



# MUSICRL

Dopasowywanie Generacji Muzyki  
do Preferencji Ludzkich

Krzysztof Sawicki  
Natalia Safiejko  
Wojciech Grabias

# WPROWADZENIE

- MusicRL to wstępnie przetrenowany autoregresyjny model MusicLM, dostrojony za pomocą uczenia ze wzmocnieniem w celu maksymalizacji nagród na poziomie sekwencji. Funkcja nagród zwraca uwagę na zgodność z zapytaniem oraz na samą jakość dźwięku przy pomocy wybranych oceniających.



# GUSTA I GUŚCIKI

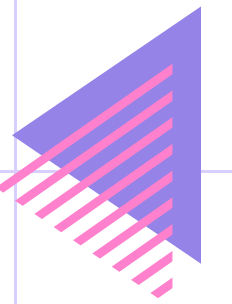
- Ocena muzyki jest subiektywna
- Problemy z dywersyfikacją odpowiedzi
- Brak wiedzy eksperckiej
- Problem z pogodzeniem muzykalności, akustyczności i zgodności z promptem



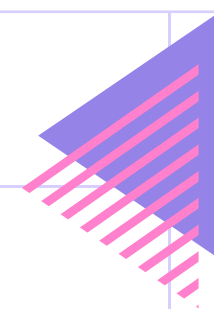
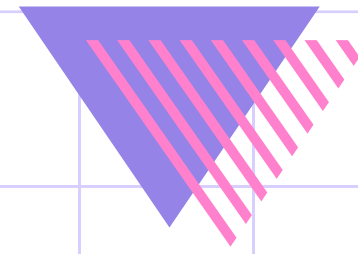
**Twoja opinia**  
**(Cringowa, niebieskopigułowa)**



**Moja opinia**  
**(Bazowana, czerwonepigułowa, fajna)**

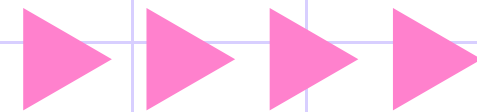


# WCZEŚNIEJSZE PRACE



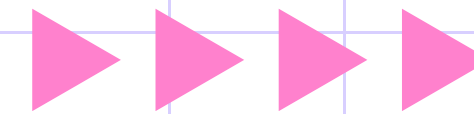
Niezadowalająca  
jakość outputu

2020



Dobra, lecz zbyt  
krótka ścieżka  
dźwiękowa

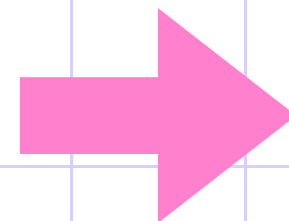
2022



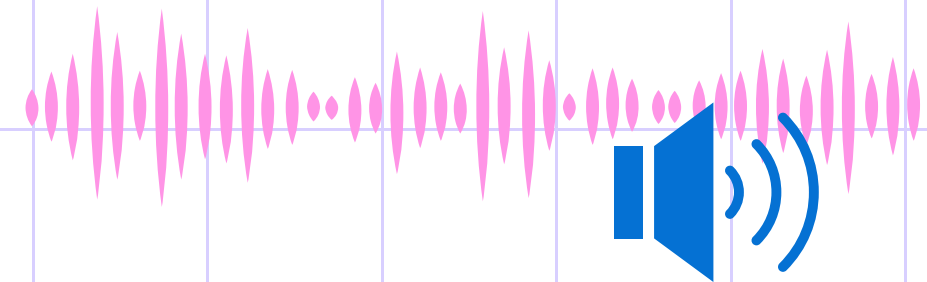
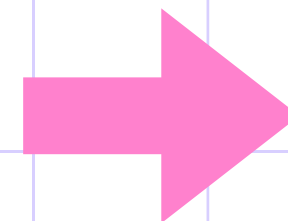
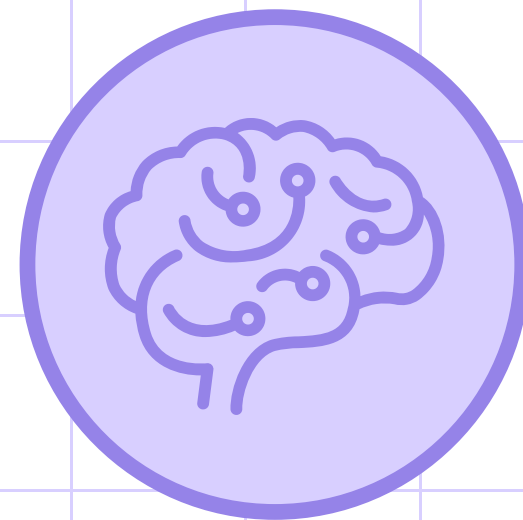
Szkolenie  
stochastycznymi  
metodami na  
zbiorach offline  
(niepewność  
wyników)

2023

Muzyka do  
nauki w  
deszczowy  
dzień



MusicLM



Funkcja  
nagrody

Reinforce preferred behavior



# METODYKA

MusicLM opiera się na dwóch różnych rodzajach reprezentacji audio do generacji: tokenach semantycznych oraz tokenach akustycznych. Początkowo został wprowadzony jako 3-stopniowy model autoregresyjny oparty na transformatorkach

