Revolutionizing Early Literacy: A Cognitively-Inspired, Dual-Process Architecture for Children's Speech Recognition

John McDonagh, Yaroslav Nedashkovskyi, Nataliya Tetruyeva
eKidz.eu, Steinstrasse 3, 81667 Munich, Germany
john.mcdonagh@ekidz.eu, yaroslav.nedashkovskyi@ekidz.eu, nataliya.tetruyeva@ekidz.eu

Executive Summary

Traditional methods for assessing children's reading skills are failing our schools. They are time-consuming for educators, providing infrequent and subjective feedback that hinders student development. This paper introduces the eKidz AI Voice Engine: a proven, AI-powered diagnostic solution that is revolutionizing literacy assessment by moving beyond conventional ASR.

Our technology is built on a **cognitively-inspired, dual-process architecture** that functionally mirrors how human cognition works. Deployed with over 250,000 students across Europe, the US, and Latin America, our system automates the analysis of oral reading fluency. By integrating an AI that combines a big-picture, intuitive engine for contextual flow with a laser-focused, analytical engine for phonemic precision, we provide objective and immediate feedback to students, teachers, and parents that is both nuanced and diagnostically rich.

The superior performance of our foundational sub-models (our phoneme-based ASR and noise & voice detection algorithms) is demonstrated by their achieving **10-15% greater accuracy** against manual transcriptions in US-based validation studies. This level of underlying precision is the foundation that will allow the interplay of our two cognitive engines to deliver even more promising results in the future. This provides the personalized and hyper-detailed diagnostics necessary for truly adaptive and effective literacy instruction, saving educators invaluable time and empowering students on their learning journey.

Table of Content

1. The Challenge in Modern Literacy Education

The ability to read fluently is a critical prerequisite for academic success and social participation. To foster this skill, educators must continuously assess each student's progress to provide targeted support. This formative evaluation is essential for selecting appropriate texts and teaching methods that build upon a child's existing abilities.

However, the reality in most elementary school classrooms presents significant obstacles:

Time Constraints: One-on-one reading assessments are highly time-consuming, making it impossible for teachers to conduct them regularly for every student.

Resource Scarcity: Many educators lack the specialized training and standardized instruments needed for deep, systematic diagnosis of reading issues. While measuring reading speed is straightforward, analyzing the root causes of errors (e.g., decoding difficulties, poor phrasing) during a busy school day is impractical.

Delayed Feedback Loop: Due to these constraints, students receive feedback on their reading development too infrequently. Small but hard-won improvements go unnoticed, which can be deeply demotivating, especially for struggling readers who fail to experience a sense of competence and self-efficacy.

Involvement of Non-Professionals: Parents, guardians, and reading mentors are crucial partners in a child's literacy journey but typically lack the professional knowledge to accurately diagnose reading levels and provide targeted guidance.

These challenges contribute to a persistent gap where many students, particularly those from disadvantaged backgrounds, fall behind early and struggle to catch up.

A Case Study in Assessment Challenges: Oral Reading Fluency

A prime example of these challenges can be seen in the assessment of **Oral Reading Fluency (ORF).** Oral reading (the ability to read connected text quickly, accurately, and with appropriate expression) has 30 years of evidence to back up its use as one of the most reliable and efficient indicators of student reading comprehension. Because of its strong evidence base, repeatability and brevity, ORF tests are used for universal screening for early intervention across grades 1 through 8 in America. They are typically conducted three times per school year, often even more frequently for younger students, to monitor reading progress.

Until very recently, ORF assessments had to be carried out by a human examiner, who sat down with each student with a timer and pencil in hand. The examiner listened as the student read a grade-level passage aloud for one minute and noted down key data points. The resulting scores were then compared to national fluency norms to determine whether the student was on target. This manual process is how all the most widely used measures of ORF are currently carried out.

Though efficient compared to many other assessments, there is still a non-trivial time cost involved in administering these tests. To put it into perspective, there are roughly 35.5 million public K-8 students in the United States, and an ORF takes about two minutes per student. Conducted three times a year, the total time spent on administering ORFs adds up to **213,000,000 minutes, or 3,550,000 hours.** This does not include the additional time spent comparing and reporting each student's performance against national norms.

Teachers are required to conduct numerous student assessments due to government and district policies. This creates a significant conflict, as educators are caught between the demands of testing and the core job of teaching. They face pressure from both parents, who want more instructional time, and policies that require more assessment.

2. Our Solution: The eKidz AI Voice Engine

eKidz has successfully addressed these shortcomings through the development and deployment of our automated, app-based reading analysis platform. Our system leverages advanced artificial intelligence, specifically our proprietary AI technology and strategy, to transform the way reading skills are measured and nurtured.

