engineering proceedings

MDPI

Proceeding Paper

TTS and STT in Service of Education †

Zakaria El Fakir *, Oussama Kaich, El Habib Benlahmar, Sanaa El Filali and Omar Zahour

Faculty of Science Ben M'Sick, Hassan University, Casablanca 20660, Morocco; oussama.kaich-etu@etu.univh2c.ma (O.K.); elhabib.benlahmar@etu.univh2c.ma (E.H.B.); sanaa.elfilali@univh2c.ma (S.E.F.); omar.zahour@univh2c.ma (O.Z.)

* Correspondence: zakaria.elfakir1-etu@etu.univh2c.ma

† Presented at the 7th edition of the International Conference on Advanced Technologies for Humanity (ICATH 2025), Kenitra, Morocco, 9–11 July 2025.

Abstract

This article explores how Text-to-Speech (TTS) and Speech-to-Text (STT) technologies are being harnessed in education to enhance accessibility, language development, and overall learner engagement. Drawing upon theoretical frameworks in linguistics and educational psychology, we highlight the benefits TTS and STT can offer to diverse student populations, including students with disabilities, language learners, and those seeking personalized or self-paced instruction. We discuss methods for integrating TTS and STT into the classroom (hardware, software, and practical considerations) and offer case studies of effective implementations in areas such as literacy support, foreign language acquisition, and assessment. We then address the pedagogical benefits these tools provide—such as differentiated instruction, immediate feedback, and a heightened sense of learner autonomy—along with limitations and challenges that educators may encounter. In conclusion, we suggest future directions for research and practice, underscoring the importance of teacher training, ethical considerations, and ever-evolving advancements in natural language processing.

Keywords: text-to-speech (TTS); speech-to-text (STT); machine learning; natural language processing (NLP); assistive technology (AT); inclusive education; differentiated instruction

check for updates

Academic Editors: Brahim El Bhiri and Ayman S. Mosallam

Published: 14 October 2025

Citation: El Fakir, Z.; Kaich, O.; Benlahmar, E.H.; El Filali, S.; Zahour, O. TTS and STT in Service of Education. *Eng. Proc.* **2025**, *112*, 4. https://doi.org/10.3390/engproc2025112004

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

As digital technologies continue to reshape teaching and learning, Text-to-Speech (TTS) and Speech-to-Text (STT) tools have emerged as powerful solutions with wide-ranging educational applications. TTS involves converting written text into synthetic speech, while STT, also referred to as automated speech recognition (ASR), captures spoken input and transcribes it into readable text. Although once limited by rudimentary technology and high cost, recent developments in natural language processing (NLP), cloud computing, and machine learning have greatly improved their accuracy, availability, and affordability [1].

The utility of TTS and STT in education spans multiple domains:

- **Accessibility and inclusion:** For students with visual impairments, reading difficulties (e.g., dyslexia), or physical constraints affecting writing, TTS can dramatically facilitate access to learning materials. Similarly, STT supports students with hearing impairments (through real-time captioning) or motor impairments that complicate the use of standard keyboards.
- **Language acquisition and literacy:** TTS offers model pronunciation for language learners, while STT can provide immediate feedback on pronunciation or spelling errors in second-language contexts [2].

- **Enhanced engagement and autonomy:** Both tools can enable learners to engage with content in a variety of modes, promoting self-paced study, differentiated instruction, and higher motivation [3], .

This article examines TTS and STT from a theoretical, methodological, and practical perspective. We first review the theoretical framework, grounding these tools in educational psychology and inclusive pedagogy. Next, we discuss implementation methods, including hardware, software, and best practices for integrating TTS/STT in diverse classrooms. We then present case studies illustrating effective usage in literacy, language learning, and assessment. Finally, we discuss pedagogical benefits, outline limitations and challenges, and suggest future perspectives for leveraging TTS and STT in education.