1

**Mapowanie  
pomiędzy  
tokenami  
a MuLan**

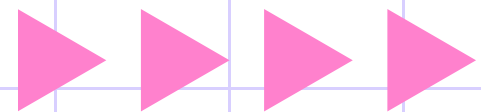
2

**Pierwszy stage  
predykcji  
SoundStream  
RVQ**

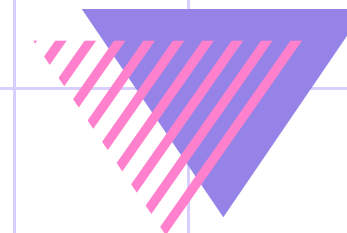
3

**Drugi stage  
predykcji  
SoundStream  
RVQ**



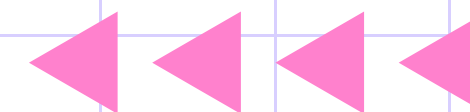
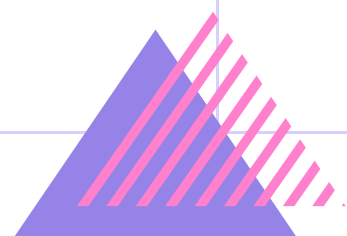


# PROCEDURA RL



$$\mathbb{J}(\theta) = (1 - \alpha) \left[ \sum_{t=0}^T \log \pi_{\theta}(a_t | s_t) \left( \sum_{i=t}^T r(s_i) - V_{\phi}(s_t) \right) \right] \\ - \alpha \sum_{t=0}^T \sum_{a \in A} [\log(\pi_{\theta}(a | s_t) / \pi_{\theta_0}(a | s_t))],$$

$$\min_{\phi} \mathbb{E}_{\pi_{\theta}} \sum_t \left( \sum_{k=t}^T r(s_k) - V_{\phi}(s_t) \right)^2.$$





# CO BRANO POD UWAGĘ W DOSKONALENIU MODEŁU?



## **zgodność z tekstem**

---

uśredniany  
MuLan score dla  
trzech 10-  
sekundowych  
fragmentów

## **jakość akustyczna**

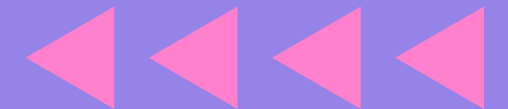
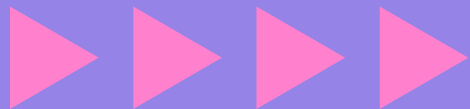
---

ocena  
zanieczyszczenia klipu  
estymator do  
przewidywania MOS -  
Mean Opinion Score

## **preferencje użytkownika**

---

zbieranie opinii  
poprzez  
porównywanie  
klipów parami



# EXPERIMENTAL SETUP





# Dane

- **model LaMDA** - generowanie opisów popularnych piosenek
- **MusicCaps** - zbiór danych z opisami utworów
- odpowiedzi użytkowników

<div>▲ aspect_list</div> <div>A list of aspects describing the music.</div>	<div>▲ caption</div> <div>A multi-sentence free text caption describing the music.</div>
<b>5518</b> unique values	<b>5521</b> unique values
['low quality', 'sustained strings melody', 'soft female vocal', 'mellow piano melody', 'sad', 'soul...]	The low quality recording features a ballad song that contains sustained strings, mellow piano melod...
['guitar song', 'piano backing', 'simple percussion', 'relaxing melody', 'slow tempo', 'bass', 'coun...]	This song features an electric guitar as the main instrument. The guitar plays a descending run in t...



