Next Wave of Neural TTS: A Review of Efficiency, Zero-Shot Adaptation, and Expressiveness

Yuvraj Sinha
Department of Computer Science & Engineering, Sharda University, Greater Noida, Uttar Pradesh, India

Dr. Sandeep Kumar
Department of Physics, Guru Jambheshwar University of Science and Technology, Hisar, Haryana, India

Abstract

Neural Text-to-Speech (TTS) synthesis has become remarkably natural, making the research frontier transition to specialized and real-world use. A decade of (2025) recent contributions are intersumed in this review to define key trends and serious gaps in research. We examine developments in three main dimensions: (1) Efficiency and Accessibility, (2) Data Efficiency and Adaptation, (3) Expressiveness and Robustness, in the context of emotion classification, linguistic sensitivity in low-resource languages, and security watermarking. Our synthesis indicates that there is a gap in research: individual models are doing great in a specific area (e.g., efficiency or zero-shot), but there are no unified frameworks, which are efficient (on device), data-scarce (zero-shot), and expressive models (prosody/emotion controlled).

Keywords

Neural Text-to-Speech, Data-Efficient Learning, Expressive Speech Synthesis, Zero Shot Adaptation, Speech Model Efficiency.