

Statistics for audio with text, and multi-modal models

Adam Boulton (www.bou.lt)

March 23, 2024

Contents

I	Audio and Natural Language Processing	2
1	Voice recognition	3
2	Song recognition	4
3	Audio-to-text / text-to-audio	5

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