# DOSKONALENIE MUSICLM

Ten sam algorytm RL i te same hiperparametry

MusicRL-R

20 000 iteracji

Kombinacja liniowa  
nagrody MuLan i  
jakości

MusicRL-U

5000 iteracji

Model nagród  
preferencji  
użytkownika

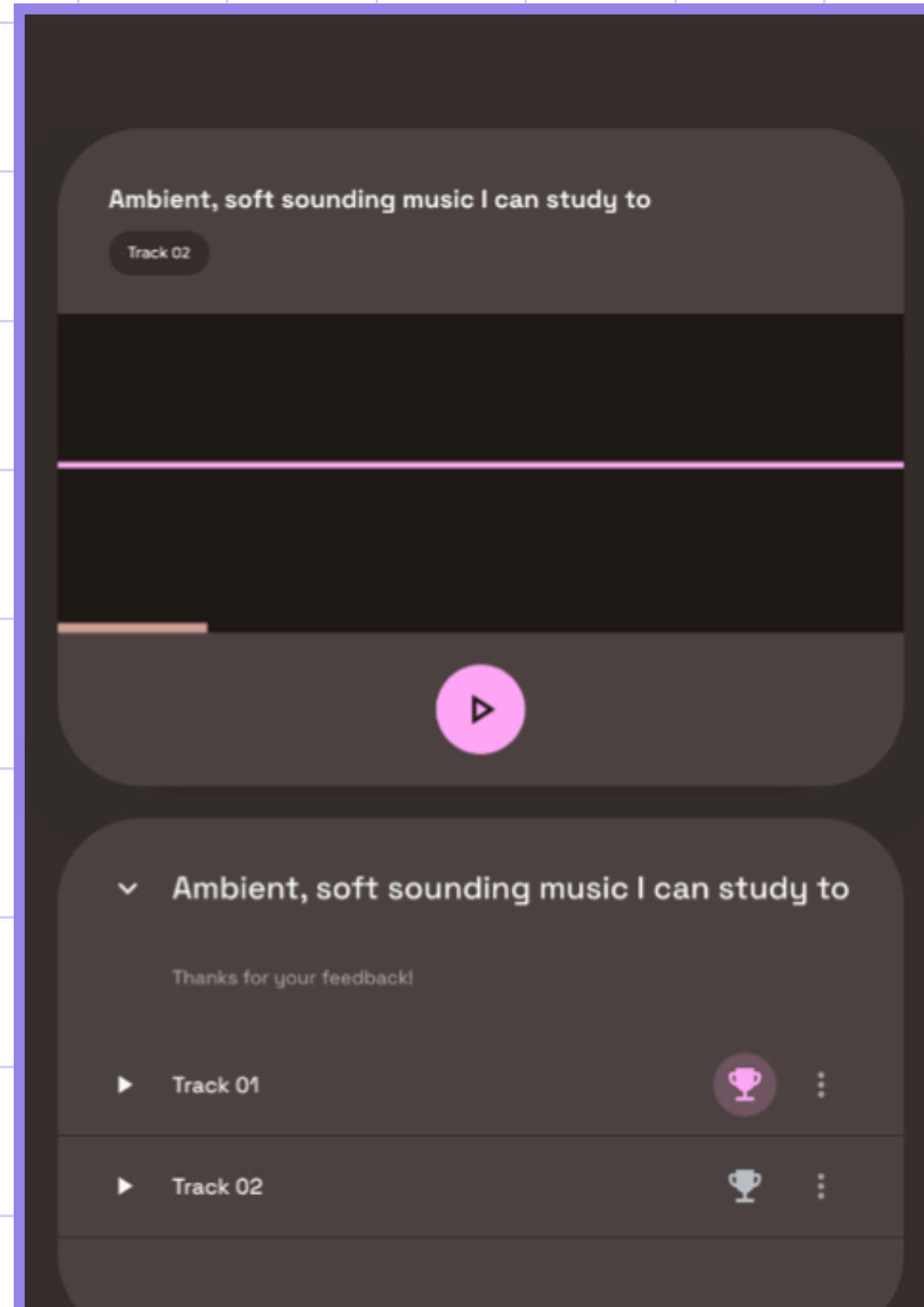
MusicRL-RU

1000 iteracji na RL-R

Sekwencyjne podejście  
MuLan i jakość  
nagroda preferencji  
użytkownika



# EWALUACJA



- oceniający z doświadczeniem słuchania różnorodnych stylów muzycznych i biegli w języku angielskim
- ocena pod względem zgodności z tekstem promptu, jakość akustyczną i ogólną atrakcyjność dźwięków





# WYNIKI



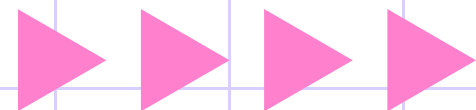
Czy RL może polepszyć jakość modeli generujących muzykę?

Czy systemy nagród można łączyć by uzyskiwać całościowo lepsze rezultaty?

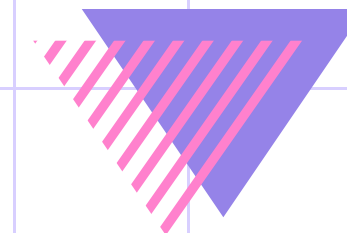


Czy RLHF pomoże dostosować się do ogólnych preferencji?

