

A thick black L-shaped frame is positioned on the left and bottom sides of the slide, framing the central text.

# WYKORZYSTANIE SI W MUZYCE

Najpopularniejsze problemy i metody rozwiązywania

# Muzyka a sztuczna inteligencja

## Muzyka

- Jedna ze sztuk pięknych
- Sztuka organizacji struktur dźwiękowych w czasie
- Kreatywność w procesie twórczym i odtwórczym
- Reguł teoretyczne ukształtowane na przestrzeni lat

## Sztuczna Inteligencja

- Zdolność maszyn do wykazywania ludzkich umiejętności (np. uczenie się, kreatywność)

# Muzyka – podejście ścisłe

- Muzyka – sekwencja dźwięków
- Dźwięk – fala dźwiękowa/fala akustyczna



# REPREZENTACJA DANYCH MUZYCZNYCH



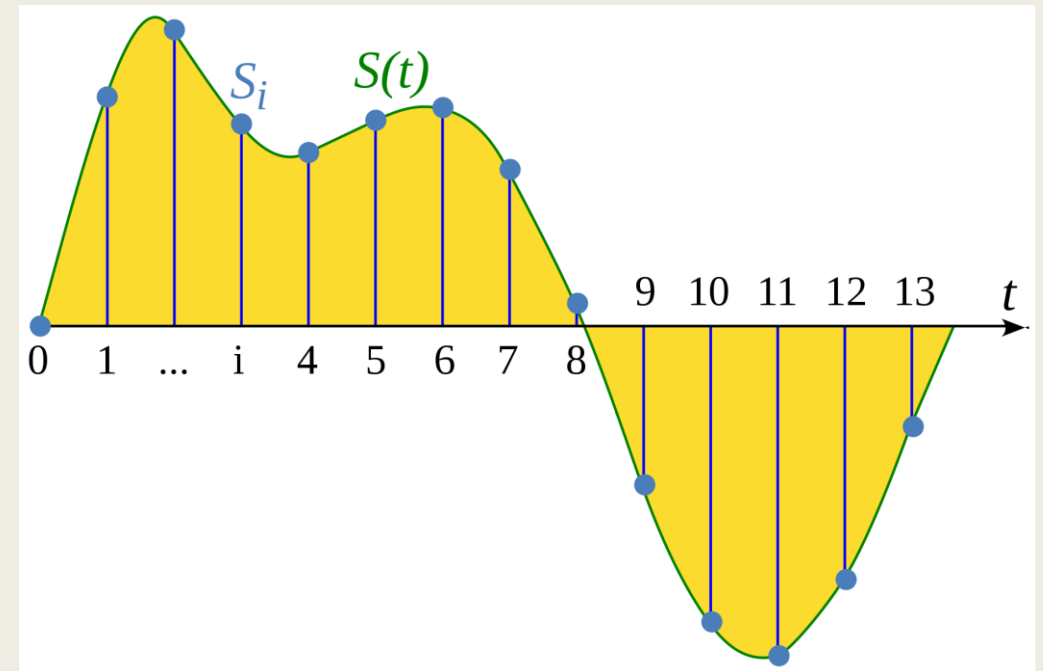
# Reprezentowanie danych muzycznych

Najpopularniejsze typy reprezentacji danych muzycznych:

- zapis fali dźwiękowej
- standard MIDI
- zapis nutowy

# Zapis fali dźwiękowej (format *wave*)

- Dyskretny zapis ciągłej fali
- Dokładność danych zależna od próbkowania



# Zapis fali dźwiękowej – wady i zalety

## Zalety



















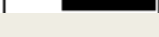




- Najbardziej naturalny
- Najłatwiejszy do uzyskania
- Najlepiej oddający rzeczywistość

## Wady

- Najtrudniejszy do pracy

# Standard MIDI (format *mid*)

- MIDI – *Musical Instrument Digital Interface* (ang. cyfrowy interfejs instrumentów muzycznych)
- Ustrukturyzowany zapis dźwięku
- Nie zawiera informacji o fali
- Zawiera informacje o poszczególnych dźwiękach

MIDI number		Note name	Keyboard	Frequency	
21	22	A0		27.500	
23		B0		30.868	29.135
24	25	C1		32.703	
26	27	D1		36.708	34.648
28		E1		41.203	38.891
29	30	F1		43.654	
31	32	G1		48.999	46.249
33	34	A1		55.000	51.913
35		B1		61.735	58.270
36	37	C2		65.406	
38	39	D2		73.416	69.296
40		E2		82.407	77.782
41	42	F2		87.307	
43	44	G2		97.999	92.499
45	46	A2		110.00	103.83
47		B2		123.47	116.54
48	49	C3		130.81	
50	51	D3		146.83	138.59
52		E3		164.81	155.56
53	54	F3		174.61	
55	56	G3		196.00	185.00
57	58	A3		220.00	207.65
		B3		246.94	233.08



# Standard MIDI – wady i zalety

## Zalety

- Bez zakłóceń – czysty dźwięk
- Łatwy do analizy

## Wady

- Cięższy do uzyskania

# Zapis nutowy (notacja nutowa)

- Język symboli
- Zawiera cechy:
  - dźwięków
  - rytmiki
  - melodii
  - harmonii
  - itp.

7. *Andantino.*  
*p dolce e semplice*

The image shows a musical score for piano, consisting of three systems of music. Each system has a treble and bass staff. The key signature is two sharps (F# and C#), and the time signature is 3/4. The tempo is marked 'Andantino.' and the dynamics are 'p dolce e semplice'. The score includes various musical notations such as notes, rests, and fingerings. The first system starts with a treble clef and a bass clef, with a key signature of two sharps and a time signature of 3/4. The music is written in a simple, elegant style, with a focus on melody and harmony. The second system continues the piece, and the third system concludes it with a double bar line. The score is presented in a clear, legible format, suitable for educational purposes.

# Zapis nutowy – wady i zalety

## Zalety

- Łatwa analiza zasad teoretycznych

## Wady

- Przystosowany tylko do konkretnych typów zadań
- Praca z tekstem, a nie dźwiękiem

# Właściwości wykorzystywane w analizie i uczeniu

Przykładowe właściwości dźwięków:

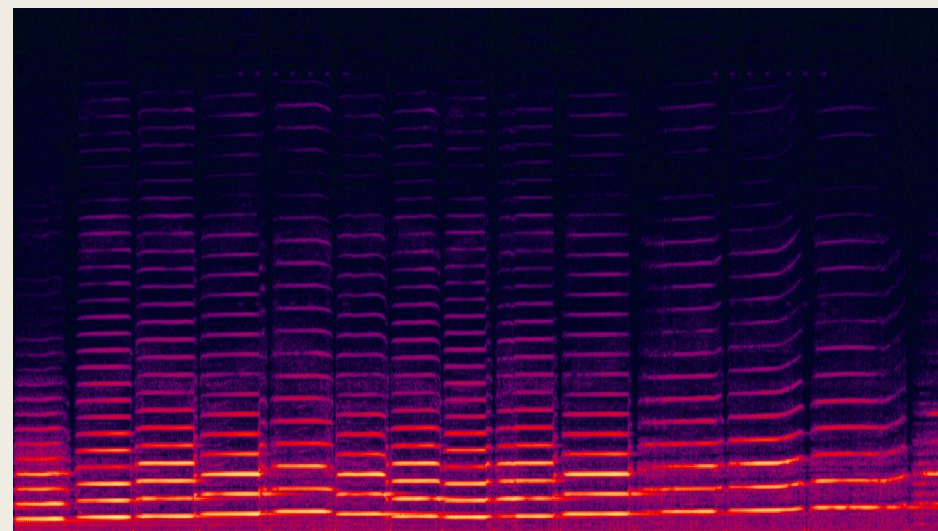
- wysokość
- czas rozpoczęcia
- czas trwania

Właściwości barwy dźwięków:

- intensywność
- jasność

# Właściwości wykorzystywane w analizie i uczeniu cd

- *Mel-frequency spectrogram*
- *Cepstral analysis/cepstral coefficients*
- *Zero-crossing rate*



