output probability for the following sequence of characters. In speech emotion classification, LSTMs or their more advanced versions are often used when analyzing MFCC features as time series data. These models can capture feature changes for a given speech sample over time to predict emotion class. Convolutional neural networks (CNNs) can work on MFCC data in one-dimensional form or learn to classify spectrograms using two-dimensional filters. CNNs are particularly effective at recognizing visual patterns and can be adapted to analyze acoustic patterns, which makes them helpful in identifying complex emotional patterns in voice signals. The development of these technologies is of great importance not only in a technical context but also in practical applications, such as supporting therapeutic and medical systems, where recognizing patients' emotions can play a crucial role.

Conclusions

Analyzing emotions, voices, and other sounds in therapy sessions with children with autism spectrum disorders (ASD) is crucial to understanding their unique communication patterns and emotional states. These children often encounter difficulties in traditional forms of expressing feelings, which poses a challenge for them and therapists. Sound and facial expression analysis techniques, supported by modern technologies such as machine learning and artificial intelligence, allow for more accurate identification of subtle signals that may escape human observers. The introduction of rehabilitation robots and software for analyzing speech and facial expressions opens up new possibilities in the individualization of therapy, adapting therapy to the specific reactions and needs of the child. Thus, the use of these tools not only increases the effectiveness of treatment but also helps build more trusted therapeutic relationships, which is the foundation for adequate support for the development of children with ASD. Regular monitoring of progress and modifications in therapeutic approaches, supported by automation and data analysis, is essential to more effective and empathetic care for children with developmental disabilities.

References

Al-Talabani, Abdulbasit & Sellahewa, Harin & Jassim, Sabah. (2015). Emotion recognition from speech: Tools and challenges. Proceedings of SPIE - The International Society for Optical Engineering. 9497. 10.1117/12.2191623.

Bluestone, J. (2010). Materia autyzmu, przeł. M. Dąbrowska-Jędral, A. Klimek, Warszawa.

Cao, H., Cooper, D.G., Keutmann, M.K., Gur, R.C., Nenkova, A., Verma, R. (2014). CREMA-D: Crowd-sourced Emotional Multimodal Actors Dataset IEEE Trans Affect Comput. 2014 ; 5(4): 377–390. doi:10.1109/TAFFC.2014.2336244.

Landowska, A., Karpus, A., Zawadzka ,T., Robins, B., Erol, Barkana, D., Kose, H., Zorcec, T., Cummins, N. (2022). Automatic Emotion Recognition in Children with Autism: A Systematic Literature Review. Sensors (Basel). 22(4):1649. doi: 10.3390/s22041649. PMID: 35214551; PMCID: PMC8875834.

Niderla, K., Maciejewski, M. (2021).: Construction of a zoomorphical robot for rehabilitation of autistic children Journal of Physics: Conference Series 1782 (012024 IOP Publishing doi:10.1088/1742-6596/1782/1/012024

Parsons, L., Cordier, R., Munro, N., et al. (2019). A Play-Based, Peer-Mediated Pragmatic Language Intervention for School-Aged Children on the Autism Spectrum: Predicting Who Benefits Most. J Autism Dev Disord. 10.

Perez-Meana, Hector. (2007). Advances in Audio and Speech Signal Processing: Technologies and Applications. 10.4018/978-1-59904-132-2.

Puglisi, A, Caprì, T, Pignolo, L, Gismondo, S, Chilà, P, Minutoli, R, Marino, F, Failla ,C, Arnao, A.A, Tartarisco, G, Cerasa, A, Pioggia, G. (2022).Social Humanoid Robots for Children with Autism Spectrum Disorders: A Review of Modalities, Indications, and Pitfalls. Children (Basel). 9(7):953. doi: 10.3390/children9070953

Schiavo, F., Campitiello, L., Todino, M.D. (2024). Di Tore, P.A. Educational Robots, Emotion Recognition and ASD: New Horizon in Special Education. Educ. Sci. 14, 258. https://doi.org/10.3390/educsci14030258

Smith, T. (2012). Evolution of Research on Interventions for Individuals with Autism Spectrum Disorder: Implications for Behavior Analysts. Behavior Analysis in Practice, 1 (35): 101–113.

Szymona, B. , Maciejewski, M., Karpiński, R., Jonak, K., Radzikowska-Büchner, E., Niderla, K., Prokopiak, A. (2021). Robot-assisted autism therapy (RAAT). Criteria and types of experiments using anthropomorphic and zoomorphic robots. Review of the research. Sensors 21, 3720. https://doi.org/10.3390/s21113720

Zielonka, M., Piastowski, A., Czyzewski, A., Nadachowski, P., Operlejn, M., Kaczor, K. (2022). Recognition of Emotions in Speech Using Convolutional Neural Networks on Different Datasets. Electronics 11, 3831. https://doi.org/10.3390/ electronics11223831