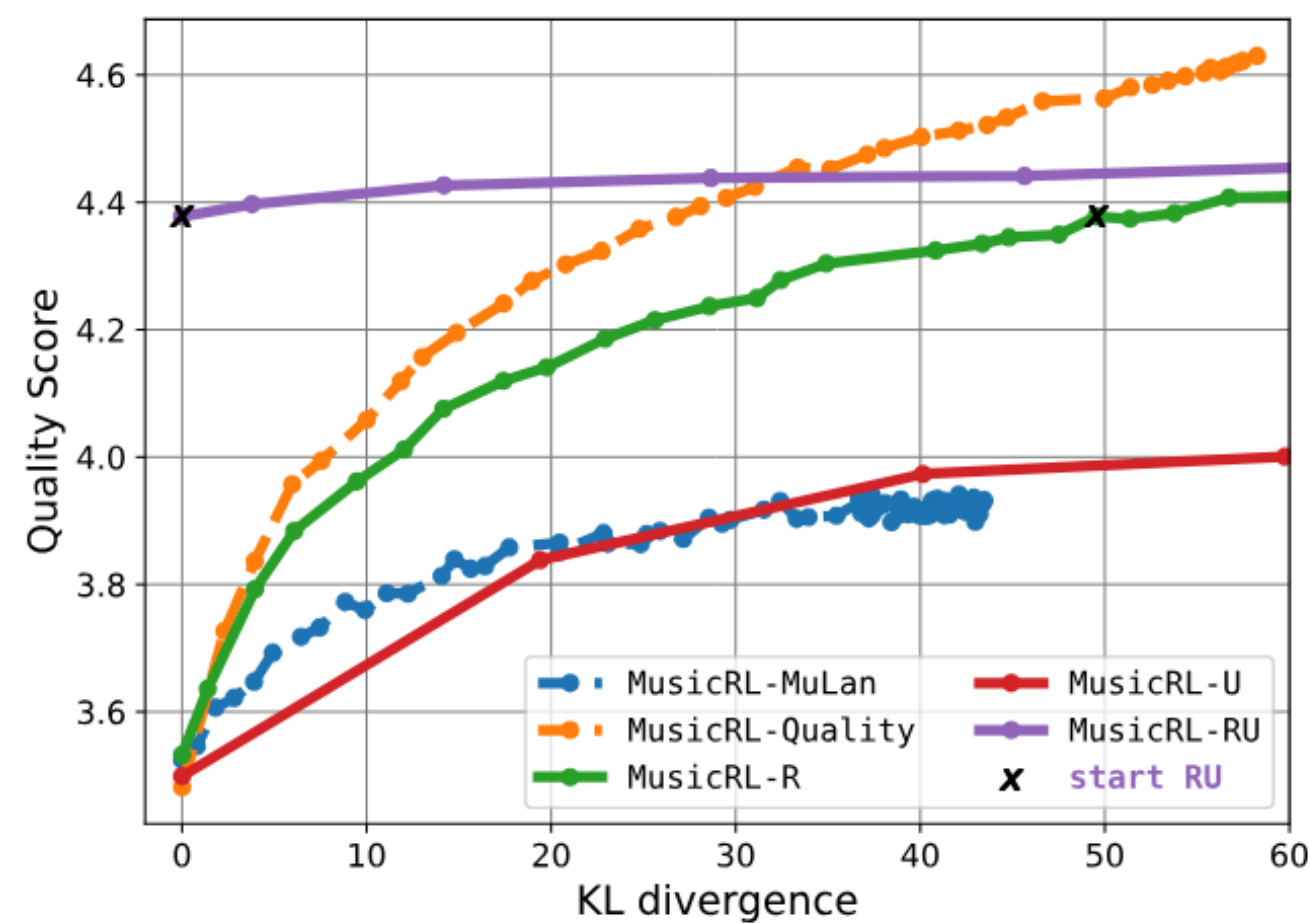




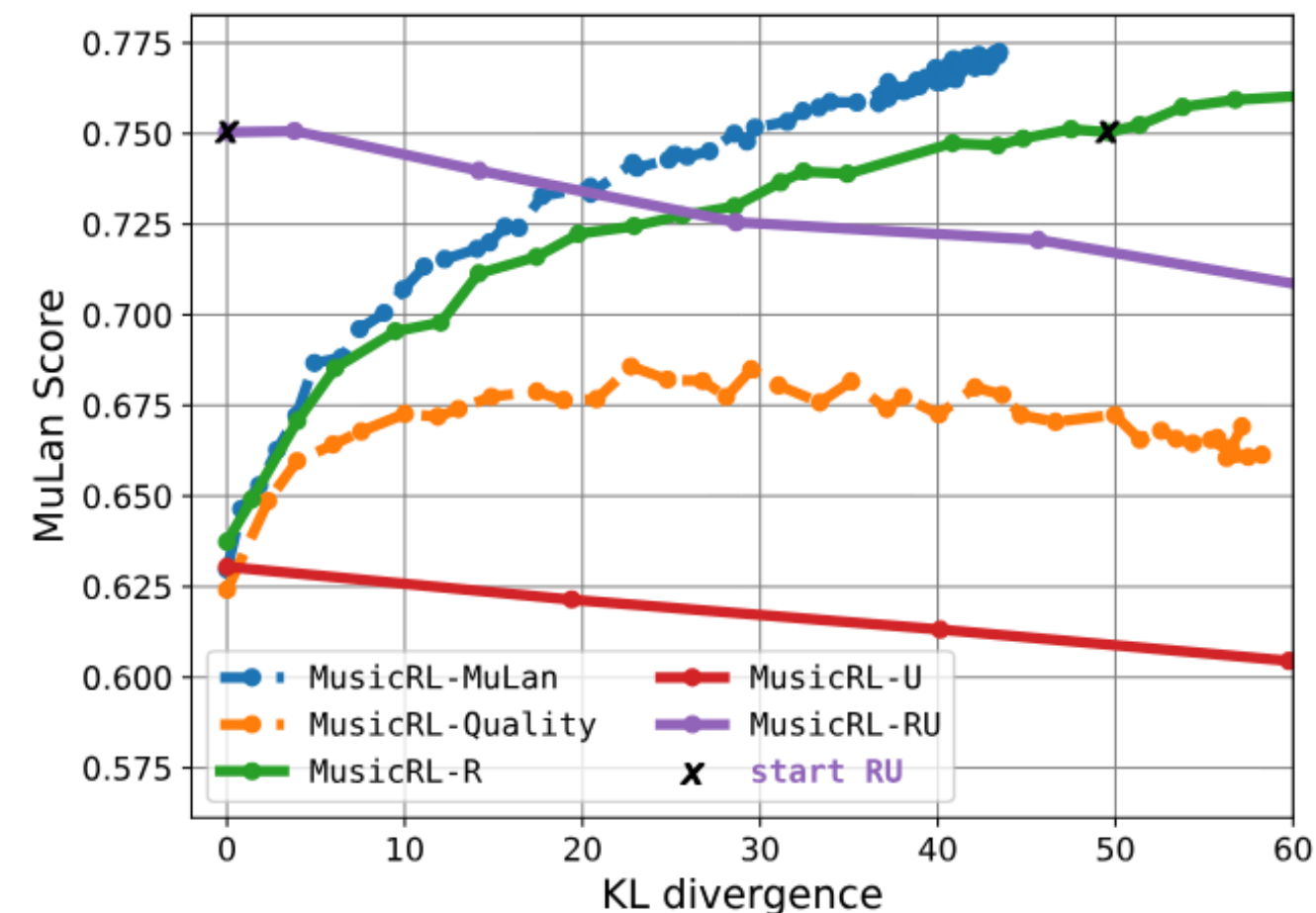
# WYNIKI IŁOŚCIOWE



## Quality



## MuLan

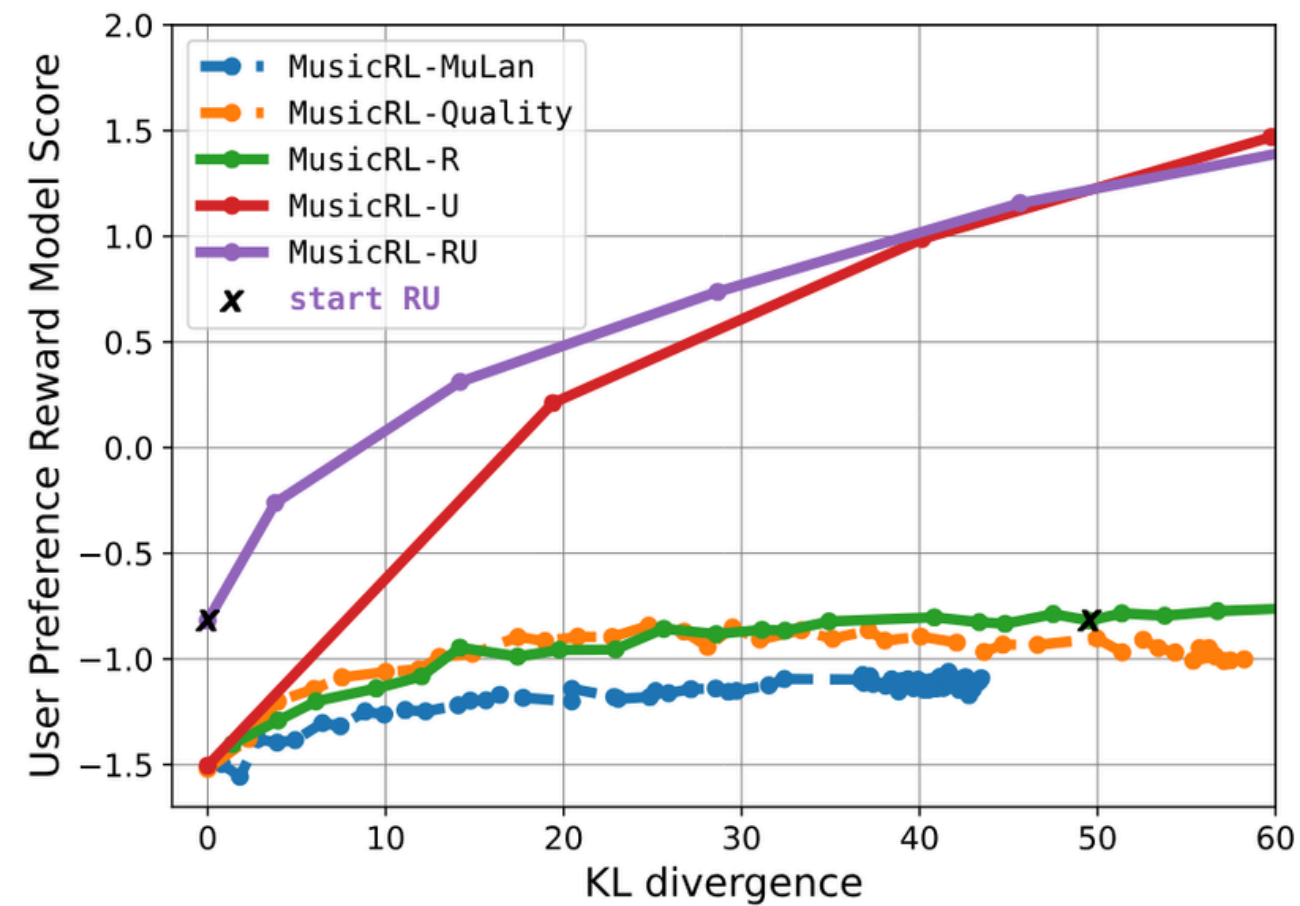