Validated with 250,000+ students across diverse educational environments in Europe, Latin America and selected US schools, our solution has demonstrated consistent, robust performance in real-world classroom settings.

eKidz Technology: A Cognitively-Inspired Dual-Process ASR for Children

Standard Automatic Speech Recognition (ASR) systems, while improving for adult speech, consistently falter when faced with the unique challenges of children's voices. The higher vocal pitch, greater acoustic variability, and developmental speech patterns of young learners create a level of ambiguity that conventional, single-mode ASR architectures cannot resolve with the precision required for meaningful educational assessment.

eKidz has solved this challenge by pioneering a specialized ASR framework inspired by **Dual-Process Theory** (Evans & Stanovich 2013; Gronchi et al. 2024; Gawronski, Luke, Creighton 2024), a foundational concept in cognitive science that is now at the forefront of advanced AI development. This theory posits that complex cognition arises from the interplay of two distinct systems: a widely focused and highly contextual **System 1** and a sharply focused, non-contextual **System 2**. Our architecture functionally mirrors this powerful model to achieve a high level of diagnostic depth and accuracy.

Our breakthrough lies in the moderated dialogue between two core components:

- **The System 1 Engine: Intuitive Contextual Processing.** This engine functions like a fast, intuitive processor, attuned to the holistic context, relationships, and prosodic flow of language. Our System 1 engine comprises models which use context and language models to form an understanding of the utterance. This is achieved through technologies like context-aware ASR and multi-scale prosody analysis, allowing it to appreciate and accommodate a child's unique accent and analyze high-timeframe prosody, providing a «big picture» view of reading fluency.
- **The System 2 Engine: Deliberative Phonemic Analysis.** This engine functions like a methodical, analytical processor, laser-focused on details. It comprises models which ignore context and grasp onto each detail. It includes a phoneme stream model, which faithfully deconstructs the speech signal to its fundamental sound units (phonemes). Unlike conventional ASR that targets word accuracy, our validated System 2 engine meticulously analyzes each sound in time. This deliberative, phoneme-level process enables the hyper-detailed, rule-based analysis required for true diagnostic insight.

The two engines are moderated through an evolving mutual-feedback "Metacognition" layer. Preprocessing "buffer" models and an output "Diagnostic" layer for final articulation of the analysis round out our architecture.

The real power of the eKidz platform lies in the interplay between the two engines. This regulatory layer allows for a profound and faithful analysis of a child's speech by enabling the systems to cross-validate and «sanity-check» each other's outputs. For instance, the System 2 engine can evidence granular phonemic details to flag a potential ASR «hallucination» from the System 1 engine. Conversely the System 1 engine can provide the contextual evidence needed to interpret corrupt or missing details from System 2. This dynamic collaboration, analogous to deliberation models

in AI research, enables much deeper and more meaningful diagnosis by identifying the specific nature of reading errors:

- **Decoding Accuracy:** Pinpointing errors in grapheme-phoneme correspondence through precise, rule-based analysis.
- **Accent Analysis and Tolerance:** The System 1 engine's contextual awareness of accent informs the System 2 engine's phonemic targets, enabling accurate assessment that is tolerant of diverse speech patterns.
- **Reading Fluency Statistics:** Automatically calculating correctly read words per minute while simultaneously tracking specific, analytically verified error types including substitutions, omissions, inclusions, and repetitions.
- **Hesitations and Self-Corrections:** Detecting moments of difficulty that indicate a lack of automaticity.
- **Handling of Unknown Vocabulary:** The metacognitive dialogue allows the system to identify neologisms and esoteric filler-words and internalize them, where a standard ASR would force an incorrect match.

This cognitively-inspired, dual-process architecture provides hyper-detailed insight that enables accurate and objective feedback. It improves teachers' ability to track progress and identify learning opportunities, some of which may otherwise go unnoticed. Voice-enabled reading assessments become highly scalable and can be performed more frequently than conventional assessments. Students can read text independently, with teachers reviewing the system-generated reports in their own time.

Our phoneme-based model and noise detection model have demonstrated superior performance, achieving 10-15% closer accuracy to manually transcribed recordings against datasets of American children's voices across diverse accents and reading assessment scenarios in validation studies with one of a leading U.S. curriculum and assessment company.

eKidz' Core Technology Architecture

Advanced Phoneme-Based ASR Engine: Our proprietary ASR framework, built on enhanced Wav2Vec 2.0 architecture, has been specifically fine-tuned for children's speech patterns. Our architecture is structured as a multi-layer system where each layer performs a specialized function, mirroring the principles of Dual-Process Theory.