# POPULARNE PROBLEMY W MUZYCE

i metody ich rozwiązywania



# Grupy problemów

## Klasyfikacja

- Gatunek muzyczny
- Kompozytor
- Wykonawca

## Generowanie

- Dźwięk
- Zapis nutowy

## Rekomendacja

- Rozpoznawanie emocji
- Systemy rekomendujące

KLASYFIKACJA





# Klasyfikacja gatunków, klasyfikacja wykonawców

Bergstra J., Casagrande N., Erhan D., Eck D., Keggl B.: *Aggregate features and ADABOOST for music classification*. Machine Learning (2006)

- Zbudowany algorytm do klasyfikacji gatunków i wykonawców
- Dane wejściowe: zapis fali dźwiękowej
- Wykorzystanie ADABOOST

Database	Magnatune		USPOP		Tzanetakis
	Genre	Artist	Genre	Artist	Genre
Number of training files	1005	1158	940	1158	800
Number of test files	510	642	474	653	200
Number of classes	10	77	6	77	10
Average song length		~ 200s			30s

# Klasyfikacja gatunków, klasyfikacja wykonawców

Bergstra J., Casagrande N., Erhan D., Eck D., Keggl B.: *Aggregate features and ADABOOST for music classification*. Machine Learning (2006)

Rank	Participant	Overall	Magnatune	USPOP
1	AB.TREE	82.34%	75.10%	86.92%
2	AB.STUMP	81.77%	74.71%	86.29%
3	Mandel & Ellis	78.81%	67.65%	85.65%
4	West, K.	75.29%	68.43%	78.90%
5	Lidy & Rauber [1]	75.27%	67.65%	79.75%
6	Pampalk, E.	75.14%	66.47%	80.38%
7	Lidy & Rauber [2]	74.78%	67.65%	78.48%
8	Lidy & Rauber [3]	74.58%	67.25%	78.27%
9	Scaringella, N.	73.11%	66.14%	75.74%
10	Ahrendt, P.	71.55%	60.98%	78.48%
11	Burred, J.	62.63%	54.12%	66.03%
12	Soares, V.	60.98%	49.41%	66.67%
13	Tzanetakis, G.	60.72%	55.49%	63.29%

Rank	Participant	Mean performance	Magnatune	USPOP
1	Mandel & Ellis	72.45%	76.60%	68.30%
2	AB.STUMP	68.57%	77.26%	59.88%
3	AB.TREE	66.71%	74.45%	58.96%
4	Pampalk, E.	61.28%	66.36%	56.20%
5	West & Lamere	47.24%	53.43%	41.04%
6	Tzanetakis, G.	42.05%	55.45%	28.64%
7	Logan, B.	25.95%	37.07%	14.83%

# Klasyfikacja kompozytorów

Herremans D., Kenneth S., Martens D.: *Classification and Generation of Composer-Specific Music Using Global Feature Models and Variable Neighborhood Search*. Computer Music Journal (2015)

- Dane wejściowe – plik MIDI
- Stworzone 4 modele:
  - zbiór reguł *if-then* (RIPPER)
  - drzewo decyzyjne (C4.5)
  - regresja logistyczna (LogitBoost)
  - SVM

<i>Variable</i>	<i>Feature Description</i>
$x_1$	Chromatic motion frequency: fraction of melodic intervals corresponding to a semitone
$x_2$	Frequency of melodic fifths
$x_3$	Frequency of melodic octaves
$x_4$	Frequency of melodic thirds
$x_5$	Most common melodic interval prevalence
$x_6$	Most common pitch prevalence
$x_7$	Most common pitch class prevalence
$x_8$	Relative strength of most common intervals: fraction of intervals belonging to the most common and second most common melodic intervals
$x_9$	Relative strength of top pitch classes
$x_{10}$	Relative strength of top pitches
$x_{11}$	Repeated notes: fraction of notes that are repeated melodically
$x_{12}$	Stepwise motion frequency

*Pitch refers to an absolute pitch, e.g., C in the 7th octave. Pitch class refers to a note without the octave, e.g., C.*

# Klasyfikacja kompozytorów

Herremans D., Kenneth S., Martens D.: *Classification and Generation of Composer-Specific Music Using Global Feature Models and Variable Neighborhood Search*. Computer Music Journal (2015)

Table 4. Confusion Matrix for RIPPER

<i>a</i>	<i>b</i>	<i>c</i>	<i>Classified as</i>
145	42	67	a (Haydn)
41	<b>110</b>	45	b (Beethoven)
25	17	553	c (Bach)

Table 6. Confusion Matrix for C4.5

<i>a</i>	<i>b</i>	<i>c</i>	<i>classified as</i>
174	31	49	a (Haydn)
63	<b>106</b>	27	b (Beethoven)
40	14	541	c (Bach)

Table 9. Confusion Matrix for Logistic Regression

<i>a</i>	<i>b</i>	<i>c</i>	<i>classified as</i>
190	30	34	a (Haydn)
57	<b>119</b>	20	b (Beethoven)
25	15	555	c (Bach)

Table 10. Confusion Matrix for Support Vector Machine

<i>a</i>	<i>b</i>	<i>c</i>	<i>classified as</i>
204	26	24	a (Haydn)
49	<b>127</b>	20	b (Beethoven)
22	10	563	c (Bach)

<i>Method</i>	<i>Accuracy (%)</i>	<i>WAUC (%)</i>
RIPPER rule set	77 (3.82)	82 (3.51)
C4.5 decision tree	79 (2.97)	88 (2.55)
Logistic regression	83 (3.27)	94 (2.12)
Support vector machines	<b>86</b> (4.15)	<b>96</b> (2.12)

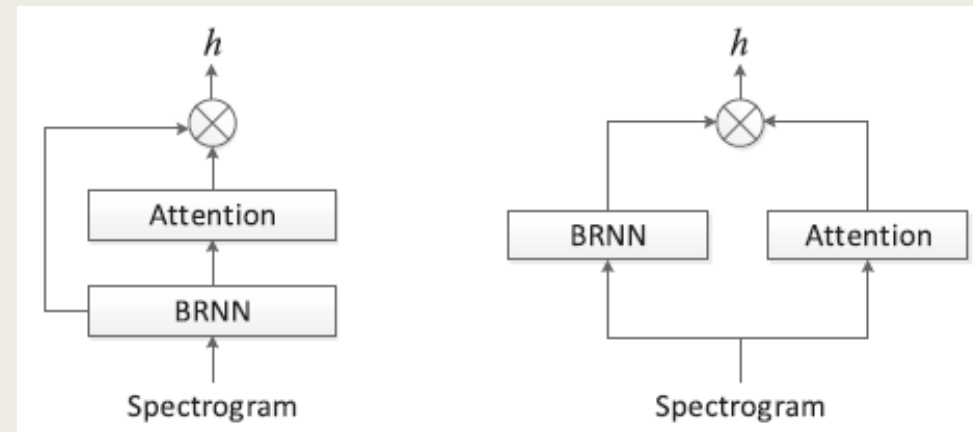
```

if (Most Common Melodic Interval Prevalence) ≤ 0.2688
  and (Melodic Octaves Frequency ≥ 0.06399) then
  Composer = BE
else if (Most Common Melodic Interval Prevalence ≤ 0.2823)
  and (Most Common Pitch Prevalence ≤ 0.07051)
  and (Melodic Octaves Frequency ≥ 0.02489)
  and (Relative Strength of Top Pitches ≤ 0.9754) then
  Composer = BE
else if (Most Common Melodic Interval Prevalence ≤ 0.328)
  and (Repeated Notes Frequency ≥ 0.07592)
  and (Most Common Pitch Prevalence ≤ 0.1076) then
  Composer = HA
else if (Stepwise Motion Frequency ≤ 0.5732)
  and (Chromatic Motion Frequency ≥ 0.1166)
  and (Repeated Notes Frequency ≥ 0.3007) then
  Composer = HA
else
  Composer = BA
end if
    
```

