Review on the Use of Speech Synthesis Technology in Education

Yiming Wang[1], Zai Xu[2], Zikang Zheng[3], Zifei Zheng[3], Jihe Wu[3]

1. Beijing 21st Century International School, Haidian District, Beijing, 100089

2. Beijing Shijia Primary School, Dongcheng District, Beijing, 100010

3. Shenzhen Experimental School Primary Division, Shenzhen, Guangdong Province, 518028

Abstract: Speech synthesis technology, also known as text to speech conversion technology, can convert text information into natural and smooth speech output. In the field of education, the application of speech synthesis technology is becoming increasingly widespread, providing great convenience for teaching activities and learning processes. This article will provide an overview of the application of speech synthesis technology in education, analyze its effectiveness in teaching aids, learning aids, and assessment exams, and explore its future development trends.

Keywords: Speech synthesis technology; Education application

1. Introduction

With the rapid development of technology, speech synthesis technology has become an important auxiliary tool in the field of education. This article provides an overview of the application of speech synthesis technology in education, including its applications in auxiliary teaching, learning assistance, and evaluation and examination, and explores its effectiveness evaluation and future development trends. Through this study, the aim is to provide educators and researchers with a comprehensive understanding of the application of speech synthesis technology in education, and to provide reference for its further application and development in the field of education.

2. Overview of Speech Synthesis Technology

Speech synthesis technology, also known as text to speech conversion technology, is a technique that converts text information into natural and fluent speech output. With the rapid development of artificial intelligence and deep learning, speech synthesis technology has made significant progress. Speech synthesis technology is based on the theoretical foundations of multiple disciplines such as linguistics, acoustics, and signal processing. By constructing complex algorithms and models, it achieves precise understanding of text content and flexible control of speech generation. The core of this technology lies in its ability to simulate the movement patterns of human vocal organs, generating speech with naturalness and comprehensibility. In practical applications, speech synthesis technology has broad application prospects. In the field of education, it can serve as a teaching aid to help students improve their language learning and reading abilities. In the field of public services, it can be used to provide services such as voice navigation and information retrieval, making people's lives more convenient. In addition, speech synthesis technology plays an important role in entertainment, advertising, and other fields, providing personalized speech content for various scenarios. With the continuous progress of technology and the expansion of application scenarios, speech synthesis technology will play a greater role in the future, providing people with more convenient and efficient services[1].

3. The application of speech synthesis technology in education

3.1 Teaching assistance

Speech synthesis technology has played an important role in teaching assistance, providing educators with a new teaching method. Specifically, the application of this technology enables teachers to impart knowledge to students more efficiently and vividly. Through speech synthesis technology, teachers can easily record and explain audio, conveying complex concepts and knowledge points to students in a clear and vivid voice form. This form of explanation not only allows students to learn anytime and anywhere, but also to some extent alleviates the teaching pressure of teachers. Compared to traditional blackboard or PPT teaching methods, speech synthesis technology makes teaching content more vivid and interesting, which helps to stimulate students' interest and enthusiasm in learning. Speech synthesis technology can

also assist teachers in classroom teaching management. For example, in the classroom questioning session, teachers can use speech synthesis technology to play questions to students, avoiding comprehension barriers caused by accent or speed issues. At the same time, teachers can also use speech synthesis technology to provide real-time feedback on student answers, helping students correct errors in a timely manner and improve learning effectiveness. In addition, speech synthesis technology can also be combined with online teaching platforms to provide strong support for distance education. By recording audio explanations and uploading them to online platforms, teachers can provide learning resources for students who are unable to attend school, achieving the sharing and optimization of educational resources.

3.2 Learning assistance

Speech synthesis technology offers unique benefits for learning. It provides convenient tools and strong support for personalized learning, helping students access accurate learning materials, including pronunciation, text reading, and explanations of terms. The technology can tailor content to individual needs, like pronunciation guidance or targeted exercises, boosting effectiveness, motivation, and confidence. Plus, it integrates with mobile devices for anytime, anywhere learning, freeing students from time and location constraints.

