Statistics for audio with text, and multi-modal models

Adam Boult (www.bou.lt)

April 29, 2024

Contents

Part I

Audio and Natural Language Processing

Chapter 1

Voice recognition

1.1 Audio

Chapter 2

Song recognition

2.1 Audio

Chapter 3

Audio-to-text / text-to-audio

3.1   Audio