# Klasyfikacja gatunków

Yu Y., Luo S., Liu S., Qiau H., Liu Y., Feng L.: *Deep attention based music genre classification*. Neurocomputing (2020)

- Dane wejściowe – spektrogram
- Schemat kodowanie-dekodowanie
- Rozpoznawanie oparte o:
  - *głębokie sieci neuronowe*
  - *mechanizm uwagi* (szeregowy, równoległy)

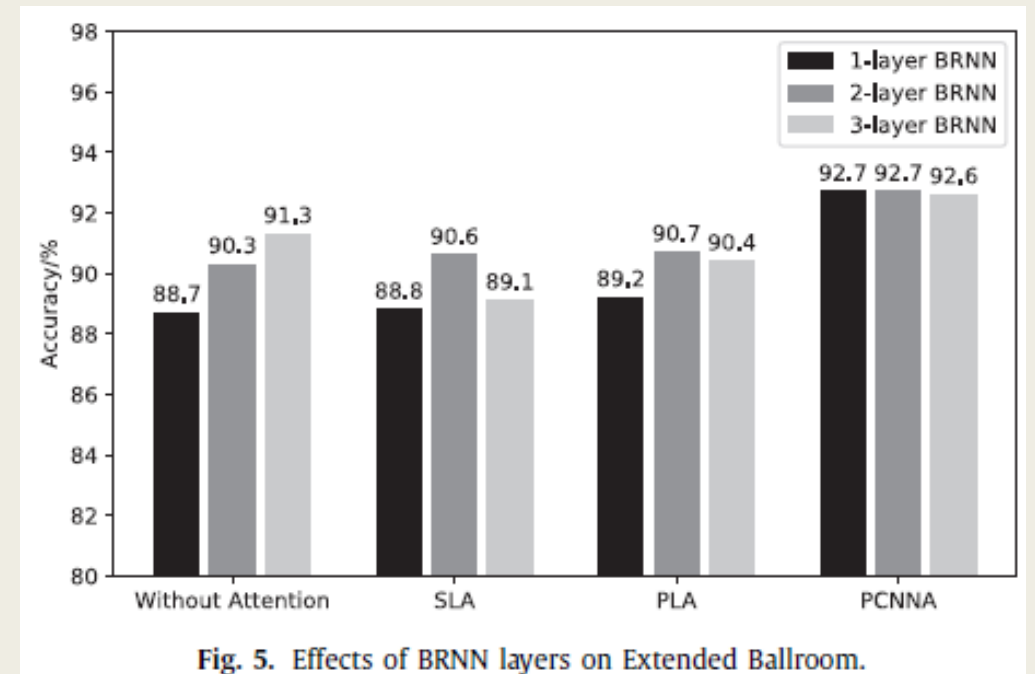
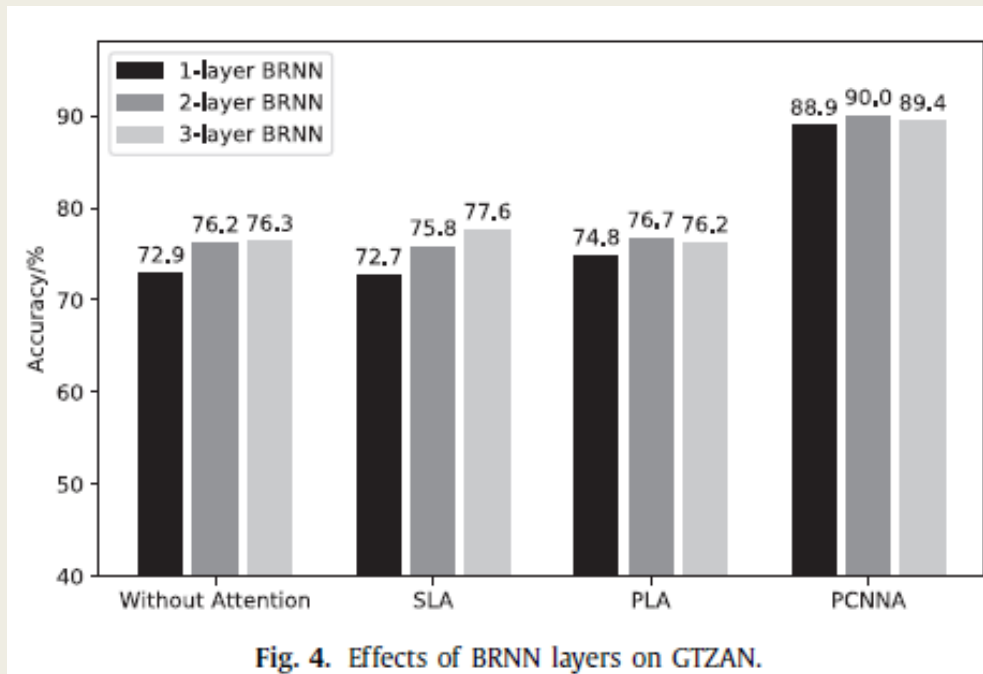


Genre distribution of Extended Ballroom.

Genre	Track number	Genre	Track number
Chacha	455	Wcswing	23
Jive	350	Slowwaltz	65
Quickstep	497	Salsa	47
Rumba	470	Pasodoble	53
Samba	468	Foxtrot	507
Tango	464	Waltz	529
Viennesewaltz	252	<b>Total</b>	<b>4180</b>

# Klasyfikacja gatunków

Yu Y., Luo S., Liu S., Qiau H., Liu Y., Feng L.: *Deep attention based music genre classification*. Neurocomputing (2020)



# Klasyfikacja muzyki ekspresyjnej

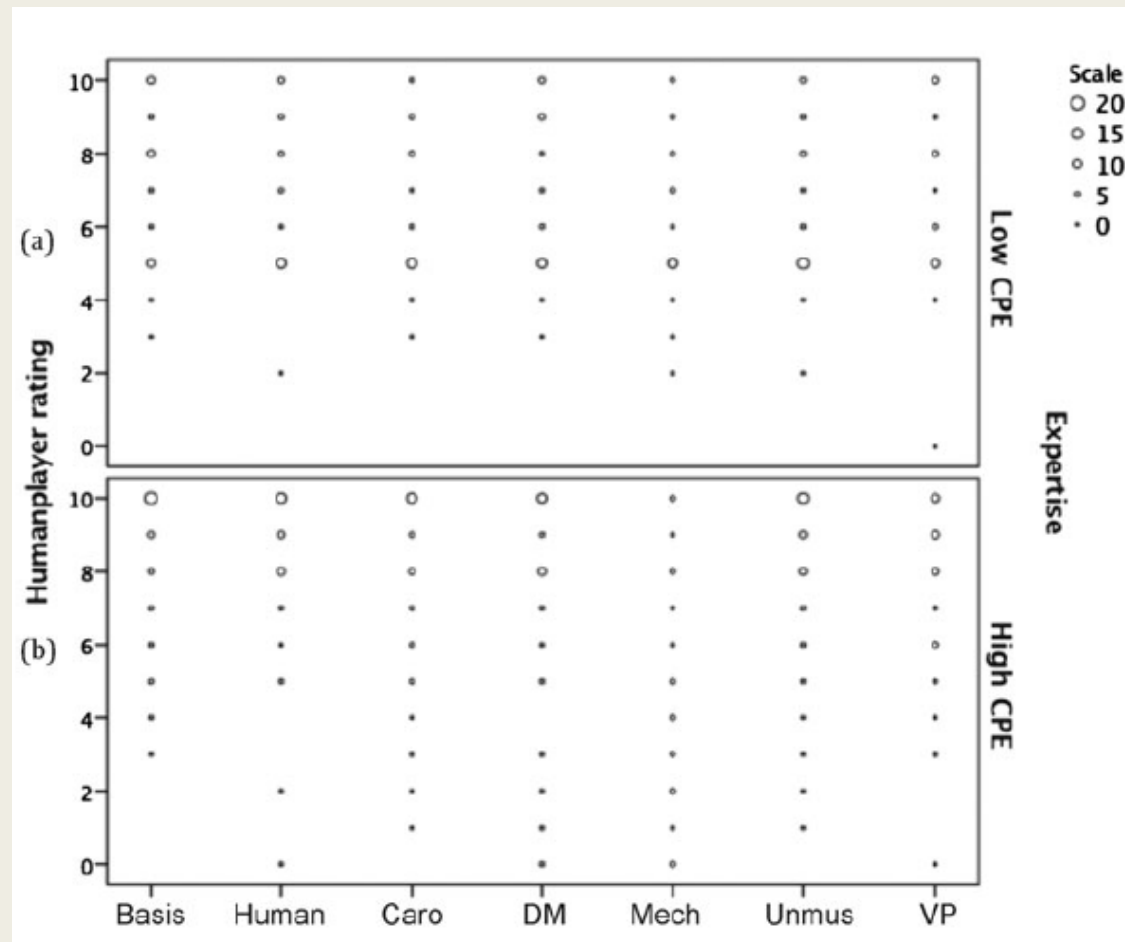
Schubert E., Canazza S., De Poli G., Rodà A.: *Algorithms can Mimic Human Piano Performance: The Deep Blues of Music*. *Journal of New Music Research* (2017)

