

Immersivaudio: audio generation based on video features.

Group 01

Michele Vitale
ist1111558

Daniele Avolio
ist1111559

Teodor Chakarov
ist1111601

Abstract—This document is a model and instructions for \LaTeX . This and the `IEEEtran.cls` file define the components of your paper [title, text, heads, etc.]. ***CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract.**

Index Terms—component, formatting, style, styling, insert

I. INTRODUCTION

The topic of media generation has been exponentially growing during the last few years. Since the release of models and services based on state-of-art AI techniques, such as StableDiffusion and ChatGPT, it has been frequent to have media generative applications, with the most important part being that they can be easily accessed even by users that do not have competences and knowledge on Artificial Intelligence. Our proposal is a pipeline composed of different models, trained or open-weights, to generate audiovisual multimedia.

II. PROBLEM DESCRIPTION

Main goal of our project is to provide a Che vogliamo fare

III. PROBLEM IMPORTANCE

Persone ipovedenti o cieche Audio enhancer per contenuti (video di persona che parla, musica di sottofondo aggiunta in post)

IV. ARCHITECTURE

architecture schema picture

A. Feature extraction

video \rightarrow features extraction

B. Prompt enhancer

features \rightarrow well constructed prompt

C. Audio generator

prompt \rightarrow generated audio

D. Reconstructor

video + audio \rightarrow output multimedia

REFERENCES

REFERENCES

- [1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529–551, April 1955.