Towards Automated Game-Based Early Screening for Language Disorder

Hamdan Al-Ali[1], **Elsa Soares**[2], **Gonçalo Leal**[2], **Ana Rita Valente**[3], **Nicole Agrela**[2], **Alexandra Marquis**[4], **Hanan Aldarmaki**[1]
[1]MBZUAI, [2]SpeechCare Center, [3]University of Aveiro, [4]UAE University
[1]{hamdan.alali; hanan.aldarmaki}@mbzuai.ac.ae

Abstract

This paper examines the potential of gamifying early childhood language disorder screening to make the process more accessible and scalable. We provide an overview of current practices in screening and assessment, and a description of our on-going work towards automation of early screening. By integrating developmental milestones into a video game format and employing automatic speech recognition and natural language processing, this approach aims to enhance the efficiency and reach of early screening in order to identify children who need further professional assessment.

1 Introduction

Language development is a crucial aspect of early childhood development, significantly impacting future academic success and social integration (Sunderajan and Kanhere, 2019). Traditional screening methods for developmental language disorders involve one-on-one sessions that, while effective, are resource-intensive, lengthy and not easily scalable (Eriksson et al., 2010). This process, combined with the global shortage of experienced Speech-Language Pathologists (SLPs) (Squires, 2013), presents a challenge in efficiently identifying children who could benefit from early intervention on a wide scale.

Gamification, the application of game-design elements in non-game contexts, is a powerful tool that can engage and motivate children, potentially transforming the screening process into an enjoyable, playful activity. Additionally, recent advancements in technology, particularly in Automatic Speech Recognition (ASR) and Natural Language Processing (NLP), offer new potential for automating the language screening process.

We propose the development of a game that integrates these technological advancements with established developmental milestones to create a screening tool for early childhood language disorders. The proposed game aims to target children aged 3 to 4 years, a critical developmental window for identifying potential disorders (Ward, 1999). By embedding screening parameters into a game environment, we aim to gather comprehensive data on a child's language capabilities in a setting that is both natural and engaging. This approach aims to address the challenges posed by the shortage of SLPs and increase the probability of early screening by extending the reach and efficiency of screening processes from the outset.

In this paper, we describe work in progress exploring related projects at the intersection between speech-language pathology and computer science, identifying screening methodologies that are amenable to automation, and proposing game activities that have the potential to probe the target developmental milestones.

2 Communication Disorders

Communication disorders encompass a range of impairments in the ability to receive, send, process, and comprehend concepts through verbal, nonverbal, and graphic symbol systems (Fogle, 2022). A communication disorder may be evident in the processes of hearing, language, and/or speech. These disorders vary significantly in severity from mild to profound, and can be either developmental or acquired (Cooper, 2018). For the purposes of this paper, the focus will be specifically on language disorders (Owens, 2020).

Language disorders may affect different aspects of language, including phonology (the sound system of a language), morphology (the structure of words), and syntax (the arrangement of words to form sentences). They may also involve the content of language, which pertains to semantics, or the meanings of words and sentences. Furthermore, language disorders can influence the function of

language in communication, known as pragmatics, which involves the social uses of language (ASHA, 1993). Individuals with communication disorders often face difficulties in various aspects of life (McCormack et al., 2009). Studies involving large groups of children with communication disorders have shown that they tend to have lower academic achievement, struggle more with reading, experience increased bullying, have weaker peer relationships, and encounter more psychosocial challenges than their typically developing peers (Lewis et al., 2016).

2.1 Language Disorder Screening

Early identification of language disorders in young children is crucial for timely intervention and support (Ward, 1999). This review evaluates several screening tools that, while not specifically designed exclusively for children aged 3 to 4 years, cover this critical developmental period within their scope. We examine their methodologies, effectiveness, and clinical implications based on recent studies and evaluations.

