# Summary of Research Papers

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Name of Paper | Author | Year of Publication | Publication Name | Implemented System | Advantages/Features | Limitation/Research Gap |
| Voicebox: Text-Guided Multilingual Universal Speech Generation at Scale | Le et al. | 2023 | NeurIPS | Generalized multilingual TTS and editing model | Handles multilingual, denoising, and in-context TTS | Not specialized for real-time assistant deployment |
| Neural Codec Language Models are Zero-Shot Text-to-Speech (VALL-E) | Wang et al. | 2023 | Microsoft / arXiv | Neural codec model for natural speech generation | Produces high-quality, speaker-specific TTS; low data requirement | High computational cost; limited open-source access |
| User Experience and Usability of Voice User Interfaces: A Systematic Literature Review | A. M. Deshmukh, R. Chalmeta | 2024 | Information (MDPI) | Review of voice UI design studies and usability challenges | Identifies usability factors, personalization importance, and accessibility gaps | No implementation; focuses mainly on survey data |
| Desktop Voice Assistant for Elderly People using Feed-Forward Neural Network for Intent Recognition | Kashish Garg, Taj Alam | 2024 | IC3 Conference | Feed-forward neural network–based intent recognition model for desktop commands | Achieved ~78.5% intent accuracy; tailored for elderly users | Dataset limited; lacks multilingual and adaptive learning |
| Execution-guided Within-Prompt Code Generation | Gust Verbruggen, Ashish | 2025 | ICLR 2025 Conference | A within-prompt search system | Combines code generation | Focused on straight-line code generation |