Porównanie wyników jakościowych i MuLan względem KL divergence

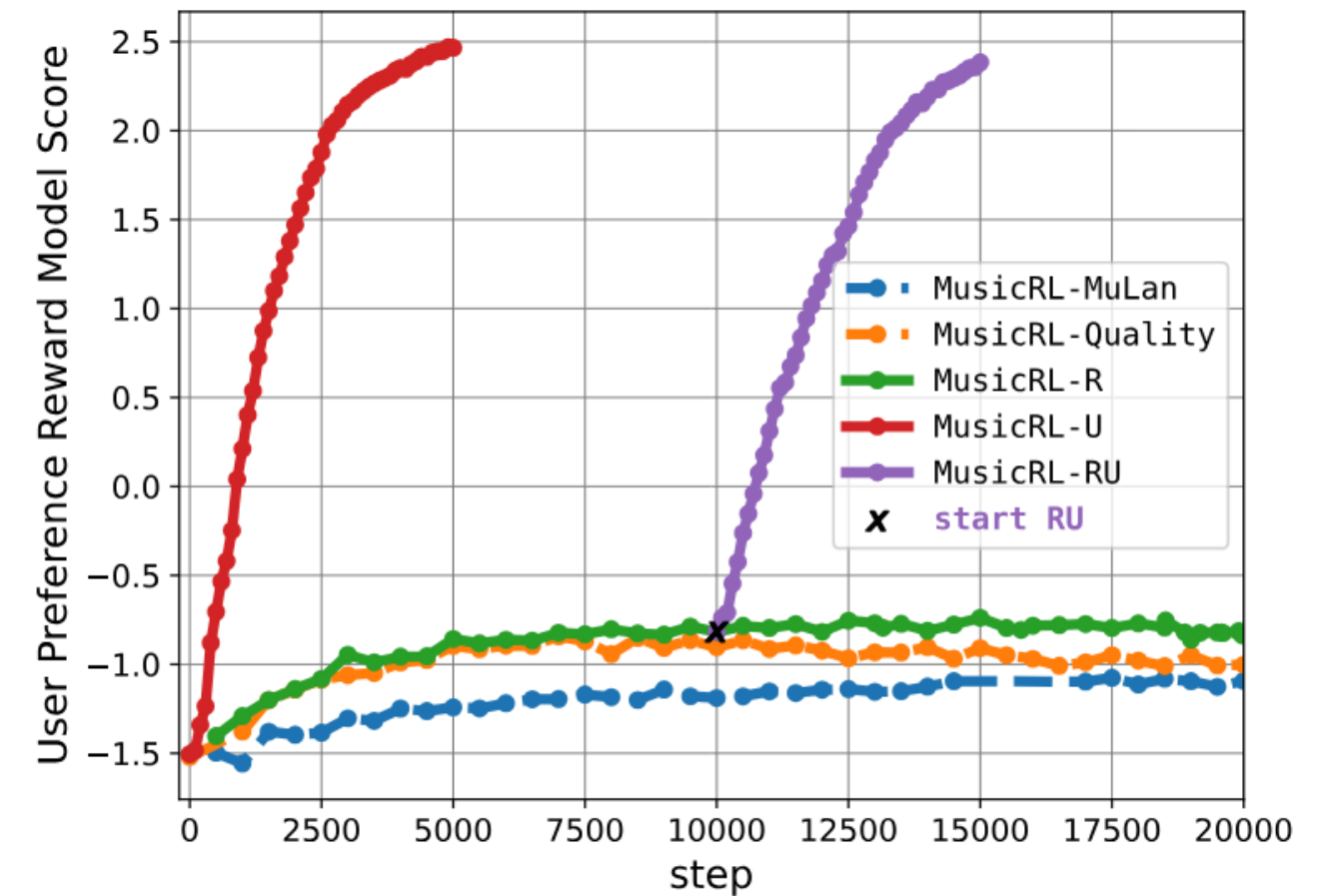
[1]

# WYNIKI ILOŚCIOWE

KL Divergence

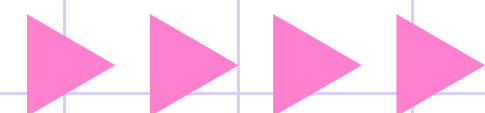