- 172 uczestników
- Uczestnicy mający, jak i niemający związku z muzyką

Abbreviation	Performer information
1. Basis	Basis mixer (Grachten & Widmer, 2012) for dynamics, tempo and articulation [duration 84s]
2. Human	A high skilled, internationally renown concert pianist [duration 76s]
3. Caro	CaRo 2.0, developed by the Sound and Music Computing group, Department of Information Engineering, University of Padova (Italy) (Canazza, De Poli, Drioli, Rodà, & Vidolin, 2004; Canazza, De Poli, & Rodà, 2015) [duration 80s]
4. DM	DirectorMusices (accent-based formulation), developed by the Music Acoustics Group, KTH Royal (Bisesi, Pamcutt, & Friberg, 2011) [duration 100s]
5. Mech	Deadpan, mechanical MIDI score [duration 73s]
6. Unmus	Performed by CaRo 2.0 with parameters tuned to opposite values (e.g. $-0.5$ instead of $0.5$ ), assumed to sound unmusical (van Oosten, 1993). The detuned parameters control agogic (accelerando-ritardando), dynamic (crescendo-diminuendo), and articulation (legato-staccato)
7. VP	VirtualPhilharmony, developed by the Katayose Laboratory, Department of Human and Systems Interaction, Kwansai Gakuin University (Japan) (Baba, Hashida, & Katayose, 2010; Katayose & Okudaira, 2004) [duration 84s]

# Klasyfikacja muzyki ekspresyjnej

Schubert E., Canazza S., De Poli G., Roda A.: *Algorithms can Mimic Human Piano Performance: The Deep Blues of Music*. Journal of New Music Reaserch (2017)





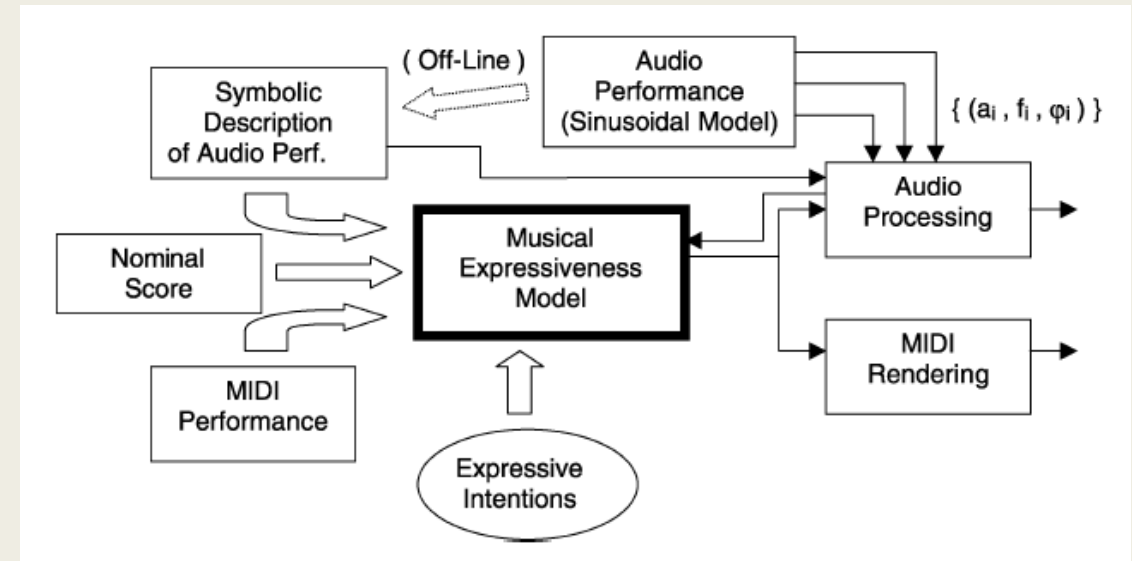
GENEROWANIE



# Generowanie muzyki ekspresyjnej

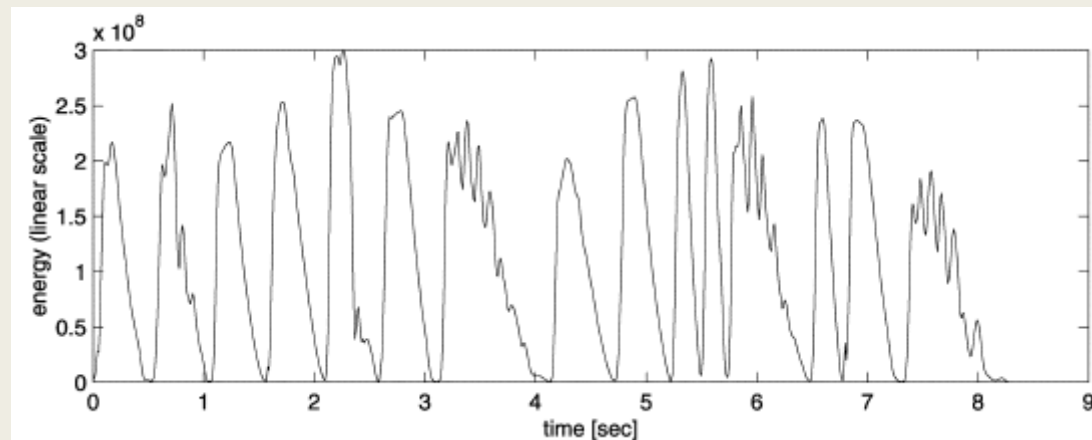
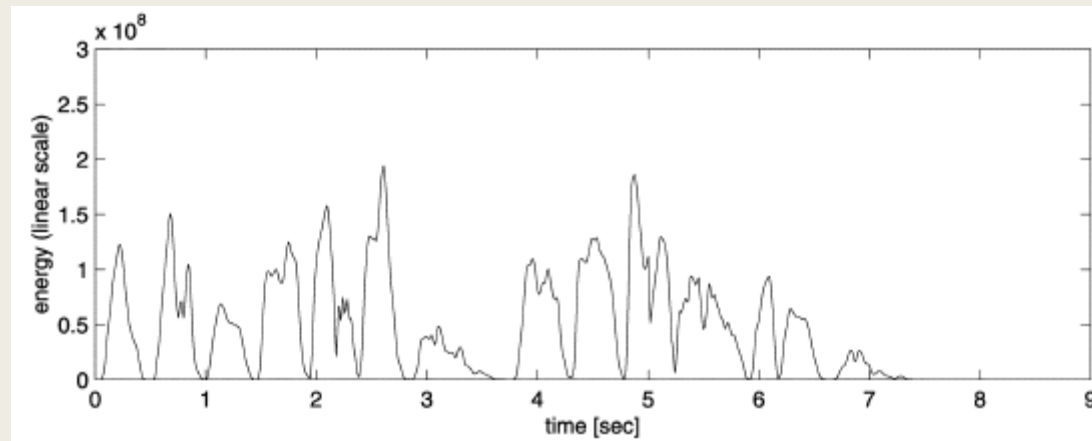
Canazza S., De Poli G., Drioli C., Roda A., Vidolin A.: *Modeling and Control of Expressiveness in Music Performance*. Proceedings of the IEEE (2004)