2.1.1 Northwestern Syntax Screening Test

The Northwestern Syntax Screening Test (NSST), (Lee, 1971), designed for children from 36 to 47 months, evaluates both receptive and expressive language abilities. Modifications have reduced the original test from 20 to 11 items while maintaining 95% of the variance observed in total test scores (Ratusnik et al., 1980). This revised version now requires approximately 10 minutes for administration and provides norms in six-month intervals, enhancing its sensitivity and specificity for this age group. Ratusnik et al. (1980) conducted cross-validation with a sample of 301 children, demonstrating its reliability in maintaining consistent clinical decisions across both the original and shortened versions, thus emphasizing its utility in clinical and educational settings.

2.1.2 Developmental Profile-II

The Parent Language Checklist and The Developmental Profile II (DP-II) (Alpern et al., 1980), particularly its Academic scale, serves as a parent-report tool assessing developmental milestones from birth to 7 years. The scale, when tested on 94 children between 36 and 39 months, revealed significant deficiencies in detecting developmental issues; only 21% of children with identified problems were correctly flagged. However, alternative cutoff scores suggested by Alpern et al. (1980) have shown potential in improving its diagnostic sensitivity. This tool underscores the challenges and importance of accurate parent-report measures and the need for rigorous standardization and validation to ensure reliability.

2.1.3 Minnesota Child Development Inventory

The Minnesota Child Development Inventory (MCDI), (Ireton and Thwing, 1974), offers a comprehensive assessment of various developmental domains, including expressive language and comprehension, for children from 24 to 87 months. It includes a detailed inventory that profiles eight developmental scales and provides norms based on a sample of 796 children. The results categorize development as retarded, borderline, or within normal limits, facilitating early detection of language and other developmental delays. Its extensive age range and detailed developmental scales make the MCDI a valuable tool for early childhood educators and clinicians.

2.1.4 ASHA's Developmental Milestones

American Speech-Language-Hearing Association (ASHA)'s Developmental Milestones[1] provide guidelines on expected communication and feeding skills from birth to 5 years. These milestones are intended to assist parents and professionals in identifying potential delays and initiating discussions for further assessment or referral. It is crucial for raising awareness and guiding early interventions based on observed developmental progress.

3 Use of Technology for Language Development

Various applications have made use of gamification, NLP techniques, or Machine Learning (ML) to assist with communication disorders. Some works focused on creating educational solutions, such as Sztahó et al. (2018), Bogach et al. (2021), and Prasanna and Perera (2019), which all utilized speech processing techniques and automated their evaluation processes without gamification. In contrast, work such as Lyytinen and Louleli (2023) demonstrated gamification without the use of automated evaluation or NLP techniques. Few studies focused on automating early screening for Developmental Language Disorder (DLD). For example, Rvachew et al. (2017) developed a computer-based

[1] http://www.asha.org/public/speech/development/chart/

tool for screening literacy delays but without gamification or NLP/speech processing techniques. On the other hand, the work most closely related to ours is Beccaluva et al. (2024), which introduced MARS, a web-based tool for screening DLD by engaging children in rhythmic babbling exercises to record their vocal productions, which are then analyzed using ML. They evaluated their solution on forty-seven children, 17 diagnosed with DLD and 30 with typical development (TD), collecting additional demographic information (i.e., age, gender, typicality) along with corresponding audio. After preprocessing the data, they trained models using Support Vector Machine, Random Forest, and Logistic Regression, achieving an overall accuracy of 83% in detecting DLD. Specifically, for DLD cases, they achieved 87% precision and 70% recall.

4 Proposed Method: An interactive Game for Early Language Screening

Our proposed approach aims to synthesize these technological advancements, particularly gamification, speech processing, NLP, and ML, into a comprehensive tool for early detection of language disorders. Screening for language disorders is a broad topic, encompassing various sub-categories of screening, such as semantics, morphosyntax, pragmatics, and phonology.

