

Crafting Emotional TTS Conversation Responses Based on User Preferences

LAYA IYER, Stanford University, USA

SANMI KOYEJO, Stanford University, USA

The emotionality of current text-to-speech systems can be critical to the usefulness and use of the experience of conversation artificial intelligence systems. Nevertheless, many existing systems lack emotional valence. This paper proposes a model pipeline created to encapsulate the tonal preference of response to a query. The tone of the response is based on indicated user preferences, which are estimated using a feedback procedure consisting of pairwise comparisons of two provided responses with different tones. A predictive model is trained using this preference data, and the response tones change to reflect preferences. This model examines the relationship between user message tonality and preferred output tonality of the text-to-speech (TTS) model. While this paper is a preliminary exploration of tonality analysis, we highlight some of the potential of our approach and propose extensions to support tonal shifts over the trajectory of a conversation.

Additional Key Words and Phrases: LLMs, Tonality, Text to Speech, User Preferences, Emotional Discernibility, ETTS