- Dane wejściowe – MIDI i zapis fali
- Modyfikacja ekspresyjności poprzez:
  - przesunięcie zdarzenia
  - rozciągnięcie/kompresję zdarzenia
- Możliwe intencje: jasne, ciemne, twarde, miękkie, ciężkie, lekkie
- Regresja liniowa



# Generowanie muzyki ekspresyjnej

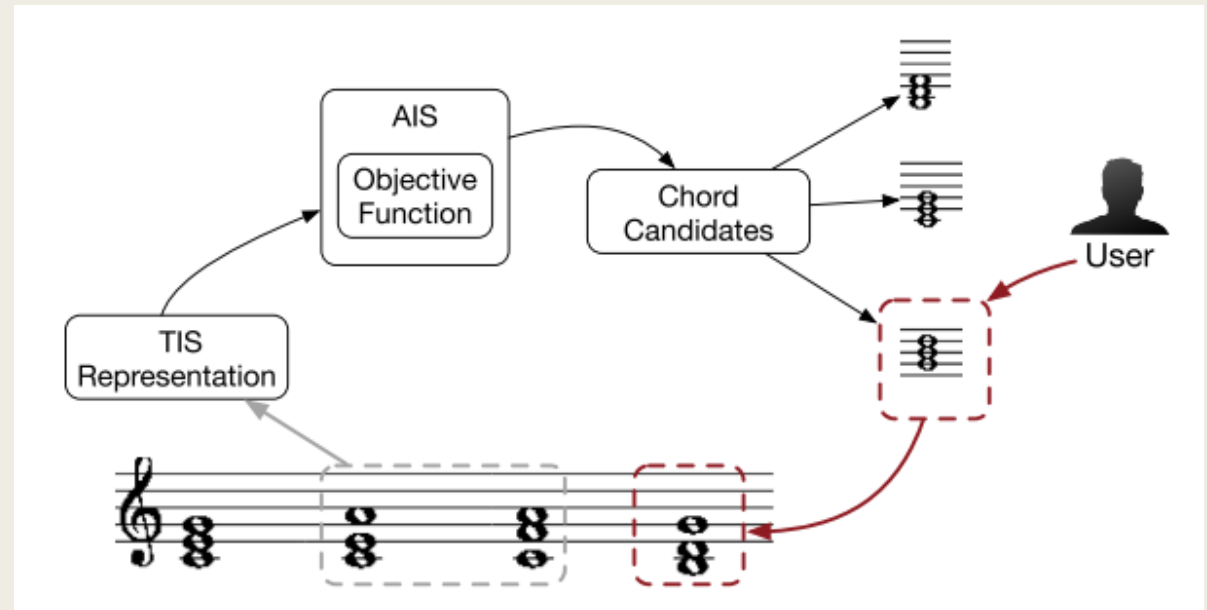
Canazza S., De Poli G., Drioli C., Roda A., Vidolin A.: *Modeling and Control of Expressiveness in Music Performance*. Proceedings of the IEEE (2004)



# Generowanie akordów

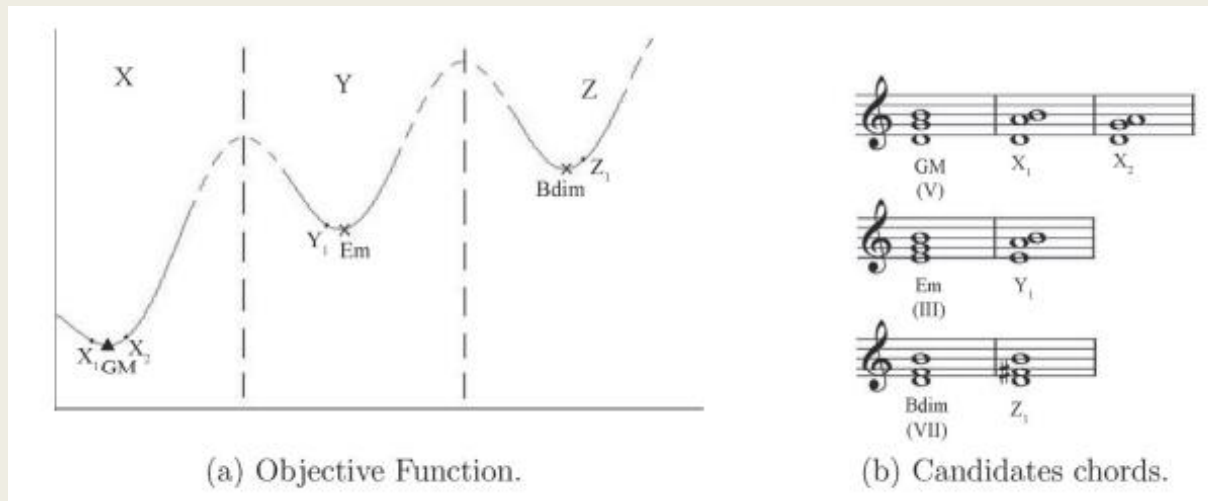
Navarro-Cáceres M., Caetano M., Bernardes G., de Castro L.: *ChordAIS: An assistive system for the generation of chord progressions with an artificial immune system*. Swarm and Evolutionary Computation (2019)

- Problem – generowanie progresji
- TIS – *Tonal Interval Space*
- AIS – *Artificial Immune System*
- Funkcja celu – pokrewieństwo kluczy i funkcji harmoniczných, dysonans



# Generowanie akordów

Navarro-Cáceres M., Caetano M., Bernardes G., de Castro L.: *ChordAIS: An assistive system for the generation of chord progressions with an artificial immune system*. Swarm and Evolutionary Computation (2019)



Two staves of musical notation showing a sequence of chords. The first staff contains 8 measures of chords. The second staff starts at measure 9 and contains 8 measures of chords, ending with a double bar line.

# Generowanie melodii

Wu J., Liu X., Hu X., Jun Z.: *PopMNet: Generating structured pop music melodies using neural networks*. Artificial Intelligence (2020)

- Generowanie melodii na podstawie muzyki popularnej
- Wejście: standard MusicXML
- Dwa etapy: generowanie struktury (sieć splotowa), generowanie melodii (sieć rekurencyjna)
- Generowanie struktury:
  - *generator*
  - *krytyk*
- Generowanie melodii:
  - *stworzona struktura*
  - *sekwencja akordów*

# Generowanie melodii

Wu J., Liu X., Hu X., Jun Z.: *PopMNet: Generating structured pop music melodies using neural networks*. Artificial Intelligence (2020)