- **Acoustic Pre-processing Layer:** This initial layer handles advanced noise and voice detection from the raw recording, ensuring a clean audio signal is passed to the cognitive engines for analysis.

- **System 2 Engine (Deliberative Phonemic Analysis):** This layer performs the precision-focused, analytical tasks. It is responsible for the detailed acoustic analysis, the core phoneme recognition, and the analysis of low-timeframe prosody (the fine details of speech rhythm and intonation).
- **System 1 Engine (Intuitive Contextual Processing):** This layer processes the holistic, «big picture» aspects of the speech. It uses fine-tuned ASR models to understand the broader context, language flow, and high-timeframe prosody (the overall melodic contour of the reading).
- **AI Metacognition Layer:** This is the moderating layer where the System 1 and System 2 engines mutually inform and correct each other. It allows the system to resolve ambiguities and «sanity-check» outputs, ensuring a more robust and faithful analysis.
- **Diagnostic Output Layer:** This final layer synthesizes the processed data from the cognitive engines. It performs conclusive lexical and semantic analysis to generate user-facing reports and potential feedback stimuli.

Examples of features enabled by this architecture include:

- Verifying the target utterance within the audio file.
- Aligning the speech contained in the audio file to the target text and returning a score.
- Generating highly accurate transcriptions of read-aloud texts by children using our specialized phoneme-based System 2 engine.
- Calculating real-time error classifications and fluency metrics by integrating the outputs of both the System 1 and System 2 engines.

Deployment-Ready Infrastructure:

- Cloud-native architecture supporting unlimited concurrent assessments
- Cross-platform compatibility (iOS, Android, web browsers)
- GDPR compliant data processing and storage

How the eKidz AI Voice Engine System Works

When a child reads an age-appropriate text aloud into any mobile device or computer, our eKidz AI Voice Engine:

1. **Captures and Processes:** Records speech with advanced noise detection
2. **Analyzes in Real-Time:** Our phoneme-based ASR engine transcribes and analyzes speech patterns with 94% accuracy on recording of any length, this rate is comparable to that of human-to-human scoring and that exceeds market standards.

3. **Classifies Errors:** Automatically identifies and categorizes reading behaviors including substitutions, omissions, inclusions, and repetitions. Our error classification accuracy is currently at 92% for the above metrics.
4. **Calculates Metrics:** Generates comprehensive fluency statistics including words per minute, decoding accuracy, and prosodic features
5. **Delivers Insights:** Provides instant, actionable feedback through intuitive dashboards for students, teachers, and parents

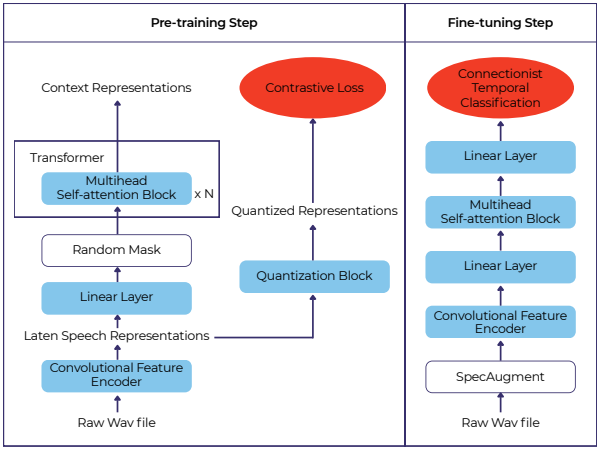
Fine-Tuning for Children's Voices: The eKidz Advantage

The eKidz AI Voice Engine is an advanced speech processing technology designed to support language and literacy development in young children. While built upon state-of-the-art deep learning models like Wav2Vec 2.0, a powerful foundation is not enough for the unique challenges of educational technology.

The Necessity of Fine-Tuning for Children

Off-the-shelf, pre-trained models like Wav2Vec 2.0 are typically trained on vast amounts of adult speech. This creates a significant performance gap when applied to children. Children's voices have different acoustic properties, and their speech is filled with unique patterns, including mispronunciations, disfluencies, and the accents of non-native speakers. A generic model simply cannot interpret this variability with the precision required for meaningful educational feedback. The result of finetuning is a more responsive and accurate voice engine that is better suited to provide meaningful feedback in real-world educational settings. The following diagram details the pre-training and fine-tuning methodology.

Chart 1. The pre-training and fine-tuning methodology.

