Mean Error Histogram

Accompaniment Loudness Estimation Absolute Error

Vocal Loudness Estimation Absolute Error

# Mixture Vocal Loudness Estimation and Genre Analysis

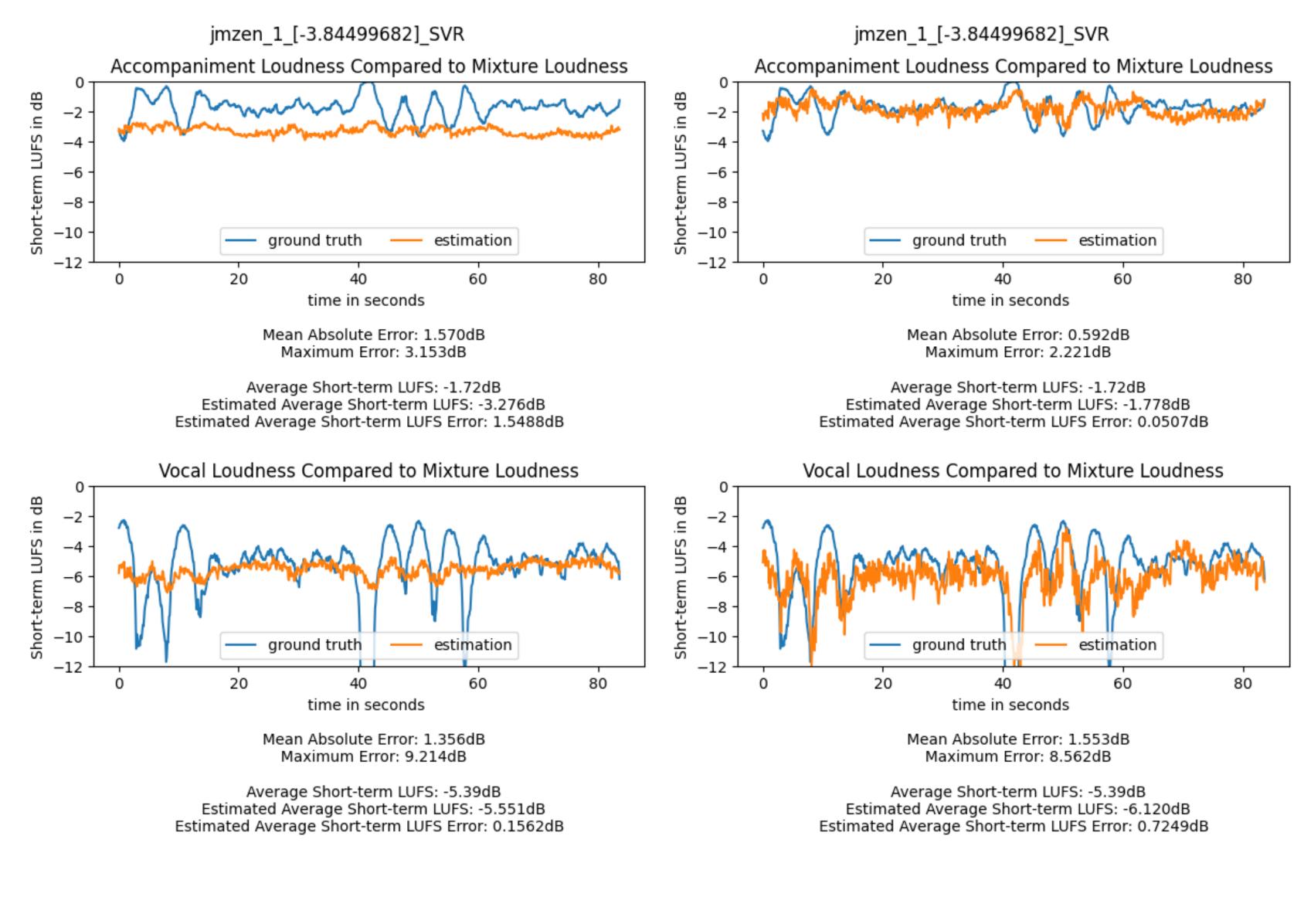
### Introduction

With the isolated vocal and accompaniment tracks available for training, we can use machine learning to estimate the vocal loudness from the mixture signal.

The ground truth is the short-term loudness of isolated vocal loudness. The audio signals are sliced into overlapped segments of 3 seconds length with the hop size of 100 ms following the EBU recommendation

The input features to the SVR model are the first 20 Mel-Frequency Cepstrum Coefficients (MFCC), 128 VGGish embeddings, and the short-term loudness of the mixture signal.

#### ള് 4.0% 4.0% 3.0% 3.0% 2.0% Because the mixture loudness can be measured directly from the audio signal, estimating 1.0% the relative vocal loudness (vocal-to-mixture ratio) is equivalent to estimating the absolute vocal loudness. The target vocal loudness values are transformed from the absolute shortterm loudness to the loudness difference compared with the mixture loudness, Average Loudness Estimation Error Histogram (file Level) Accompaniment Loudness Estimation Error Vocal Loudness Estimation Error 10.0% $STL_{\Delta vox} = STL_{vox} - STL_{mixture}$ 6.0% <u>പ്</u> 4.0% 2.0% where STL∆vox is the relative vocal loudness, STLvox is the short-term loudness of the Short-term LUFS Error in dB Short-term LUFS Error in dB Mean Absolute Error: 0.767dB vocal signal, and STLmixture is the short-term loudness of the mixture signal. Maximum Error: 2.374dB Maximum Error: 4.807dB Estimation Genre Analysis jmzen\_1\_[-3.84499682]\_SVR jmzen\_1\_[-3.84499682]\_SVR Accompaniment Loudness Compared to Mixture Loudness Accompaniment Loudness Compared to Mixture Loudness -10



- The left figure uses epsilon of 0.3, and the right figure uses epsilon of 0.1.
- Epsilon = 0.3 yields lower MAE, but unable to predict the loudness outside of a small range.