Time signature: C

Chord: C

Bar: Gsus4, Am

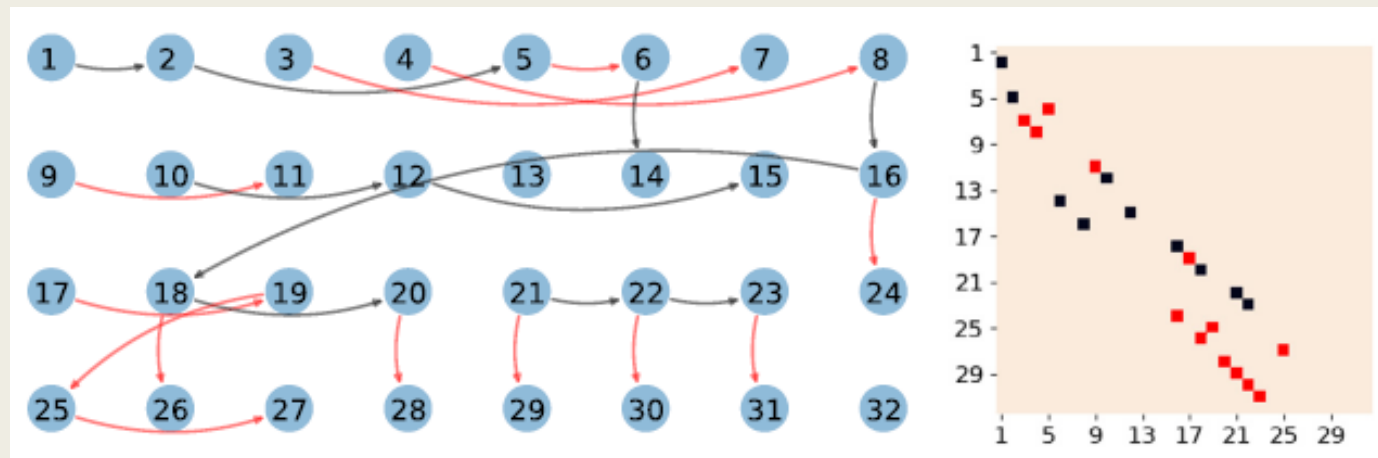
Beat: F

Chord progression: C, Gsus2, Am7, F

Chord progression: Am, Dm, G, C, G/B, Am, Dm, G, C, G/B, Am, Dm, G, Am, Bb, F/G

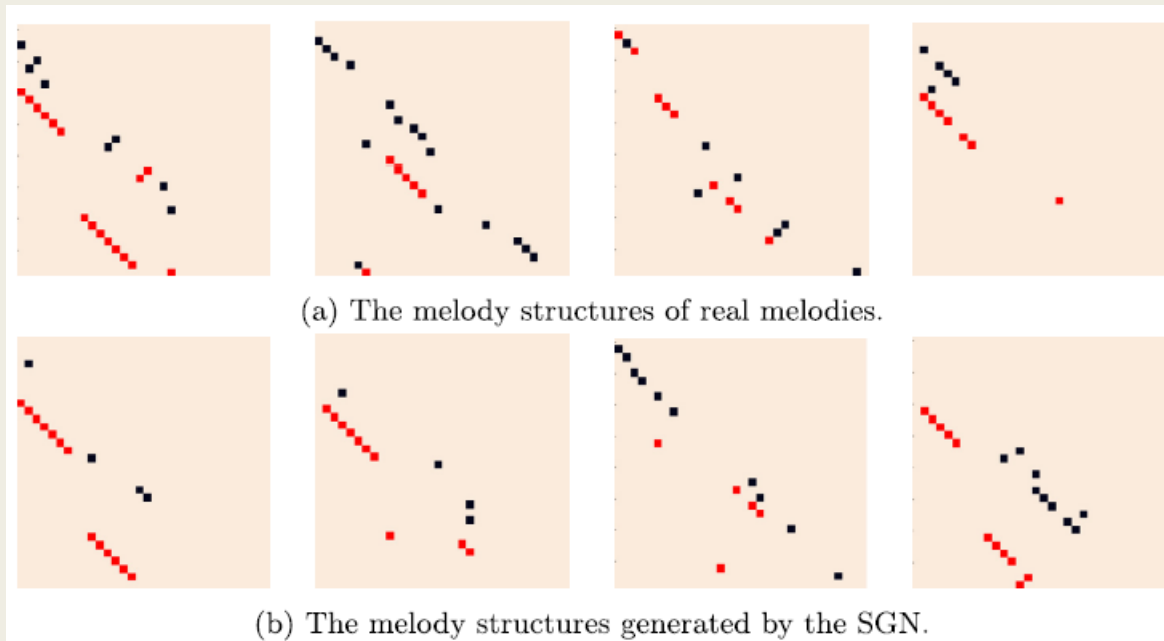
Chord progression: C, G/B, Am, Em, F, C/E, Dm7, F/G

Chord progression: C, G/B, Am, Em, F, C/E, Dm7, F/G



# Generowanie melodii

Wu J., Liu X., Hu X., Jun Z.: *PopMNet: Generating structured pop music melodies using neural networks*. Artificial Intelligence (2020)



The percentages of relations in the melodies in the dataset and the melodies generated by different models.

	Repetition	Rhythmic sequence	No relation
Real data	29.06%	32.64%	38.30%
AttentionRNN	0.42%	24.73%	74.85%
LookbackRNN	9.13 %	37.78%	53.09%
MidiNet	18.75%	18.752%	62.29%
Music Transformer	35.64%	25.15%	39.21%
PopMNet	27.05%	28.91%	44.04%
PopMNet-Real	31.45%	31.33%	37.22%
PopMNet-NC	26.28%	29.45%	44.27%

Human evaluation scores of melodies in Experiment 2 (mean±std, all 18 subjects).

	Pleasure	Reality	Smooth	Integrity
Music Transformer	2.98 ± 1.08	3.01 ± 1.16	3.10 ± 1.97	3.02 ± 1.23
PopMNet	3.41 ± 0.83	3.29 ± 0.82	3.39 ± 0.94	3.31 ± 0.94
Human	3.92 ± 0.92	3.89 ± 0.94	3.85 ± 0.91	3.98 ± 0.89



# Generowanie harmonizacji

De Prisco R., Zaccagnino G., Zaccagnino R. : *EvoComposer: An Evolutionary Algorithm for 4-Voice Music Composition*. Evolutionary Computation (2019)

- Generowanie harmonizacji dla niefigurowanego basu
- Wejście: zapis nutowy
- Algorytm ewolucyjny
- Dwa podproblemy: generowanie harmonii, generowanie akordów

The image displays a musical score illustrating the process of harmonicization. It is divided into two parts, a) and b). Part a) shows a single bass line in a bass clef with a key signature of one flat and a common time signature. The notes are: G2, A2, B2, C3, D3, E3, F3, G3, A3, B3, C4, D4, E4, F4, G4, A4, B4, C5. A white arrow points downwards from this line to part b). Part b) shows a four-voice setting in a grand staff with a key signature of one flat and a common time signature. The voices are labeled: Soprano (treble clef), Alto (treble clef), Tenor (bass clef), and Bass (bass clef). The Soprano line contains the notes: G4, A4, B4, C5, D5, E5, F5, G5, A5, B5, C6, D6, E6, F6, G6, A6, B6, C7. The Alto line contains the notes: G4, A4, B4, C5, D5, E5, F5, G5, A5, B5, C6, D6, E6, F6, G6, A6, B6, C7. The Tenor line contains the notes: G3, A3, B3, C4, D4, E4, F4, G4, A4, B4, C5, D5, E5, F5, G5, A5, B5, C6. The Bass line contains the notes: G2, A2, B2, C3, D3, E3, F3, G3, A3, B3, C4, D4, E4, F4, G4, A4, B4, C5.

# Generowanie harmonizacji

De Prisco R., Zaccagnino G., Zaccagnino R. : *EvoComposer: An Evolutionary Algorithm for 4-Voice Music Composition*. Evolutionary Computation (2019)

Composer	C1	C2	T1	T2	T3	E1	E2	Mean
<i>EvoBassComposer</i>	3.21	3.37	3.77	3.73	4.17	3.79	3.17	3.60
<i>EvoComposer</i>	5.19	5.41	5.74	5.72	6.12	5.72	5.13	5.58
<i>TabuSearch</i>	4.20	4.36	4.78	4.74	5.16	4.79	4.21	4.60
<i>SimulatedAnnealing</i>	3.80	3.99	4.36	4.33	4.77	4.38	3.77	4.20
<i>ParticleSwarm</i>	3.41	3.56	3.97	3.93	4.35	3.99	3.42	3.80

The image shows a musical score for four voices: Soprano, Alto, Tenor, and Bass. The music is in 3/4 time with a key signature of one sharp (F#). Below the staves, guitar chords and Roman numerals are provided for each measure. The chords are: D Bm7 A G D D A7 Em D G A7 D A G A7 D Bm7 D G Bm G D. The Roman numerals are: I vi V IV I I V ii I IV V I V IV V I vi I IV vi V I.