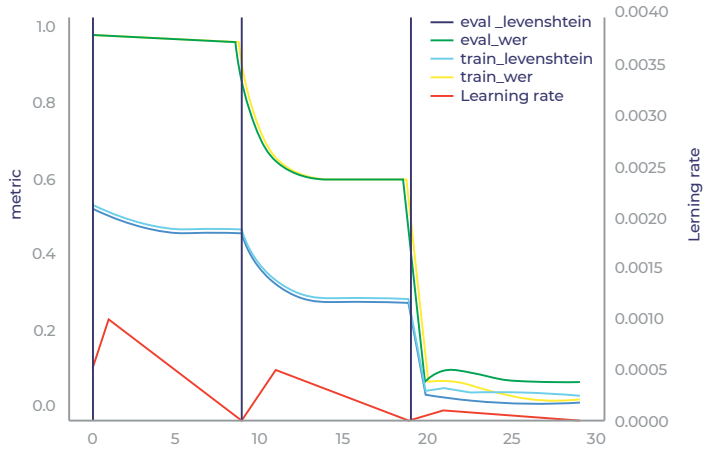


eKidz' Success Through Specialized Fine-Tuning

To solve this, we engineered a critical fine-tuning process. We retrained our base model using our own curated, child-specific audio datasets. This intensive process teaches the model the specific nuances of young speakers' voices.

The results of this process are clear and measurable. As the model undergoes fine-tuning, its performance gains are steady and significant with each training epoch. While early epochs show high variability, especially with non-native pronunciations, the model quickly learns to capture key linguistic features. This dramatically reduces substitution and deletion errors, aligning the AI's transcription much more closely with human benchmarks. The model's learning process eventually converges, resulting in a stable, highly accurate, and responsive voice engine specifically optimized for real-world educational settings.

Chart 2. Model's performances over training epochs



Validated Educational Impact of eKidz' AI Voice Engine

A comprehensive, 6-week scientific evaluation validates the significant educational impact of the eKidz app, which integrates our proprietary AI Voice Engine and adaptive feedback mechanisms for literacy intervention.

The study found that our technology achieves results comparable to established, evidence-based paper programs, demonstrating its power to significantly enhance the effectiveness of reading interventions. The core driver of this success is the automated feedback mechanism, powered by real-time ASR assessment, which proved highly effective in accelerating reading development.

While reading fluency improved across all study groups, the cohort receiving instant feedback from the eKidz AI assessment demonstrated the most substantial progress, achieving effect sizes ranging from d = .29 to .49. These clear, measurable improvements provide scientific validation that our ASR assessment, when applied for formative evaluation, delivers tangible educational outcomes that meet and often exceed traditional intervention methods.

In essence, our digital solution successfully automates the principles of established analog reading assessments, creating an efficient and scalable platform that matches—and frequently surpasses—the diagnostic precision of a professional educator.

3. State of Science and Technology

Classification of Reading Competence

Reading forms the basis for independent knowledge acquisition and personal development, making it a central component of school education. The goal of promoting reading competence is to enable children to participate autonomously in social and cultural life (KMK, 2005). Reading competence is a multi-faceted construct that encompasses different levels and facets (Lenhard 2019, Schilcher & Wild 2018; Rosebrock & Nix 2008). Basal reading processes initially include the decoding of words by establishing grapheme-phoneme correspondences and forming local semantic connections. In the further process, the formation of global coherence across multiple sections is an important characteristic of a competent reader. At the beginning of the acquisition process, increasing the degree of reading accuracy and the progressive increase of reading speed are the most important tasks to be mastered. Only when a threshold is exceeded in these areas (e.g., 95% decoding accuracy and 100 words per minute (WpM) as a benchmark for reading speed) does the reading process become automated, freeing up cognitive resources for processing the content. Sufficient decoding accuracy (Rasinski, 2003) and reading speed form the foundation upon which reading comprehension can develop (Wild & Schilcher 2019, Lenhard 2019).

Currently, this is assessed by teachers using analog tests such as read-aloud protocols (e.g., Johns & Berglund, 2006) or the Oral Reading Fluency Scale (Pinnell, 1995). Standardized tests are available, for example, in the form of the Salzburger Lesescreening 2-9 (Wimmer & Mayringer, 2016), ELFE II (Lenhard, Lenhard & Schneider, 2017), or the Lesegeschwindigkeits- und -verständnistest 6-12 (Schneider, Schlagmüller & Ennemoser, 2007). Because read-aloud protocols require the teacher to record parameters

individually for each child on an observation sheet, they demand extensive time resources. With standardized tests, the evaluation requires additional work time at home. The project initially focuses on reading fluency, which is particularly laborious to assess. Its assessment—in contrast to measuring reading comprehension through multiple-choice tests—can only be done in one-on-one situations. Prosodic reading is also an indicator of basal reading skills, as it allows for inferences about text comprehension (Reiss 2019; Nagler, Lindberg & Hasselhorn 2018). The development of reading competence thus follows specific developmental steps, which can differ due to genetic predispositions, cognitive abilities, and especially external factors such as class composition or socioeconomic and cultural background (Nagler, Lindberg & Hasselhorn 2018).