steps

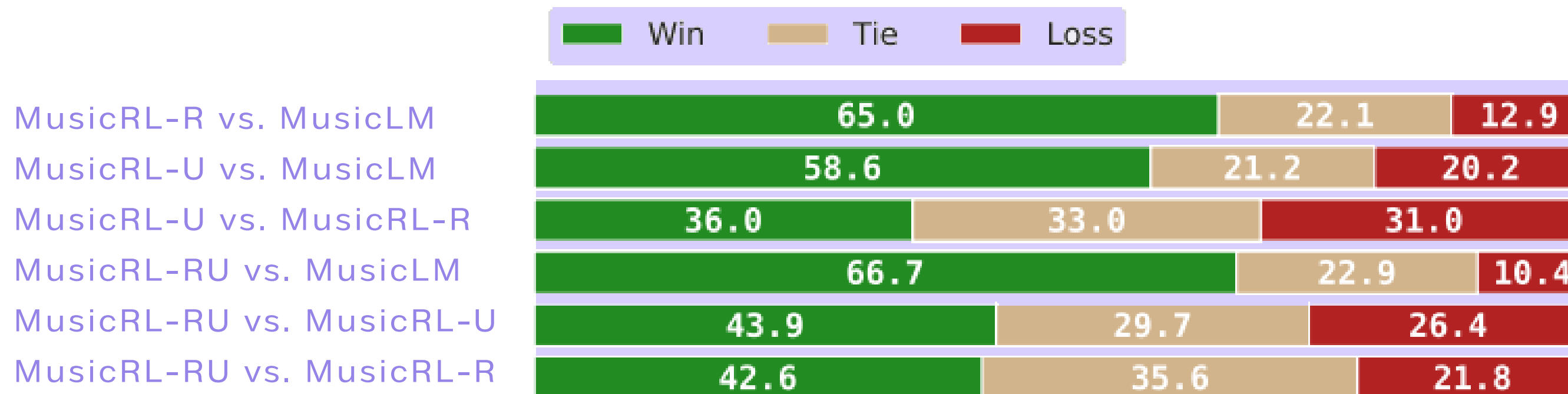
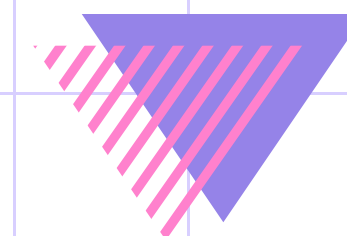


Porównanie preferencji względem KL divergence oraz kroków uczenia

[1]







