CLaMP

Contrastive Language-Music Pre-training for Cross-Modal Symbolic Music Information Retrieval

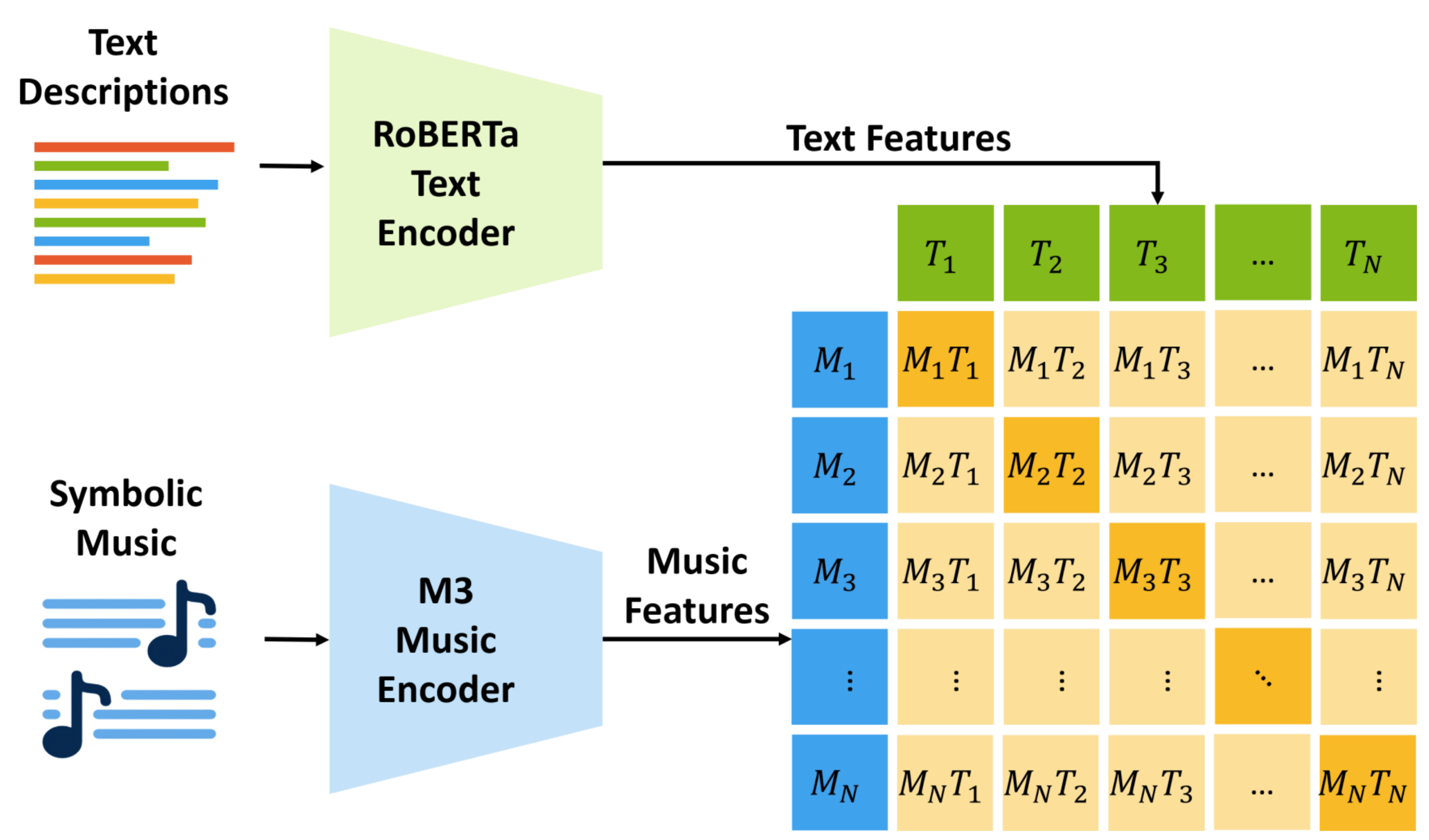
Shangda Wu[1] **Dingyao Yu**[2] **Xu Tan**[2] **Maosong Sun**[1,3]