To optimally support the reading skills of all students in their individual development process, teachers must be able to design reading instruction adaptively and oriented to the students' performance level. This, however, requires that the reading competence of each child is determined continuously, preferably using standardized procedures (Wagner 2016; Martschinke 2015, Lenhard & Schneider 2009). Such formative evaluation is a prerequisite for the teacher to derive individual learning goals (Lenhard & Schneider 2009). The feedback of performance to the children, teachers, and guardians is an important factor that promotes cooperative learning and creates positive conditions for cooperative work (Wagner 2016). The continuous monitoring of individual reading progress is a key prerequisite for the optimal, adaptive support of children. If this is not possible, learners fall short of the potential they could achieve with consistent support (Gebauer 2019; cf. Philipp 2011). The proportion of weak readers remains high, and the unfavorable correlations between social background and reading performance are stable (Bremerich-Vos et al. 2017; McElvany et al. 2017; Hippmann et al. 2019). However, even if weak readers represent the largest problem group, strong and average readers must also be adequately classified and supported at their individual performance level.

In the current research discourse, the following desiderata emerge: For effective and adaptive reading instruction, a sufficient diagnostic competence of teachers is indispensable, but it is generally not highly developed. The responsibility for diagnosis is also perceived differently. Instruments for a differentiated and precise diagnostic assessment in heterogeneous environments, where established instruments show floor or ceiling effects, are lacking.

Computer-based Assessment of Reading Performance

Using mobile applications on devices like smartphones, tablets, or PCs, the target groups could be addressed individually and supported in the process of formative assessment of their reading competence. Mobile applications offer many options and could be used to enable an objective, standardized assessment or classification and presentation of results over the course of development. The novelty of the proposed project lies in utilizing existing hardware and previous software from reading support programs and supplementing them with highly innovative aspects of automatic speech recognition and automatic reading analysis, which have not been integrated in any way so far and do not exist for the intended application. The automatic analyses include AI-based algorithms that capture and automatically classify acoustic speech events and specific features. In conjunction with linguistic prior knowledge and statistical methods, parameters for reading accuracy and fluency can be evaluated individually and automatically. This project aims to address and implement these desiderata. Through motivationally appealing design and approaches from gamification, the processes are intended to run in such a way that they do not lead to performance anxiety and demotivation, but rather that individual performance gains lead to an increased sense of self-efficacy.

Automatic Speech Recognition and Analysis for Assessing Reading Competence in Children

The measurement methods and result presentations to be developed should be implemented for children, teachers, and guardians in such a way that they receive feedback on the performance level and the development of individual competence facets. Furthermore, communication between the target groups is to be promoted. Technologies from the field of artificial intelligence, such as automatic speech recognition (ASR) and evaluation algorithms based on acoustic features and machine learning methods, are to be embedded as an automatic, computer-based analysis tool to enable the assessment of reading fluency and accuracy. This combination of speech recognition and the additional content-based evaluation via algorithms as a knowledge-based system is completely new and has not yet been developed in the form of an application beyond Europe. eKidz can apply and implement its expertise in ASR and algorithms of machine learning and AI for pattern recognition and analysis here.

Although the performance of ASR systems has increased in recent years in the area of adult speech (Li 2019), the automatic recognition of children's speech has so far been scarcely addressed, even though it is a fundamental future step in medical or pedagogical applications. Shivakumar and

Georgiou (2020) make it clear that ASR for children still achieves lower accuracy in performance compared to ASR performance for adult speech. The acoustic prerequisites, in particular, are a challenge here, as previous feature extractions in the ASR front-end are not spectrally aligned for the acoustics of children, and the acoustic models, based on DNNs, are not sufficiently trained on them. Furthermore, the inter- and intra-speaker variance in children is particularly high both acoustically and linguistically and has been difficult to model in ASR systems so far.