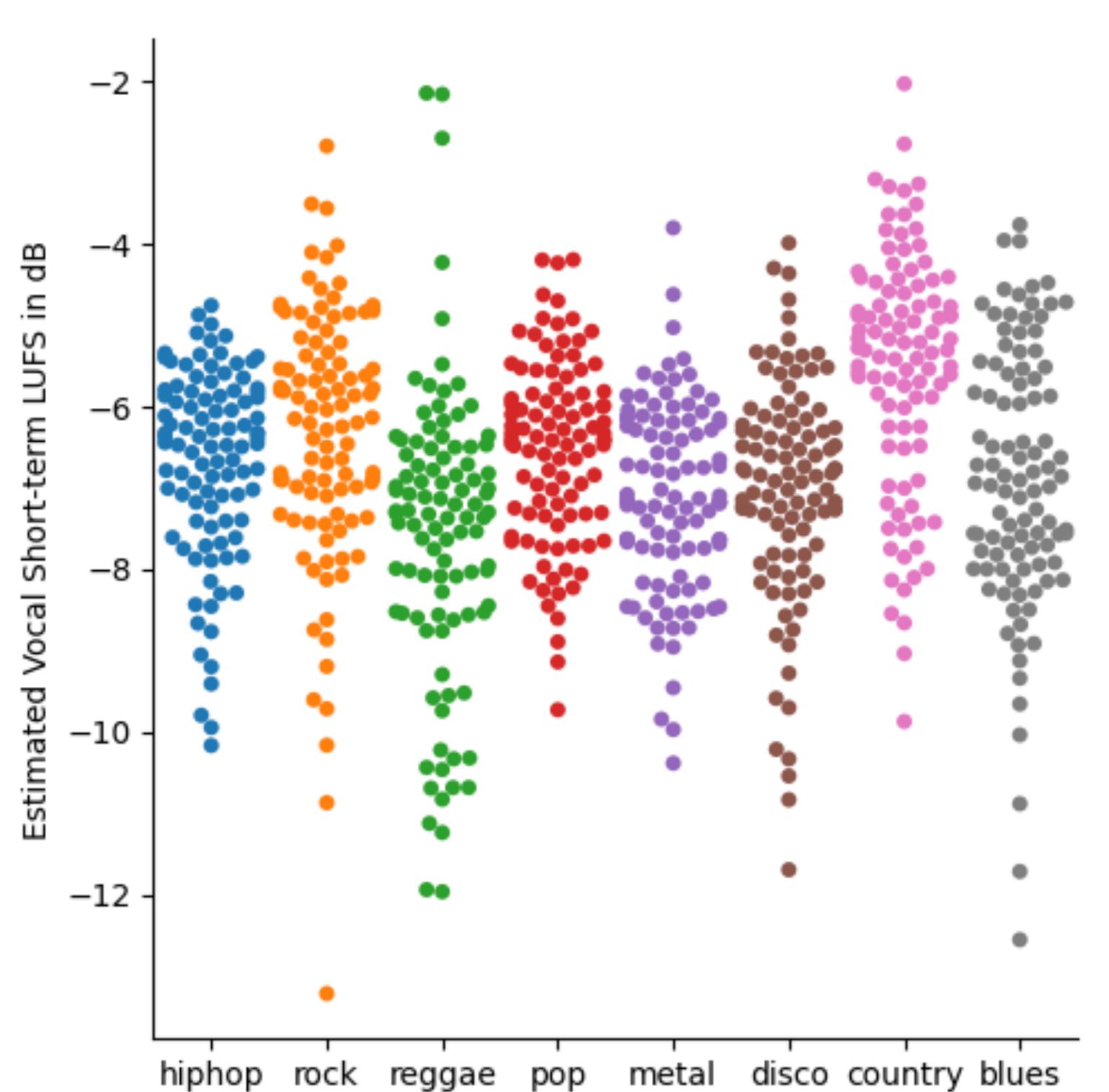
# Results

TABLE I

RESULTS OF THE VOCAL AND ACCOMPANIMENT LOUDNESS ESTIMATION

	VOX MAE(dB)	VOX ME	ACC MAE	ACC ME
Mean*	3.65	6.35	1.17	3.65
SVR	1.86	7.12	1.00	3.12

\*The baseline system.



Error Histogram

9 4.0%

3.0%

6.0%

SVR Error Histogram

6.0%

5.0%

g 4.0%

3.0%

2.0%

1.0%

7.0%

6.0%

5.0%

Accompaniment Loudness Estimation Absolute Error

Vocal Loudness Estimation Absolute Error

Genre	hiphop	rock	reggae	pop	metal	disco	country	blues
Relative Vocal Loudness (dB)	-6.69	-6.36	-7.58	-6.52	-7.13	-7.02	-5.51	-6.90

## CONTACT

Kelian (Mike) Li Music Informatics Lab Center for Music Technology kli421@gatech.edu