

SONG DESCRIBER: A PLATFORM FOR COLLECTING TEXTUAL DESCRIPTIONS OF MUSIC RECORDINGS

Ilaria Manco¹, Benno Weck², Philip Tovstogan², Minz Won², Dmitry Bodganov²

¹Queen Mary University of London, ²Universitat Pompeu Fabra



Collecting the first dataset of music captions

Despite a growing interest in integrating natural language text within MIR systems, there are currently no public datasets **with aligned audio-text data** to support this area of research

To fill this gap, we present Song Descriptor, an **open-source data annotation platform** designed to crowdsource music captions of CC-licensed recordings



Slow-paced R&B ballad with elegant vocals and elements of pop music

What is a music caption?

We define a music caption as: *a short piece of text that describes the content of a music recording*

What will the dataset be useful for?

A dataset of (music, caption) pairs will offer

1. insights into how people describe music
2. training and evaluation data for ML
3. a first benchmark for audio-and-language research in the music domain

Welcome to Song Descriptor! 🙌

The data collection protocol is composed of three stages: onboarding, annotation and evaluation



song-describer.streamlit.app

After onboarding, participants are asked to annotate a music recording (Task 1) or evaluate previously collected captions (Task 2)

Task 1: Annotation

✎ Let's get captioning!

In this task, you'll listen to a music track (up to 2 minutes) and write a short description of it.

🔊 Before pressing play, please make sure the volume of your headphones or speakers is set to a comfortable level.

► Note that some tracks may have lyrics with sensitive content. If you don't feel comfortable listening to a track, you can skip it by clicking the button at the bottom of this page.

▶ 0:00 / 2:00 🔊

Track info

How would you describe this track in one sentence?

📖 Instructions

💬 Need some examples?

0/512

How familiar are you with this kind of music?

Rate your familiarity:

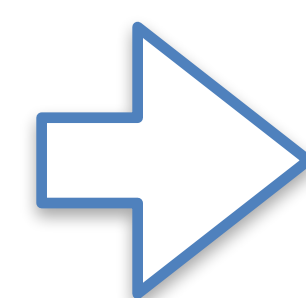
Not familiar at all

Not familiar at all

Very familiar

Submit

Skip this track



Task 2: Evaluation

⚖️ Evaluation

In this task, you'll first be shown a piece of text and asked whether it is a valid caption.

📌 A caption is valid if it's a comprehensible and grammatically correct English sentence describing a piece of music.

If the caption is valid, you'll then be asked to listen to a music track and rate how well the caption describes it.

Is the caption below valid?

Caption

Slow-paced R&B ballad with elegant vocals and elements of pop music

Yes ✓

No ✗

How well does it describe this track?

▶ 0:00 / 2:00 🔊

Track info

Rate the caption from 1 to 5

★

★★★★★



Contact: i.manco@qmul.ac.uk

Code: github.com/ilaria-manco/song-describer