In the implementation of the proposed project, it can be exploited that grammatical structures and vocabulary are less complex in children and that reading performance (as opposed to spontaneous speech) is being tested, which reduces heterogeneity (Yeung & Alwan 2018; Liao et al. 2015). The ASR of eKidz is an evolving system. It uses special feature algorithms in the front-end. The modelling of the acoustic signal to the graphemic output is implemented via DNN (Wav2Vec2) Models for adult speech already exist here. Nevertheless, it is necessary to adapt and train the acoustic models for the target group. This means that, on the one hand, the prosody of read speech differs from that of spoken speech, and on the other hand, the spectral features of children's speech must be adequately captured, and the algorithms adapted and the models retrained accordingly. Precise methods and approaches for such new training are to be developed during the project period. However, it is promising to adopt existing algorithms for training an acoustic model and to link them with new speech data and annotations, thus enabling integration into the existing system. Aspects such as dialect or accent in children whose L1 (= first language) is not their schooling language must also be considered during data collection and further development and pose a challenge. Further limitations in speech development, which can manifest primarily in phonetic and phonological errors (e.g., sigmatism), can be assumed in the target group. To counteract this, a heterogeneous audio material must be ensured for adaptation and development. Establishing semantic references within the speech material and training the models accordingly can also reduce the susceptibility to errors here. Booth et al. (2020) show that even a medium-sized amount of children's speech data is sufficient to optimize ASR performance. In their dataset, about 10 hours of audio material of individual phrases were used. Similar approaches are to be investigated in the project.

New developments in the field of ASR and statistical methods also enable semantic-lexical and phonemic analyses. Furthermore, word recognition and phoneme rates can be determined, and acoustic features of speech can be captured. Prosodic features can be captured from the speech signal via acoustic measures of pitch or loudness, providing information

on intonation and stress. These analyses can, among other things, provide clues about aspects of reading accuracy and fluency and thus cover the core areas of reading competence. For the automated assessment of reading competence, methods from statistics and machine learning will be used. These convert the natural language of children into linguistic and acoustic variables that lead to metric parameters. These, in turn, are examined using quantitative methods of computational linguistics. Here, classification systems are trained. Using tools such as TensorFlow and ScikitLearn, suitable parameters and methods are tested and trained, which then allow for automatic classification. Support Vector Machines (SVM), Naive Bayes, and Gaussian Processes are methods from the field of supervised learning and classification to be tested here (Solan, Z., Horn, D., Ruppin, E., & Edelman, S. (2005).; Daelemans, W., & Hoste, V. (2002)). In the first step, evaluations are carried out manually by experts, and so-called annotations are created. Subsequently, the manual evaluations are inserted into matrices of linguistic variables, and using shallow AI, an algorithm is then iteratively developed, and such classifiers are implemented that achieve an evaluation quality within the inter-rater reliability (Johnson, D. O., Kang, O., & Ghanem, R. (2016); Goh, Y. C., Cai, X. Q., Theseira, W., Ko, G., & Khor, K. A. (2020)).

Audio Data Quality in Digital Applications

User audio data processing and data quality have a significant impact on the sustainability of ASR performance. In the field of children's speech recognition, detection methods must be designed to suppress background voices and noise. Children's applications must use a lossy audio codec, such as AAC (Advanced Audio Coding), to save bandwidth and offer a technology suitable for everyday use. The possibilities for audio signal enhancement and their adaptation to the target analysis systems are only possible with a transparent target system, so conventional solutions from the adult speech market are not applicable. For optimal results, the device on which the data is recorded must also be optimally positioned. For this, users are sensitized with appropriate interface hints, and the correctness is confirmed with a test on the end device.

4. Benefits for Key Stakeholders

This integrated tool is designed to create a collaborative and effective ecosystem for literacy development.

For Teachers

- **Time Savings:** Frees up valuable classroom time previously spent on manual, one-on-one testing.

- **Enhanced Diagnostics:** Provides detailed, objective data on each child's specific challenges, such as incorrect phrasing, a weak sight-word vocabulary, or decoding difficulties.
- **Personalized Instruction:** Enables the creation of individualized training plans and the formation of targeted learning groups (e.g., tandem reading pairs, homogenous skill groups) based on reliable data.

For Students

- **Immediate Feedback:** Regular, automated feedback makes even small steps of progress visible, fostering intrinsic motivation and a positive sense of self-efficacy.
- **Personalized Goals:** The child understands their specific strengths and areas for improvement, encouraging them to take ownership of their learning.
- **Equitable Access:** Provides crucial diagnostic support for children who may lack sufficient academic guidance at home, promoting easier access to education.

For Parents, Guardians, and Tutors

- **Clarity and Empowerment:** Offers a clear, understandable assessment of a child's reading level without requiring professional expertise.
- **Effective Home Support:** Provides concrete recommendations for practice, ensuring that support at home is aligned with the child's needs and avoids material that is either too easy or too difficult.
- **Improved Communication:** Creates a common, data-driven basis for productive conversations between parents and teachers.

