Contribution ID: **143** Type: **Poster**

The art of audio: generating text and images

In this paper, we explore a novel multi modal approach that bridges the gap between audio, text and visual representation by utilizing the Clotho dataset, a comprehensive collection of audio recordings with detailed textual annotation. Our methodology involves two primary stages: first we employ state-of-the-art audio processing and natural language processing techniques to transcribe audio data into accurate textual representations. In the second stage, we transform these transcriptions into visual forms, creating an innovative way to visualize the content and structure of audio information.
Through this study, we aim to advance the field of multi modal data analysis by demonstrating how audio, text, and visual elements can be seamlessly integrated to offer enriched user experiences and deeper analytical capabilities. The proposed framework contributes to ongoing research in this field and opens up new possibilities for future exploration and applications.

Primary authors: Ms JAYAN, Krishnaja; PALIWAL, aashita

Presenter: PALIWAL, aashita

Track Classification: Innovation and Technology for Sustainability