2. Theoretical Framework

2.1. Accessibility and Inclusive Education

Central to the adoption of TTS and STT tools is the principle of universal design for learning (UDL), which promotes flexible pathways for instruction, engagement, and learner expression [4]. According to UDL, educational content should be made accessible through multiple means, ensuring that learners with diverse needs can fully participate. TTS can provide an auditory channel for text-based resources, while STT supplies a written alternative for auditory or spoken interactions.

In addition, Assistive Technology (AT) research underscores the importance of integrating digital tools that foster autonomy among students with disabilities [5]. TTS helps reduce barriers for individuals with reading or visual impairments, allowing them to engage with written material independently. Meanwhile, STT supports individuals with hearing impairments by generating real-time captions, creating a more inclusive classroom environment.

2.2. Language Acquisition and Cognitive Load

From the perspective of second language acquisition (SLA) research, TTS provides scaffolding for pronunciation, listening, and reading comprehension. By hearing synthesized speech while reading along, language learners can align phonological and orthographic representations, potentially accelerating vocabulary acquisition [6]. STT assists in speaking practice, offering immediate written feedback on accuracy and identifying errors for targeted remediation [3].

Moreover, Cognitive Load Theory (CLT) suggests that learners have limited working memory resources [7]. TTS and STT can help manage cognitive load by splitting information across audio and textual channels, thereby reducing the effort required to process content. For instance, a student can listen to a text while reading along, reinforcing comprehension through dual-modality input. Additionally, STT's real-time transcripts can support learners taking notes, freeing cognitive resources to focus on conceptual understanding rather than the mechanics of writing.

2.3. Motivation and Self-Determination

Self-Determination Theory (SDT) highlights the role of autonomy, competence, and relatedness in fostering motivation and engagement in learning [8]. TTS and STT support this autonomy by giving learners more control over the pace and format of their interaction with educational materials. Students who struggle with decoding texts or producing written work may feel more confident if they can rely on TTS to access content or STT to articulate ideas verbally. This sense of competence can, in turn, enhance intrinsic motivation and willingness to engage with challenging materials [9].

3. Methods of Integration

3.1. Hardware and Software Requirements

Adopting TTS and STT in educational settings generally involves minimal specialized hardware. Most modern computers, tablets, and smartphones have built-in capabilities for text-to-speech and speech recognition. However, the quality and accuracy of these tools can vary. Institutions might explore:

1. **Dedicated software solutions**: Tools like Kurzweil 3000 or Read&Write (TTS) and Dragon NaturallySpeaking or Google Speech-to-Text (STT) are widely recognized for their robust features and customization options.
2. **Built-in operating system support**: Microsoft Windows, macOS, iOS, and Android devices all come with native TTS/STT engines that can be enhanced or customized through settings and voice packs.
3. **Cloud-based services**: APIs from providers such as Google Cloud, Amazon Web Services (AWS), or Microsoft Azure offer scalable solutions for advanced speech synthesis and recognition, sometimes supporting multiple languages and specialized vocabularies.

3.2. Classroom Integration Strategies

Implementing TTS and STT effectively requires instructional design that aligns with curricular goals and student needs. Key strategies include:

- **Individualized Accommodation**: Provide TTS-enabled e-books or reading materials for students with dyslexia or visual impairments and offer STT for those with motor or hearing limitations.
- **Reading Comprehension Activities**: Encourage students to follow along with TTS while reading challenging texts. Accompany this practice with comprehension questions or note-taking tasks.
- **Language-Learning Exercises**: Use TTS to demonstrate correct pronunciation, intonation, and pacing. Deploy STT to give learners immediate feedback on their spoken output, highlighting mispronounced words or grammar issues.
- **Peer Collaboration**: Foster group activities where students read aloud, record themselves, and compare STT transcriptions or TTS renditions. This collaborative setting can promote error detection and collective problem-solving.
- **Formative Assessment**: Teachers can create low-stakes quizzes or oral exams using STT, where the system transcribes student responses for quick review. TTS can also read test items aloud to ensure clarity.