5. Conclusion and Future Developments

The transition from manual to automated reading assessment represents a paradigm shift in early literacy education. Our cognitively inspired, dual-process ASR system is a foundational element of our strategy for speech analysis. This project will create a tool that is more than just a technological novelty. It is a means to provide objective, frequent, and diagnostically rich feedback that empowers teachers, motivates students, and engages parents more effectively. This will help ensure that no child is left behind, fostering the strong literacy skills that are essential for a future of learning, participation, and opportunity.

The eKidz AI Voice Engine holds an important position in the current scientific discourse on reading promotion and the development of adaptive learning technologies by bridging empirical classroom research, cognitive psychology models, and modern technology. Reading fluency has long been considered a central goal in learning to read, as it forms the basis for understanding written texts. By developing and standardizing an automatically determined reading score based on children's speech recordings, we have created an objective, validated measure that assesses reading fluency with precision, thus allowing for diagnostic and pedagogical conclusions.

The eKidz technology allows curriculum providers and EdTech companies to integrate formative assessment and benefit from a suite of powerful capabilities. By leveraging our precise assessment technology, which is uniquely tuned to the characteristics of children's speech, partners can offer highly effective, individually adaptive feedback. This integration supports the development of self-regulated learning and demonstrably increases student motivation, grounded in established pedagogical principles like self-determination theory. Furthermore, our engine transforms complex reading fluency data into a normalized, easy-to-understand score, optimizing the usability of the technology for both teachers and learners and creating a seamless interface between advanced AI and practical classroom application.

Based on our findings in the most recent research projects and technological foundation, diverse future perspectives emerge. First, our automated reading score can be integrated into adaptive learning systems that offer individualized support paths to continuously monitor and optimize learning progress. Such systems have the potential to improve reading promotion in an evidence-based manner, especially in elementary schools, and to relieve teachers.

Furthermore, our technology supports the development of a comprehensive, digitally supported reading diagnostics system. In the long term, the eKidz AI voice engine could play an important role in inclusive teaching scenarios by identifying children with reading difficulties at an early stage and supporting them with tailored interventions. This corresponds to current educational policy goals for promoting equal opportunities. The integration of feedback based on cognitive and motivational theories can help to strengthen reading motivation in the long term. This dual-process architecture will facilitate the future development of sophisticated features such as voice isolation, simultaneous tracking and analysis of multiple voices, further semantic analysis, and real-time conversational feedback. Similarly, the rich underpinning database and data model will allow for

evermore profound lexical understanding, as well as additions of more languages and accents.

Finally, the combination of linguistic analysis and machine learning opens up new research fields, such as improving the automatic detection of reading errors and prosodic features. Scientific publications and exchange with the professional community can help to further validate the developed methods and adapt them in different linguistic and cultural contexts.

In summary, the eKidz AI Voice Engine contributes both to the scientific advancement of reading promotion and to the practical implementation of modern diagnostic and support methods, thus opening up new paths for the future of language and reading promotion in the digital age

References

Booth, E., Carns, J., Kennington, C., & Rafla, N. (2020). Evaluating and Improving Child-Directed Automatic Speech Recognition. In Proceedings of The 12th Language Resources and Evaluation Conference. 6340-6345.

Bremerich-Vos, A., Wendt, H. & Bos, W. (2017). Lesekompetenzen im internationalen Vergleich: Testkon-zeption und Ergebnisse. In Hußmann, A., Wendt H., Bos, W., Bremerich-Vos, A., Kasper,D., Lankes, E-M., McElvany, N.,Stubbe T.C., & Valtin, R. (Hrsg.), IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich (S. 79–142). Münster u.a.: Waxmann.

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. Perspectives on Psychological Science, 8(3), 223-241.

Gawronski, B., Luke, D., & Creighton, L. A. (2024). Dual-Process Theories. In The Oxford Handbook of Social Cognition (2nd ed.).

Gebauer, S. (2019). Förderung von Lehrerkompetenzen zur adaptiven Unterrichtsgestaltung. Zum Potenzial situierter Lernumgebungen in der Lehrerfortbildung. Empirische Forschung im Elementar- und Primarbe-reich. Bad Heilbrunn: Klinkhardt.

Gronchi, G., et al. (2024). Dual-Process Theory of Thought and Inhibitory Control: An ALE Meta-Analysis. Brain Sciences, 14(1), 101.

Hippmann, K., Jambor-Fahlen, S. & Becker-Mrotzek, M. (2019). Der Einfluss familiärer Hintergrundvariablen auf die Leseleistung von Grundschulkindern im Anfangsunterricht. Zeitschrift für Erziehungswissenschaft (online). Url: https://doi.org/10.1007/s11618-018-0861-8 (zuletzt geprüft: 07.02.2019).