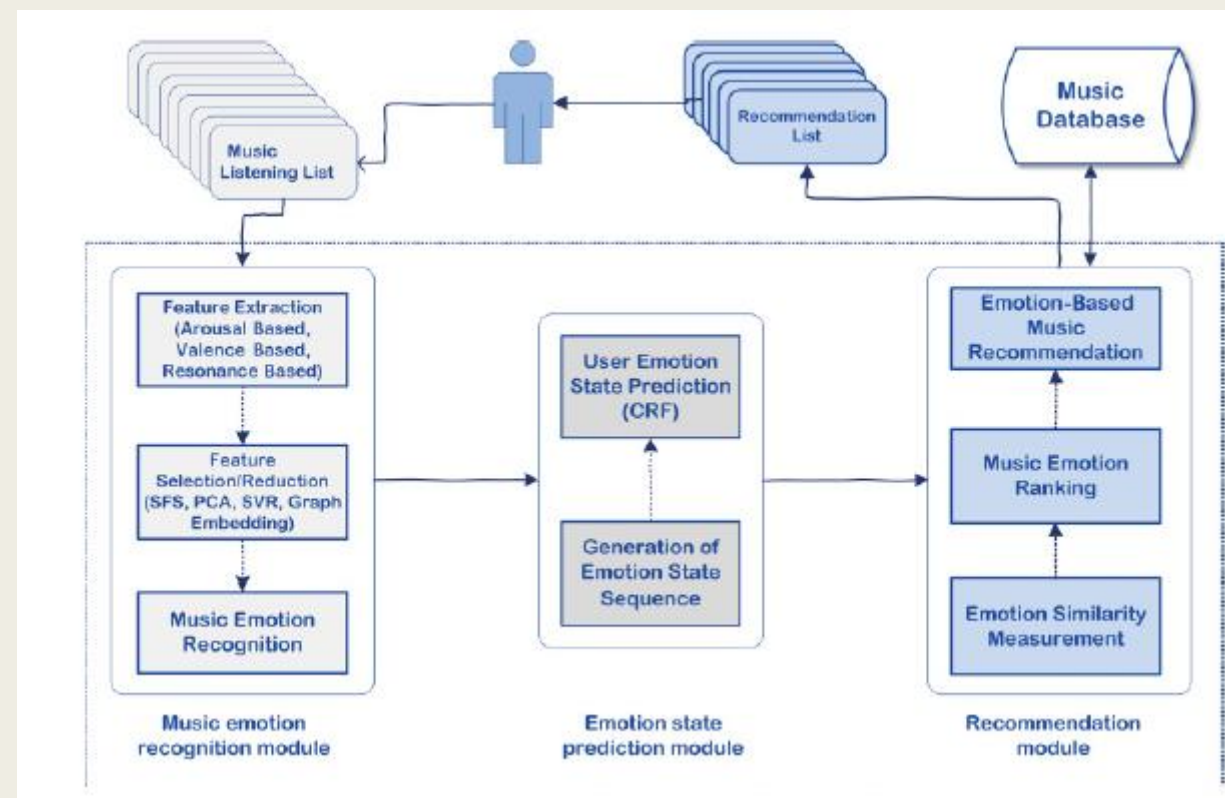
REKOMENDACJA



# Rozpoznawanie emocji – rekomendacja

Deng J., Leung C., Milani A., Chen L.: *Emotion Associated with Music: Classification, Prediction of Changes and Recommendation*. ACM Transactions on Interactive Intelligent Systems (2015)

- Dynamiczna rekomendacja na bazie rozpoznawania emocji
- Wejście: zapis fali
- Model emocji – RAV (*resonance, arousal, valence*)
- Rozpoznawanie emocji – regresja



# Rozpoznawanie emocji – rekomendacja

Deng J., Leung C., Milani A., Chen L.: *Emotion Associated with Music: Classification, Prediction of Changes and Recommendation*. ACM Transactions on Interactive Intelligent Systems (2015)

- Emocja użytkownika wynika tylko ze słuchanych utworów
- Model CRF (*Conditional random field*)

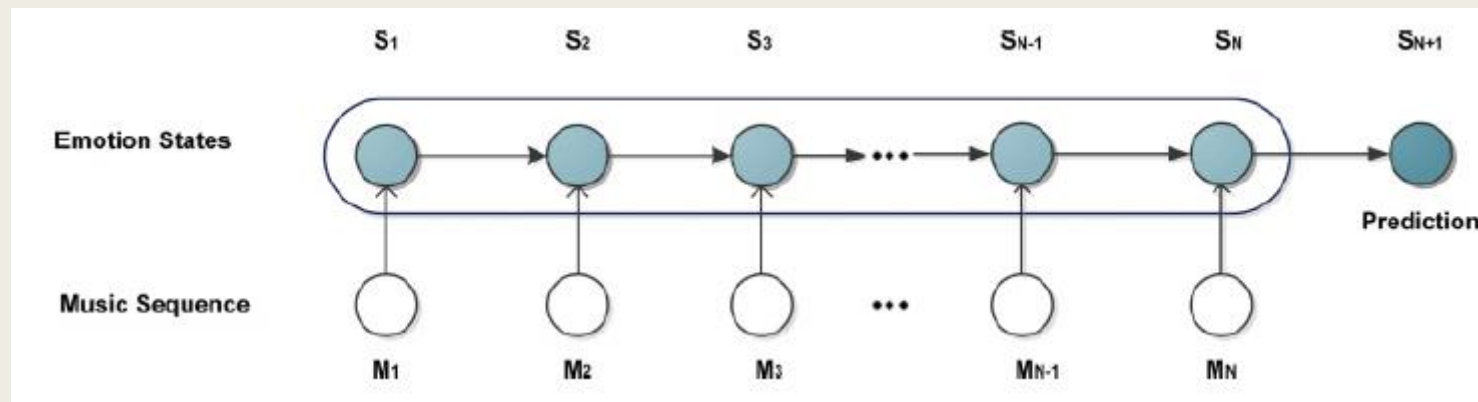


Table IX. Overall Evaluation Results (AER and AUC) of Listeners' Emotion State Prediction by CRF over Four Different Datasets

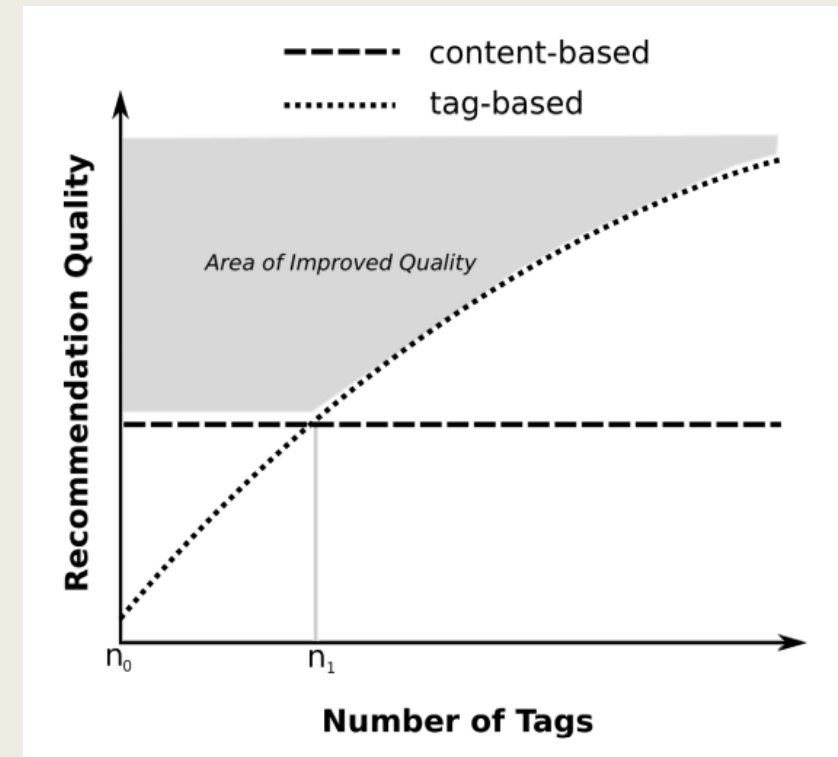
Evaluation	Dataset 1	Dataset 2	Dataset 3	Dataset 4
AER	0.12	0.15	0.18	0.16
AUC	0.84	0.81	0.79	0.82

*Note:* The smaller AER depicts a better prediction performance. The larger the AUC, the better the emotion state prediction. Dataset 1 with strong emotion intensity shows the lowest AER and the highest AUC.