3.3. Professional Development for Educators

Teachers often require specific training to fully leverage TTS and STT, including:

- Familiarity with the technology's features, benefits, and limitations.
- Strategies for troubleshooting common issues (misrecognitions, accent bias in STT, unnatural speech in TTS, etc.).
- Pedagogical approaches to embed these tools naturally into lesson planning, rather than using them as an afterthought.
- Awareness of ethical considerations, such as data privacy and the potential for misuse (e.g., students relying excessively on STT for writing tasks).

4. Case Studies and Examples

4.1. Literacy Intervention with TTS

A U.S. middle school implemented a TTS-based program to support students reading two or more years below grade level. Over one semester, participating students accessed e-texts in social studies and English classes via a TTS application. Post-intervention assessments showed:

- A statistically significant gain in reading comprehension scores compared to a control group using traditional print materials only.
- Improved learner confidence in tackling grade-level texts, as reported in teacher observations and student self-evaluations.

Qualitative feedback also indicated that students began to approach their reading assignments more independently, relying on TTS to parse unfamiliar words rather than waiting for teacher assistance.

4.2. STT for English Language Learners (ELLs)

In a South Korean high school, educators introduced a speech-to-text platform to enhance English speaking proficiency. Learners were assigned guided conversation topics and recorded themselves speaking, while the STT tool generated real-time transcripts and highlighted potential errors. Teachers then reviewed these transcripts to provide focused feedback on pronunciation, syntax, and word choice.

A post-program survey showed that 85% of participants felt more comfortable speaking English, attributing this confidence to immediate, personalized correction. In standardized oral exams, the STT group outperformed the control group in fluency and overall pronunciation scores. Teachers noted that the technology also fostered self-reflection, encouraging students to identify and address recurring mistakes on their own.

5. Educational Benefits

5.1. Accessibility and Equity

One of the core advantages of TTS and STT is that they break down barriers to learning by offering multimodal access to information. Students who cannot read printed materials due to visual or cognitive impairments can access the same content through synthetic speech. Likewise, learners with hearing impairments or physical limitations can express themselves fully via speech recognition. These accommodations help create inclusive learning spaces where all students can thrive.

5.2. Differentiated Instruction and Personalized Learning

Because TTS and STT technologies offer different modes of presentation and production, they enable teachers to adapt instruction to each student's needs. Advanced readers, for instance, might not require TTS, while struggling readers benefit significantly from hearing text while following along visually. Such differentiation aligns with individualized education program (IEP) goals and broader personalized learning strategies.

5.3. Enhanced Motivation and Engagement

By allowing learners to interact with content on their own terms—listening instead of reading, dictating instead of typing—TTS and STT can increase intrinsic motivation. Students often find these technologies novel and empowering, which can translate into more consistent study habits and deeper engagement with academic materials [8].

5.4. Immediate Feedback and Metacognition

STT offers real-time feedback on students' spoken output, helping them pinpoint errors in pronunciation or usage. Similarly, TTS enables learners to hear how passages should sound, prompting self-checking and metacognitive reflection. When students compare their own vocalized sentences to a synthesized model or a correct STT transcript, they become active agents in the feedback process [6].

6. Limitations, Challenges, and Future Perspectives

6.1. Technological Constraints

Despite significant advances, TTS voices can still sound unnatural or robotic, potentially reducing the sense of immersion. STT accuracy often varies by accent, background noise, and complexity of the vocabulary, creating bias or misrecognition issues that can frustrate learners [3]. Additionally, institutions with limited budgets may struggle to provide high-quality devices or stable internet access needed for cloud-based speech recognition.

6.2. Pedagogical and Ethical Considerations

Teachers must strike a balance between leveraging TTS/STT for support and avoiding overreliance that could undermine skill development (e.g., reading fluency, handwriting). Furthermore, privacy concerns arise when STT systems send voice data to external servers for processing. Educators and policymakers need clear guidelines on data storage, user consent, and compliance with relevant regulations [9].