3.3 Evaluation Examination

In the field of assessment examination, the application of speech synthesis technology provides a new means to assess students' language ability and academic performance, making the assessment examination more objective, fair and efficient. Speech synthesis technology provides a reliable basis for oral evaluation by generating standardized pronunciation templates. Traditional oral evaluation is often influenced by subjective and fatigue factors of raters, while speech synthesis technology can ensure the consistency and accuracy of each pronunciation, thus improving the fairness and reliability of oral evaluation. By comparing students' pronunciation with the standard template, teachers can quickly identify pronunciation errors or shortcomings and provide targeted guidance and feedback. Speech synthesis technology also plays an important role in computer-aided reading and writing test.

4. Future development trends of speech synthesis technology

4.1 Multilingual support

With the deepening of globalization, language diversity has become a practical issue that cannot be ignored. In this context, the multilingual support function of speech synthesis technology is particularly important. Multilingual support not only means that the technology can handle text to speech tasks in different languages, but also requires it to accurately grasp the pronunciation rules, prosody, and intonation changes of each language. The achievement of this goal is inseparable from the continuous progress of deep learning and natural language processing technology. In the future, with the continuous improvement of multilingual support technology, speech synthesis technology will be able to better meet the needs of global users and help break down language barriers. Whether it's English, Chinese, French, or any other language, users can obtain natural and fluent voice output through this technology. This will bring unprecedented convenience and possibilities for cross-cultural communication, international education, language learning, and other fields[2].

4.2 Personalized customization

Each person's voice has unique characteristics, and different timbres, tones, and speech speeds together constitute an individual's vocal characteristics. Therefore, users expect speech synthesis technology not only to simply convert text to speech, but also to obtain personalized outputs that match their own voice characteristics. In order to meet this demand, future speech synthesis technology will pay more attention to personalized customization services. By collecting user voice samples and analyzing sound characteristics, the technology can generate speech output that is highly similar to the user's voice. This personalized speech synthesis technology can not only be applied in entertainment, advertising, and other fields, providing users with speech content that is closer to their own image, but also playing an important role in education, auxiliary communication, and other scenarios. In the field of education, personalized speech synthesis technology can provide students with customized learning resources. According to the learning styles and needs of different students, teachers can generate phonetic tutorials or auxiliary materials that match their vocal characteristics, helping students better understand and master knowledge. This personalized learning approach will help enhance students' interest and effectiveness in learning, and promote innovation and development in educational methods.

4.3 Multimodal Interaction

Traditional speech synthesis technology mainly focuses on converting text into speech, while multimodal interaction emphasizes the fusion of multiple perceptual methods to provide users with a more natural and rich interaction experience. Multimodal interactive speech synthesis technology will combine multiple information sources such as vision, hearing, and touch to provide users with a more comprehensive way of presenting and interacting with information. In the field of education, this technology can be applied to virtual classrooms, online learning and other scenarios, creating a more vivid and interesting learning environment for students by combining various elements such as images, animations, and gestures. Students can not only receive speech information through auditory means, but also observe images and

animations visually, and even interact with virtual environments through touch, thereby gaining a deeper understanding and mastery of knowledge. In addition, multimodal interactive speech synthesis technology can also be applied in fields such as smart assistants and smart homes, providing users with a more convenient and intelligent service experience.

4.4 Intelligent emotional expression

Human speech contains rich emotional information, and traditional speech synthesis techniques often appear more mechanical and monotonous in terms of emotional expression. Future speech synthesis technology will place greater emphasis on understanding and expressing emotional information, achieving precise capture and simulation of human emotions through technologies such as deep learning. In the field of education, this speech synthesis technology with emotional expression ability can more vividly present teaching content, stimulate students' learning interest and emotional resonance. At the same time, it can also adjust parameters such as intonation and speed of speech according to students' learning status and emotional changes, in order to provide a more personalized learning experience[3].

5. Conclusion

The application of speech synthesis technology in education has achieved significant results, providing great convenience for teaching activities and learning processes. However, there are still some challenges and issues that need further research and resolution. In the future, we should continue to deepen the research and application of speech synthesis technology, unleash its greater potential in the field of education, and contribute to cultivating more outstanding talents.

References

[1] Sébastien M L, Simon K, Naomi H. The limits of the Mean Opinion Score for speech synthesis evaluation[J]. Computer Speech & Language, 2024, 84.

[2] Bang W C, Chun C. Effective Zero-Shot Multi-Speaker Text-to-Speech Technique Using Information Perturbation and a Speaker Encoder[J]. Sensors, 2023, 23(23).

[3] Gökay D. Differential convolutional network for noise mask estimation[J]. Applied Acoustics, 2023, 211.