While earlier detection and intervention is effective[2], we focus on the age group between three and four years old, as screen use is not recommended for children younger than 24 months (American Academy of Pediatrics, 2024). Through the game, we aim to collect data that allow us to analyze the child's language performance. To do so, we base our measurement on Yang et al. (2022), which indicates that evaluating a child's language abilities can be done by assessing utterance length and complexity, as well as lexical diversity. The assessment of lexical diversity focus on the total number of different words used by the child [3], and the type-token ratio, which measures the ratio of different word types (types) to the total number of words (tokens) used, providing insight into the child's vocabulary richness and variety. More specifically, as reflected by Winters et al. (2022) and Akmeşe and Kanmaz (2021), we will analyze:

(1) measures of linguistic productivity in narratives: total number of words, utterances, and lexical diversity, **(2) global measures of narrative linguistic complexity:** average and maximum sentence length in words, **(3) measures of syntactic-semantic complexity** (Frizelle et al., 2018)**:** number and proportion of simple and complex sentences, types of complex sentences, and the diversity of adverbial clauses, **(4) maintenance of referential cohesion** (Gagarina and Bohnacker, 2022)**:** problems with nominal or verbal agreement, use of regular and/or irregular inflection, inappropriate use of tense and mood.

4.1 Transferring Requirements to a Game

Botting (2002) reflected that storytelling is one of the best ways to observe and evaluate children's pragmatic skills. Several researchers, including Akmeşe and Kanmaz (2021), Orizaba et al. (2020), and Winters et al. (2022), have analyzed language skills based on storytelling. Given these observations, we propose implementing the measurement requirements above into a game that motivates storytelling, and other side activities.

The proposed game will be level-based, with a focus on avoiding repetitive and dull levels as suggested by Lövdén et al. (2010). Each level will feature a familiar and reassuring character, which has been found effective by Vona et al. (2020). The characters will present challenges to the child (player) that require assistance. For example, in one level, the player helps a character by re-arranging story images scattered by another character. The images might include a bus, a breakfast, and an alarm. The player will sequence the images to show: the alarm rang, the student ate breakfast, and then went to school. Then, the player narrates the story and records their voice, which we process using ASR to determine the content.

In another level, a curious character asks questions such as "What is this?", "What is he/she doing?", and "Which is bigger?". The player will be tasked with answering the questions. These tasks and questions are inspired by the Speech and Language Milestone Chart [4] by LD OnLine (2024).

As the game progresses, the player will see their progress through the main menu, reflecting their performance on each level and overall progression.

[2] Ward (1999) followed up with 122 children aged between 8 to 21 months diagnosed with early language delay, and concluded that early intervention is effective at preventing language delay at 3 years old.

[3] Word categories: noun, verb, adjective, adverb, preposition, pronoun, determiner, conjunction, and interjection.

[4] https://www.ldonline.org/ld-topics/speech-language/speech-and-language-milestone-chart

To motivate the player, we will introduce stars as collectibles and other incentives. However, following American Academy of Pediatrics (2024)'s recommendation that children aged 2 to 5 only use smart devices for 1 hour per day, we will ensure that sessions do not exceed this time limit.

4.2 Evaluation Methodology

We will start by collecting data needed by SLPs to analyze the child's case. This includes a short questionnaire at the start of the game about the child's age, languages spoken, and other questions important for understanding the environmental factors that can influence language development. This information will be part of our inputs. After the child completes the game, we will retrieve and process the data using ASR and NLP techniques like Part-of-Speech Tagging and syntactic parsing, to extract additional information and gain further insight into the children's capabilities. This data will be used to measure the four key points mentioned in Section 4. By combining parent-provided data with gameplay data, we will collaborate with SLPs to identify potential signs of language disorder and label the data. In **stage one**, SLPs will perform one-on-one screening using traditional methods to create gold labels. In **stage two**, independent SLPs will assess the children using only the data collected through the game, and their performance will be compared with the gold labels to validate the game's methodology. Finally, we will train ML models to predict language delays using the collected data, focusing on high recall to improve screening coverage.

4.3 Challenges & Future Work

The journey from concept to implementation is filled with technical and operational challenges, from developing engaging and educational game content to ensuring the accuracy and reliability of the AI-driven screening tools. Effective collaboration between game developers, speech therapists, technology experts, and educational institutions, will be crucial in overcoming these difficulties.