# WYNIKI CAŁOKSZTAŁTU



Wyniki porównania preferencji użytkowników, model vs. model

[1]



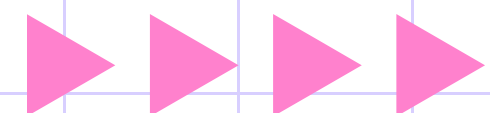
# PRÓBA ZROZUMIENIA LUDZKICH PREFERENCJI



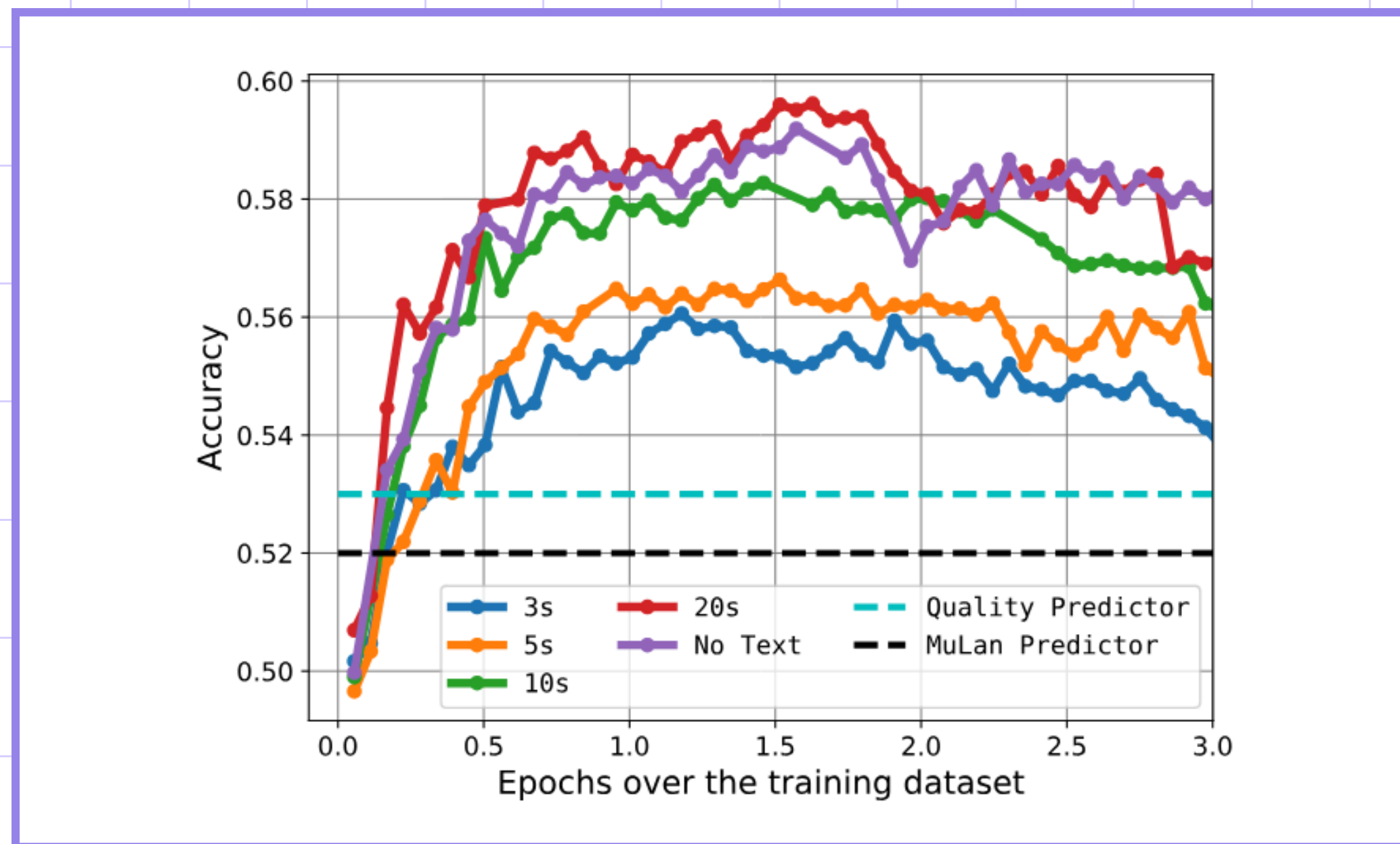
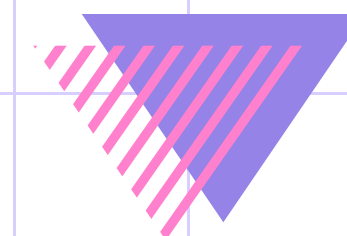
Na podstawie modelu **User Preference Reward Model**





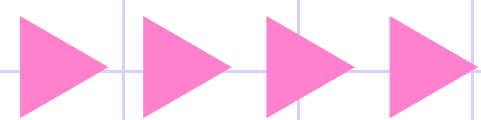


# USER PREFERENCE MODEL

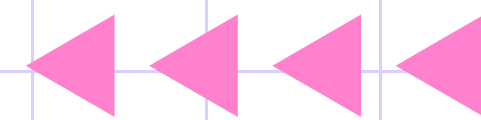


Porównanie zgodności z oceną użytkowników modelu user preference

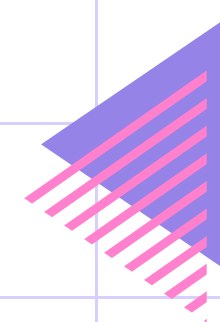
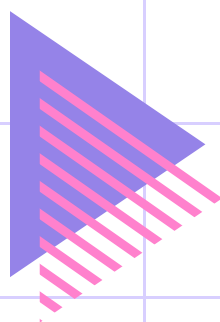
[1]