6.3. Teacher Training and Institutional Support

A major hurdle in implementing TTS and STT is the lack of training and institutional support. Educators require both pedagogical and technical know-how to seamlessly integrate these tools in the curriculum. Without adequate professional development, technologies risk becoming underutilized or misapplied, failing to deliver meaningful learning outcomes.

6.4. Emerging Trends: AI and Multilingual Support

Rapid progress in machine learning is enhancing the naturalness and fluency of TTS and improving STT accuracy across multiple languages and dialects. Future directions may include:

- Adaptive TTS voices capable of adjusting reading speed, emotional intonation, or clarity based on student feedback.
- AI-driven analytics that can interpret STT transcripts to provide targeted remediation, highlight language patterns, or predict learner progress.
- Cross-language functionalities that allow instant translations or support in bilingual classrooms.
- Integration with virtual/augmented reality, enabling hands-free educational experiences for learners of varying abilities.

7. Conclusions

Text-to-Speech and Speech-to-Text technologies hold immense promise for inclusive, personalized, and effective education. Grounded in frameworks like universal design for learning and cognitive load theory, these tools expand access to knowledge, reduce barriers, and spark higher levels of student engagement. By offering auditory and textual modes for both content delivery and student expression, TTS and STT serve the diverse needs of modern classrooms.

Nevertheless, implementing TTS and STT effectively requires proper training, infrastructure, and ethical oversight. The evolution of machine learning will likely bring even more advanced capabilities, from real-time multilingual support to adaptive voices that respond to student progress. Moving forward, rigorous research and thoughtful practice are essential to harness these technologies in ways that respect learners' individual differences, foster autonomy, and ultimately improve educational outcomes.

Author Contributions: Conceptualization, Z.E.F., O.K. and E.H.B.; Methodology, Z.E.F., O.K. and O.Z.; Software, O.K. and Z.E.F.; Validation, E.H.B., S.E.F. and O.Z.; Formal analysis, Z.E.F. and E.H.B.; Investigation, Z.E.F. and O.K.; Resources, S.E.F.; Data curation, Z.E.F.; Writing—original draft preparation, Z.E.F. and O.K.; Writing—review & editing, E.H.B., S.E.F. and O.Z.; Visualization, O.K.; Supervision, E.H.B.; Project administration, E.H.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zierau, N. The next wave of AI-driven speech technologies in education. *Educ. Technol. Res. J.* **2020**, *8*, 114–131.
2. Bashori, M.; van Hout, R.; Strik, H.; Cucchiarini, C. Integrating speech technology in foreign language learning: Effects on fluency and accuracy. *Comput. Assist. Lang. Learn.* **2022**, *35*, 1–23.
3. Chang, C.K.; Chang, C.H.; Shih, J.L. Using speech-to-text for self-regulated learning in language education. *Comput. Educ.* **2021**, *169*, 104233.
4. Rose, D.H.; Meyer, A. *Teaching Every Student in the Digital Age: Universal Design for Learning*; Association for Supervision and Curriculum Development (ASCD): Alexandria, VA, USA, 2002.
5. Bryant, D.P.; Bryant, B.R. *Assistive Technology for People with Disabilities*, 2nd ed.; Pearson: Boston, MA, USA, 2012.
6. Kim, H.; Rath, T. The effect of text-to-speech support on reading comprehension for students with reading difficulties. *J. Spec. Educ.* **2019**, *45*, 51–65.
7. Sweller, J. Cognitive load theory, learning difficulty, and instructional design. *Learn. Instr.* **1994**, *4*, 295–312. [CrossRef]
8. Deci, E.L.; Ryan, R.M. The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychol. Inq.* **2000**, *11*, 227–268. [CrossRef]
9. McKenna, E.; Oswald, D. Improving reading outcomes with text-to-speech technology: A review of evidence-based practices. *Read. Writ. Q.* **2020**, *36*, 185–202.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.