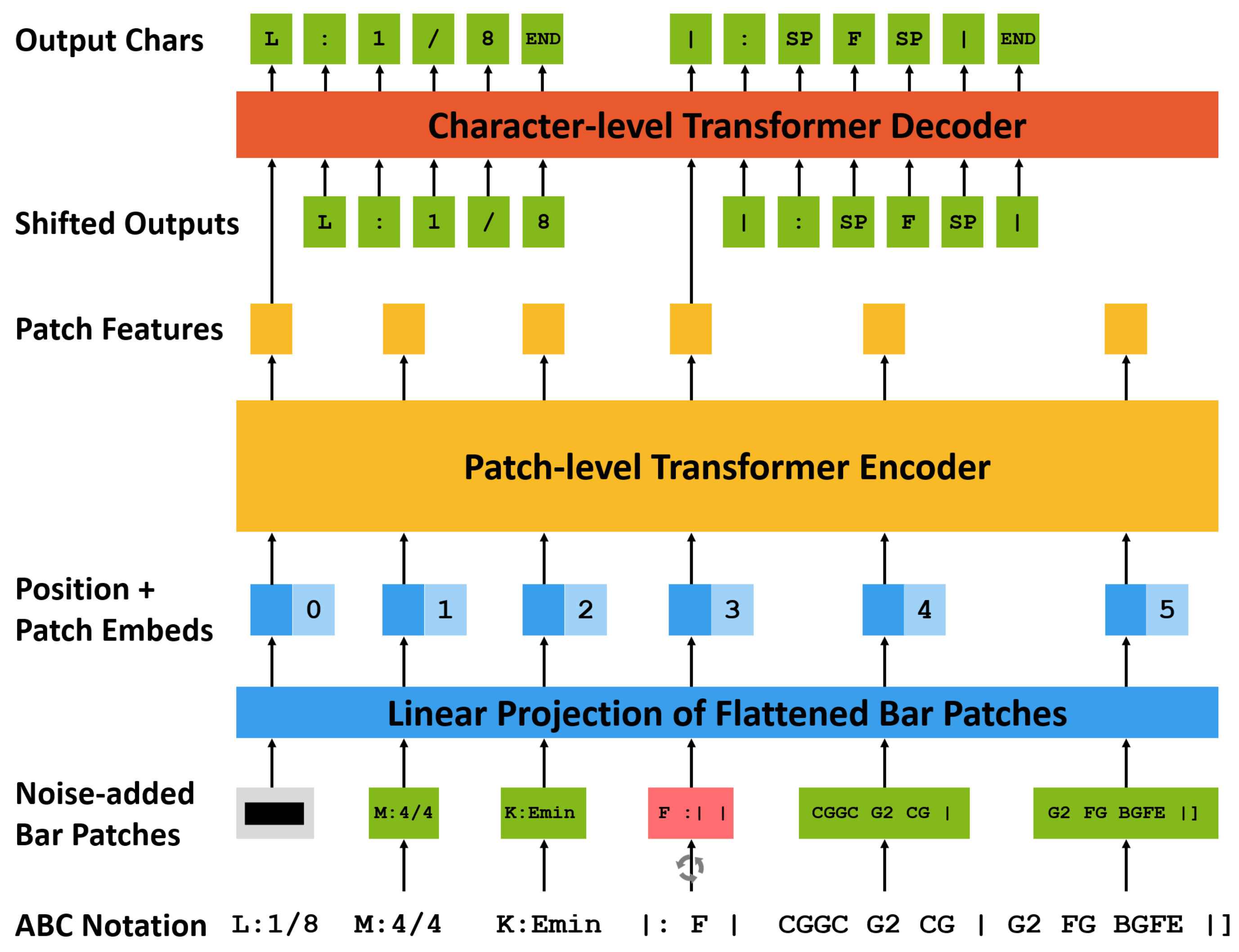
[1] Central Conservatory of Music
[2] Microsoft Research Asia
[3] Tsinghua University