Technical Challenges: The accuracy of the AI-driven screening model depends heavily on the quantity and quality of the data collected. Initially, gathering a sufficiently large and diverse dataset through field trials will be costly and time-consuming. The data must be carefully labeled and validated to ensure that models learn from accurate examples. In addition to screening-related data, larger data sets of children's speech will be needed to develop accurate ASR models if speech-related activities are deemed suitable for the game design. Children's speech is challenging for automatic processing due to its natural variability and shortage of data (Gerosa et al., 2009). For bilingual children, additional complexity is expected due to code-switching.

Operational Challenges: To validate the effectiveness of various aspects of the proposed game, several field trials will be needed. A sample with sufficient number of children with various developmental conditions needs to be collected for the first stage of thorough validation. This may require the administration of a large number of manual screenings to identify a sufficient number of children with language disorder. Collaboration with pre-schools and parents will be essential at this stage. Second, to provide norms for benchmarking the game's outcomes, a large number of participants from different regions and demographic segments are needed. Additional difficulties will be encountered in bilingual communities, for which both languages need to be assessed. Last but not least, collecting data involving children requires a well-defined and thorough ethical and legal framework to ensure children's protection against any potential misuse of the data.

5 Conclusion

The proposed game-based screening tool utilizes established developmental milestones to guide its design. By embedding these milestones into a game's mechanics, we ensure that each interaction within the game serves a dual purpose: to engage the child and to evaluate their language development. The use of NLP and ML methods for analyzing the data collected from these interactions aims to provide a preliminary screening that can help identify children who may require further evaluation by a specialist. This ensures that no child in need of further screening is overlooked, while maximizing the utilization of SLP time for the most likely cases of language delay. Early detection and intervention in language disorders are critical for the educational and social development of children. By providing a more accessible and appealing method for screening, we hope to increase the number of children who receive timely intervention, thereby improving long-term outcomes in their learning and communication abilities.

References

Pelin Pişlav Akmeşe and Serap Kanmaz. 2021. Narrative to investigate language skills of preschool children. *International Electronic Journal of Elementary Education*, 14(1):9–22.

Gerald Alpern, Thomas Boll, and Marsha Shearer. 1980. Developmental profile ii. *J Read*, 18:287–91.

American Academy of Pediatrics. 2024. https://www.aap.org/. Accessed: 2024-05-29.

ASHA. 1993. Definitions of communication disorders and variations.

Eleonora Aida Beccaluva, Fabio Catania, Fabrizio Arosio, and Franca Garzotto. 2024. Predicting developmental language disorders using artificial intelligence and a speech data analysis tool. *Human–Computer Interaction*, 39(1-2):8–42.

Natalia Bogach, Elena Boitsova, Sergey Chernonog, Anton Lamtev, Maria Lesnichaya, Iurii Lezhenin, Andrey Novopashenny, Roman Svechnikov, Daria Tsikach, Konstantin Vasiliev, et al. 2021. Speech processing for language learning: A practical approach to computer-assisted pronunciation teaching. *Electronics*, 10(3):235.

Nicola Botting. 2002. Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child language teaching and therapy*, 18(1):1–21.

Rachel Cooper. 2018. *Diagnosing the diagnostic and statistical manual of mental disorders*. Routledge.

Mårten Eriksson, Monica Westerlund, and Carmela Miniscalco. 2010. Problems and limitations in studies on screening for language delay. *Research in Developmental Disabilities*, 31(5):943–950.

Paul T Fogle. 2022. *Essentials of communication sciences & disorders*. Jones & Bartlett Learning.

Pauline Frizelle, Paul A Thompson, David McDonald, and Dorothy VM Bishop. 2018. Growth in syntactic complexity between four years and adulthood: Evidence from a narrative task. *Journal of Child Language*, 45(5):1174–1197.

Natalia Gagarina and Ute Bohnacker. 2022. A new perspective on referentiality in elicited narratives: Introduction to the special issue. *First Language*, 42(2):171–190.

Matteo Gerosa, Diego Giuliani, Shrikanth Narayanan, and Alexandros Potamianos. 2009. A review of ASR technologies for children's speech. In *Proceedings of the 2nd Workshop on Child, Computer and Interaction*, pages 1–8.