# TEXT ADHERENCE



Zgodność tekstu okazuje się być czynnikiem nieistotnym.  
Użytkownicy bardziej skupiają się na innych aspektach  
generowania muzyki



1

MuLan model - 51.6%

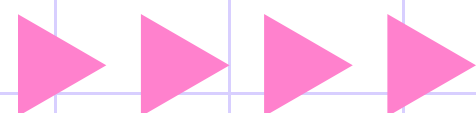
2

Ludzie nie wiedzą  
czego chcą

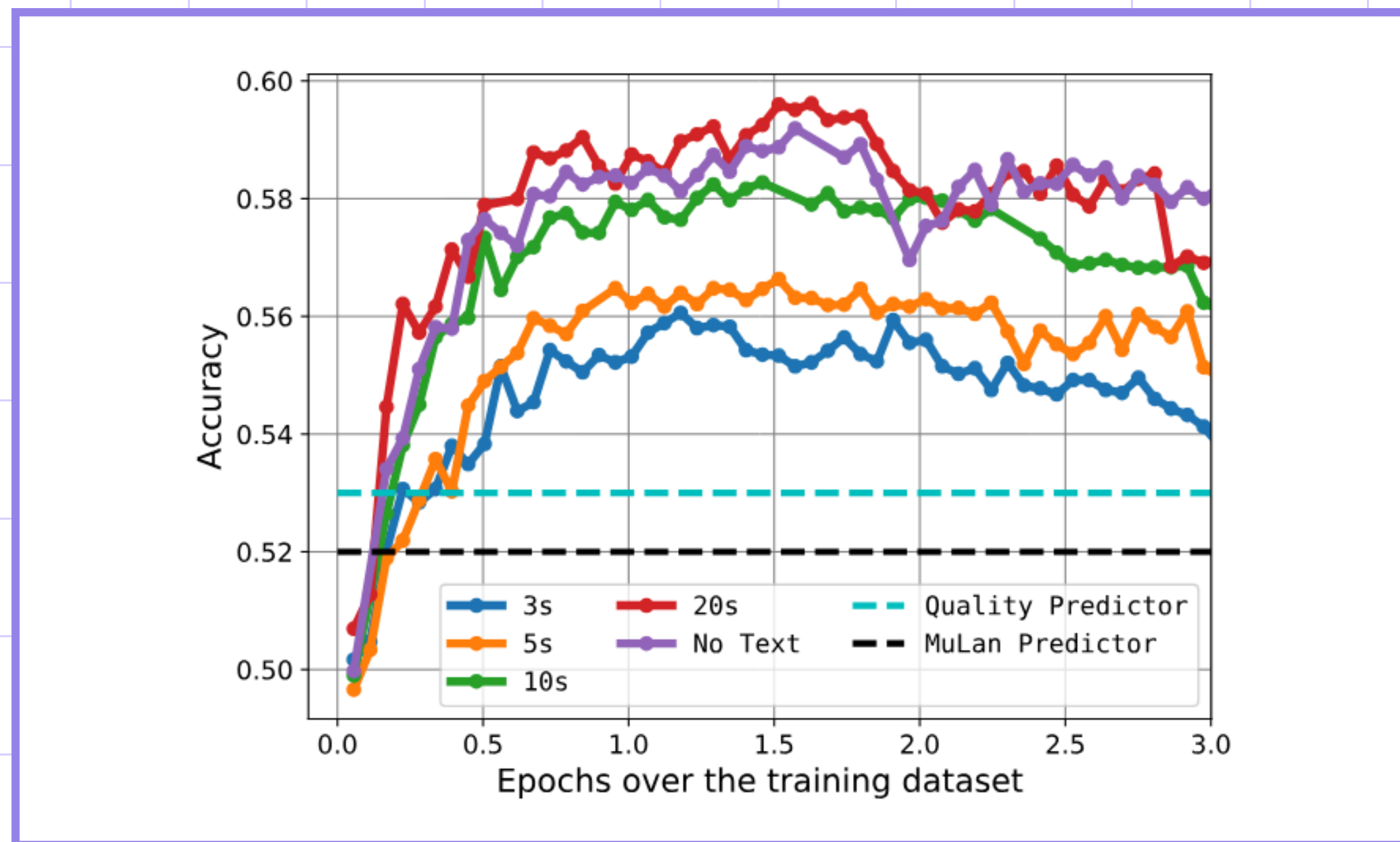
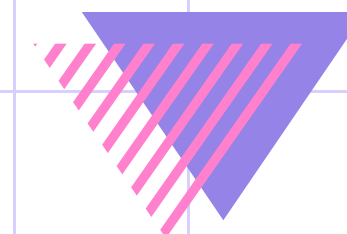
3

Nowe zjawisko - nie  
oczekujemy zbyt  
dużo lub winimy  
siebie za zły prompt





# AUDIO QUALITY



Porównanie zgodności z oceną użytkowników modelu user preference

[1]

# DZIĘKUJEMY ZA UWAGĘ

bibliografia:

- [1] Cideron, Geoffrey, et al. "MusicRL: Aligning Music Generation to Human Preferences." arXiv preprint arXiv:2402.04229 (2024).