GitHub Code　arXiv Paper　WikiMT Dataset

ISMIR 2023
Milan, Italy
Nov. 5-9, 2023
POLITECNICO MILANO 1863

I. Methodology


Text Descriptions
RoBERTa Text Encoder
Text Features
T_1
T_2
T_3
...
T_N
Symbolic Music
M3 Music Encoder
Music Features
M_1
M_2
M_3
⋮
M_N
M_1T_1 M_1T_2 M_1T_3 ... M_1T_N
M_2T_1 M_2T_2 M_2T_3 ... M_2T_N
M_3T_1 M_3T_2 M_3T_3 ... M_3T_N
M_NT_1 M_NT_2 M_NT_3 ... M_NT_N



Bridging Music with Text

- **Contrastive Learning** aligns music and text for cross-modal semantic understanding
- **Text Dropout** improves model robustness

Efficient Music Sequence Processing

- **Bar Patching** for efficient music representation based on ABC notation
- **Masked Music Model** pre-training objective for learning music features

Large Pre-training Dataset

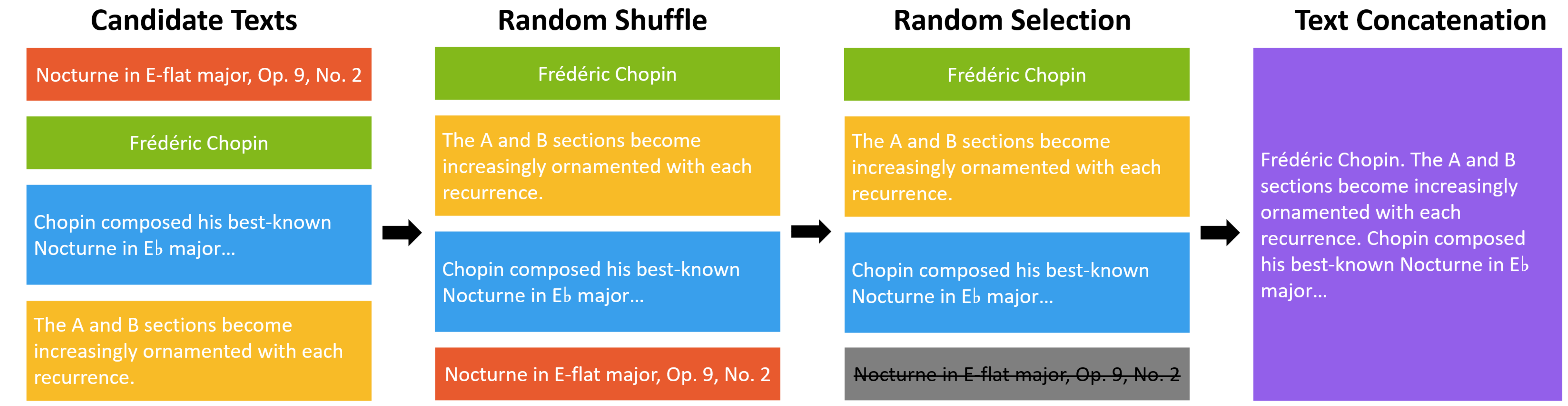
- **WebMusicText (WebMT)** consists of 1.4M music-text pairs (ABC notation), sourced from web

Bar Patching:

- Segments scores into **bars/headers**
- Converts each patch from 64x98 (tokens and vocabs) matrix to 768D embeddings
- Enables representation of **up to 512x64=32,768 tokens**

Masked Music Model (M3):

- Based on an **asymmetric encoder-decoder architecture**
- Self-supervised using bar patching
- **Introduces noise**, then reconstructs bar characters based on patch features



Candidate Texts

Nocturne in E-flat major, Op. 9, No. 2

Frédéric Chopin

Chopin composed his best-known Nocturne in E♭ major...