Harry Ireton and Edward Thwing. 1974. *Minnesota child development inventory*. Behavior Science Systems, Incorporated.

LD OnLine. 2024. Ld online: The educator's guide to learning disabilities and ADHD. https://www.ldonline.org/. Accessed: 2024-05-29.

Laura L Lee. 1971. Northwestern Syntax Screening Test (NSST).

Barbara A Lewis, Emily Patton, Lisa Freebairn, Jessica Tag, Sudha K Iyengar, Catherine M Stein, and H Gerry Taylor. 2016. Psychosocial co-morbidities in adolescents and adults with histories of communication disorders. *Journal of Communication Disorders*, 61:60–70.

Martin Lövdén, Lars Bäckman, Ulman Lindenberger, Sabine Schaefer, and Florian Schmiedek. 2010. A theoretical framework for the study of adult cognitive plasticity. *Psychological bulletin*, 136(4):659.

Heikki Lyytinen and Natalia Louleli. 2023. In search of finalizing and validating digital learning tools supporting all in acquiring full literacy. *Frontiers in Psychology*, 14:1142559.

Jane McCormack, Sharynne McLeod, Lindy McAllister, and Linda J Harrison. 2009. A systematic review of the association between childhood speech impairment and participation across the lifespan. *International Journal of Speech-Language Pathology*, 11(2):155–170.

Lorena Orizaba, Brenda K Gorman, Christine E Fiestas, Gary E Bingham, and Nicole Patton Terry. 2020. Examination of narrative language at microstructural and macrostructural levels in Spanish-speaking preschoolers. *Language, Speech, and Hearing Services in Schools*, 51(2):428–440.

Robert E. Owens. 2020. *Language Development: An Introduction*, 10th edition. Pearson.

V Prasanna and Indika Perera. 2019. Speakup-a mobile application to train and overcome stuttering. In *2019 19th International Conference on Advances in ICT for Emerging Regions (ICTer)*, volume 250, pages 1–8. IEEE.

David L Ratusnik, Thomas M Klee, and Carol Melnick Ratusnik. 1980. Northwestern syntax screening test: a short form. *Journal of Speech and Hearing Disorders*, 45(2):200–208.

Susan Rvachew, Phaedra Royle, Laura M Gonnerman, Brigitte Stanke, Alexandra Marquis, and Alexandre Herbay. 2017. Development of a tool to screen risk of literacy delays in French-speaking children: Phophlo. *Canadian journal of speech language pathology and audiology= Revue canadienne d'orthophonie et d'audiologie*, 41(3):321–340.

Katie Squires. 2013. Addressing the shortage of speech-language pathologists in school settings. *Journal of the American Academy of Special Education Professionals*, 131:137.

Trisha Sunderajan and Sujata V Kanhere. 2019. Speech and language delay in children: Prevalence and risk factors. *Journal of family medicine and primary care*, 8(5):1642–1646.

David Sztahó, Gábor Kiss, and Klára Vicsi. 2018. Computer based speech prosody teaching system. *Computer Speech & Language*, 50:126–140.

Francesco Vona, Emanuele Torelli, Eleonora Beccaluva, and Franca Garzotto. 2020. Exploring the potential of speech-based virtual assistants in mixed reality applications for people with cognitive disabilities. In *Proceedings of the international conference on advanced visual interfaces*, pages 1–9.

Sally Ward. 1999. An investigation into the effectiveness of an early intervention method for delayed language development in young children. *International Journal of Language & Communication Disorders*, 34(3):243–264.

Katherine L Winters, Javier Jasso, James E Pustejovsky, and Courtney T Byrd. 2022. Investigating narrative performance in children with developmental language disorder: A systematic review and meta-analysis. *Journal of Speech, Language, and Hearing Research*, 65(10):3908–3929.

Ji Seung Yang, Carly Rosvold, and Nan Bernstein Ratner. 2022. Measurement of lexical diversity in children's spoken language: Computational and conceptual considerations. *Frontiers in psychology*, 13:905789.