John, P., Brooks, B. & Schriever, U. (2019). Speech acts in professional maritime discourse: A pragmatic risk analysis of bridge team communication directives and commissives in full-mission simulation. Journal of Pragmatics, 140, 1/2019, 12-21.

John, P. & Brooks, B. (2014). Lingua Franca and its grammar footprint : introducing an index for quantifying grammatical diversity in writen and spoken language, Journal of Quantitative Linguistics, 21(1), 12/2013, 22-35.

John, P., Noble, A., Takagi, N. & Björkroth, P. (2015) Using Computer Dialogue Systems for Providing a Student-Centred Teaching Approach in SMCP- Based Maritime Communication (workshop). International Maritime English Conference (IMEC), Johor, Malaysia; 10/2015

Johns, J.L. & Berglund, R.L. (2006). Fluency. Strategies and assessments. Newark, DE: International Reading Association.

KMK Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2005). Bildungsstandards im Fach Deutsch für den Primarbereich. Beschluss vom 15.10.2004. München: Wolters Kluwer

Lenhard, W. & Schneider, W. (Hrsg.). (2009). Lesediagnostik und Leseförderung. In: Diagnose und Förde-rung des Leseverständnisses. Göttingen: Hogrefe

Li, A. (2019). Google speech recognition is now almost as accurate as humans. 9to5Google.

Liao, H., Pundak, G., Siohan, O., Carroll, M. K., Coccaro, N., Jiang, Q. M., & Bacchiani, M. (2015). Large vocabulary automatic speech recognition for children. In Sixteenth Annual Conference of the International Speech Communication Association.

McElvany, N. Schroeder, S., Hachfeld, A., Baumert, J., Richter, T., Schnotz, W., Horz, H., Ullrich, M. (2009). Diagnostische Fähigkeiten von Lehrkräften bei der Einschätzung von Schülerleistungen und Aufga-benschwierigkeiten bei Lernmedien mit instruktionalen Bildern. Zeitschrift für Pädagogische Psychologie; 23, 34, 223-235.

Martschinke, S. (2015). Facetten adaptiven Unterrichts aus der Sicht der Unterrichtsforschung. In Lernpro-zessbegleitung und adaptives Lernen in der Grundschule. 15-32. Springer VS, Wiesbaden.

Nagler, T., Lindberg, S., & Hasselhorn, M. (2018). Leseentwicklung in der Kindheit. Kindheit und Entwick-lung, 27(1), 5-13.

Philipp, M. (2011). Wer hat, dem wird gegeben? Individuelle sowie soziodemografische Merkmale und ihre Bedeutung für den Matthäus Effekt im Leseverstehen. Leseforum Schweiz, 2/2011, 1–14.

Pinnell, G.S., Pikulski, J.J., Wixson, K.K., Campbell, J.R., Gough, P.B. & Beatty, A.S. (1995). Listening to children read aloud. Data from NAEP's integrated reading performance record (IRPR) at grade 4. Washington, DC: Office of Educational Research and Improvement.

Rasinski, T.V. (2004). Assessing reading fluency. Honolulu, HI: Pacific Resources for Education and Learning (PREL).

Reiss, K., Weis, M., Klieme, E., & Köller, O. (Hrsg.). (2019). PISA 2018: Grundbildung im internationalen Vergleich. Waxmann Verlag.

Schilcher, A. & Wild, J. (2018): Lesen. Eine Schlüsselkompetenz im interdisziplinären Forschungsdiskurs. In: Boelmann, J. (Hrsg.), Empirische Forschung in der Deutschdidaktik. Band 3: Forschungsfelder (163-184). Baltmannsweiler: Schneider.

Shivakumar, P. G., & Georgiou, P. (2020). Transfer learning from adult to children for speech recognition: Evaluation, analysis and recommendations. Computer speech & language, 63, 101077.

Wagner, L. (2016). Adaptive und evidenzbasierte Förderung im Unterricht – Wozu braucht man das? Pots-damer Zentrum für empirische Inklusionsforschung (ZEIF), 11, 1-9.

Wild, J. & Schilcher, A. (2019): Grundlagen einer systematischen schulischen Leseförderung. In: Stückl, G. & Wilhelm, M. & Kronach, W. (Hrsg.), Lehren und Lernen in der bayerischen Grundschule (1-17). Kronach, Köln: Carl Link Verlag.

Yeung, G. and Alwan, A. (2018). On the difficulties of automatic speech recognition for kindergarten-aged children. Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH:1661–1665.

Revolutionizing Early Literacy: A Cognitively-Inspired, Dual-Process Architecture for Children's Speech Recognition

2025