The A and B sections become increasingly ornamented with each recurrence.

Random Shuffle

Frédéric Chopin

The A and B sections become increasingly ornamented with each recurrence.

Chopin composed his best-known Nocturne in E♭ major...

Nocturne in E-flat major, Op. 9, No. 2

Random Selection

Frédéric Chopin

The A and B sections become increasingly ornamented with each recurrence.

Chopin composed his best-known Nocturne in E♭ major...

~~Nocturne in E-flat major, Op. 9, No. 2~~

Text Concatenation

Frédéric Chopin. The A and B sections become increasingly ornamented with each recurrence. Chopin composed his best-known Nocturne in E♭ major...

II. Experiments

Table 2. Semantic search performance of CLaMP on WikiMT (1010 music-text pairs) under different settings.

Setting	MRR	HR@1	HR@10	HR@100
S/512	**0.2561**	**0.1931**	**0.3693**	**0.7020**
S/1024	0.2016	0.1436	0.3109	0.6554
S/512 (w/o TD)	0.1841	0.1248	0.2911	0.6188
S/512 (w/o M3)	0.1262	0.0802	0.1960	0.5119
S/512 (w/o M3, BP)	0.0931	0.0525	0.1584	0.4426

Music: **1010 lead sheets** (ABC notation) from **Wikifonia** with natural language info removed

Text:

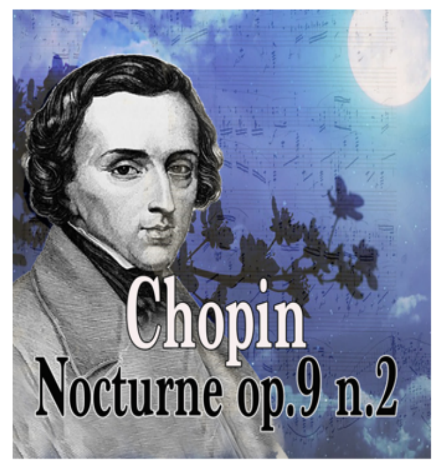
- **Title & Artist**: From scores
- **Description**: Sourced from **Wikipedia**, processed using BART-large
- **Genre**: **8 classes**, derived from Wikipedia by keyword matching

Table 3. Classification performance of different models on three datasets: WikiMT (1010 pieces, 8 genres), VGMIDI (204 pieces, 4 emotions), and Pianist8 (411 pieces, 8 composers).

Model	WikiMT		VGMIDI [11]		Pianist8 [12]	
	F1-macro	*Accuracy*	*F1-macro*	*Accuracy*	*F1-macro*	*Accuracy*
Linear Probe MusicBERT-S/1024	0.2401	**0.3507**	0.4662	0.5350	0.8047	0.8102
Linear Probe MusicBERT-B/1024	0.1746	0.3219	0.5127	0.5850	**0.8379**	**0.8413**
Zero-shot CLaMP-S/512	**0.2660**	0.3248	**0.5217**	**0.6176**	0.2180	0.2512
Zero-shot CLaMP-S/1024	0.2248	0.3406	0.4678	0.5049	0.1509	0.2390
Linear Probe M3-S/512	0.2832	0.3990	0.5991	0.6667	0.6773	0.6909
Linear Probe M3-S/1024	0.3079	0.4020	0.5966	0.6522	0.6844	0.6958
Linear Probe CLaMP-S/512	**0.3452**	0.4267	**0.6453**	**0.6866**	0.7067	0.7152
Linear Probe CLaMP-S/1024	0.3449	**0.4416**	0.6345	0.6720	**0.7271**	**0.7298**

III. Applications

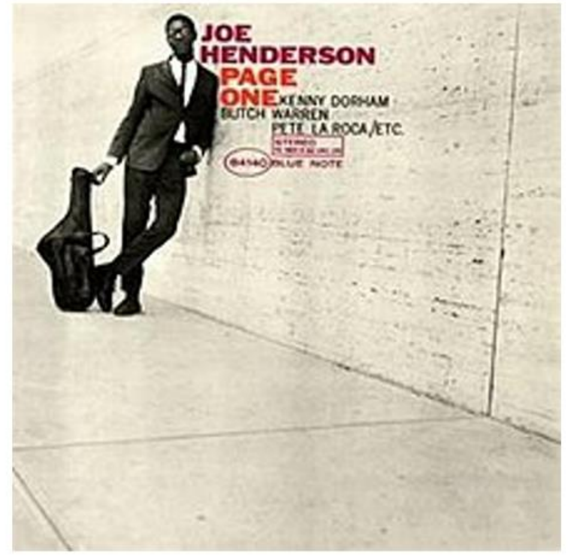
"It is ..."



Musical Form	*Mood*	*Composer*
"in rondo form"	"mysterious and haunting"	**"composed by Chopin"**
"in medley form"	"lively and energetic"	"composed by Liszt"
"in sonata-allegro form"	"dark and brooding"	"composed by Albinoni"
"in rounded binary form"	"joyful and optimistic"	"composed by Mozart"
"in theme and variations form"	**"calm and introspective"**	"composed by Prokofiev"

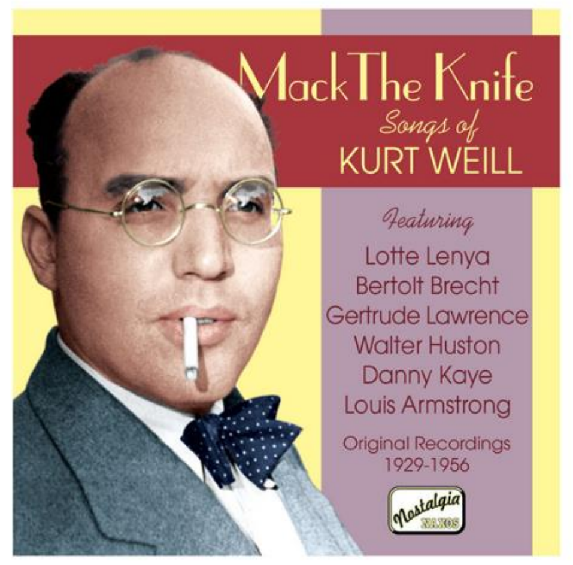
"Jazz standard in …"

"Minor key with a swing feel."



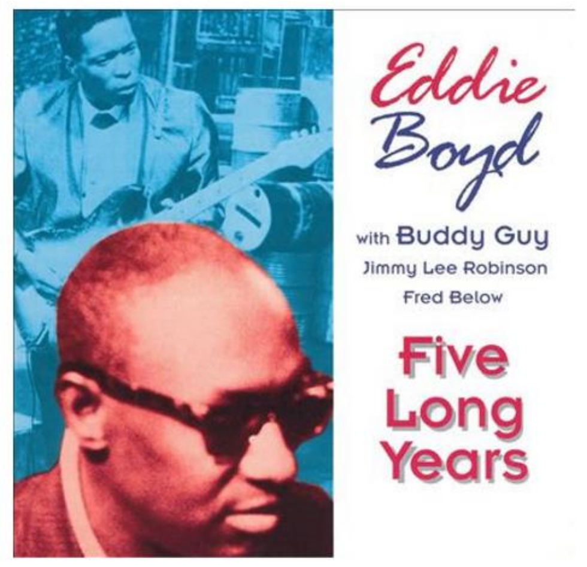
Blue Bossa

"Major key with a fast tempo."



Mack the Knife

"Blues form with a soulful melody."



Five Long Years