# Rekomendacja na podstawie tagów

Horsburgh B., Craw S., Massie S.: *Learning pseudo-tags to augment sparse tagging in hybrid music recommender systems*. Artificial Intelligence (2015)

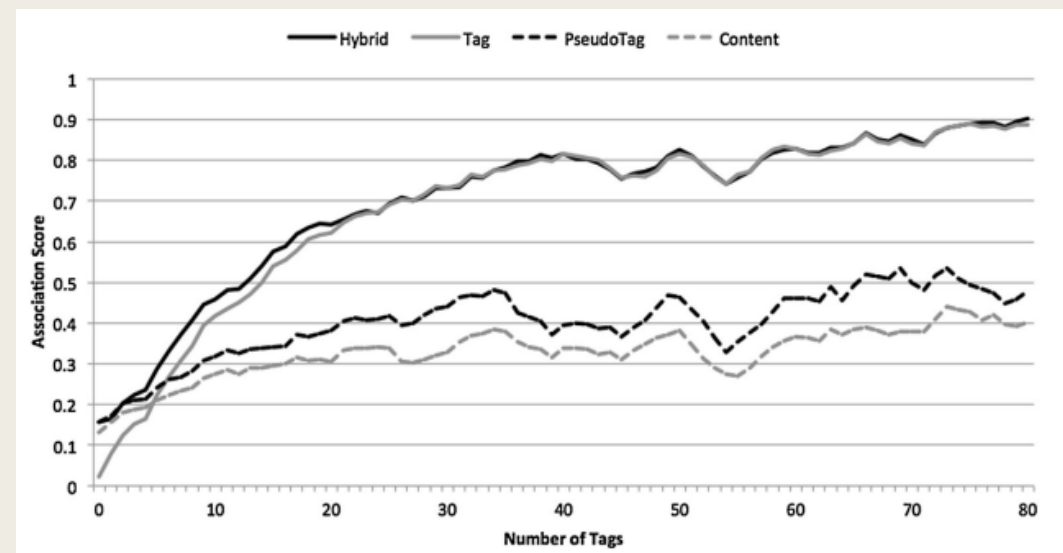
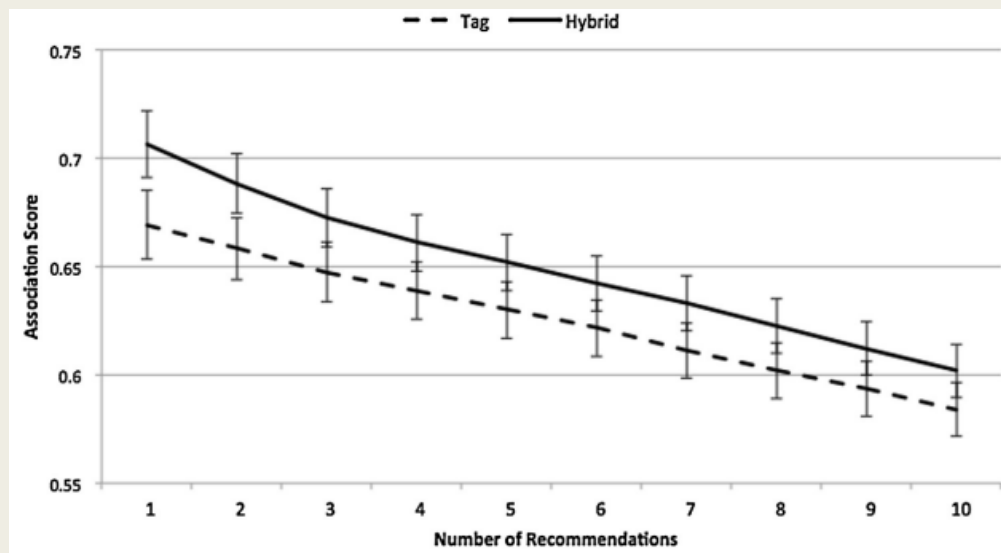
- Automatyczne tagowanie muzyki
- Wejście: zapis fali
- Stworzenie pseudotagów na bazie zawartości utworu i utworów sąsiednich
- K najbliższych sąsiadów



# Rekomendacja na podstawie tagów

Horsburgh B., Craw S., Massie S.: *Learning pseudo-tags to augment sparse tagging in hybrid music recommender systems*. Artificial Intelligence (2015)

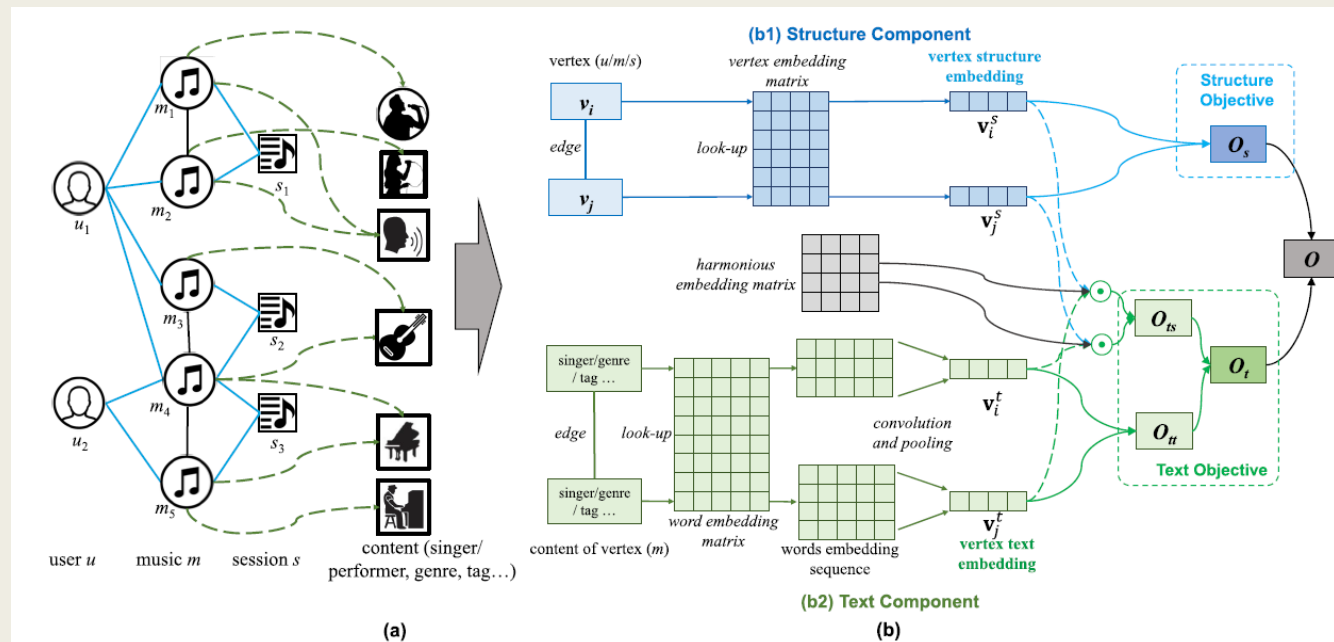
- Podejście hybrydowe – ważone łączenie tagów i pseudotagów



# Rekomendacja – kontekst i zawartość

Wang D., Zhang X., Yu D., Xu G., Deng S.: *CAME: Content- and Context-Aware Music Embedding for Recommendation*. IEEE Transactions on Neural Networks and Information Systems (2021)

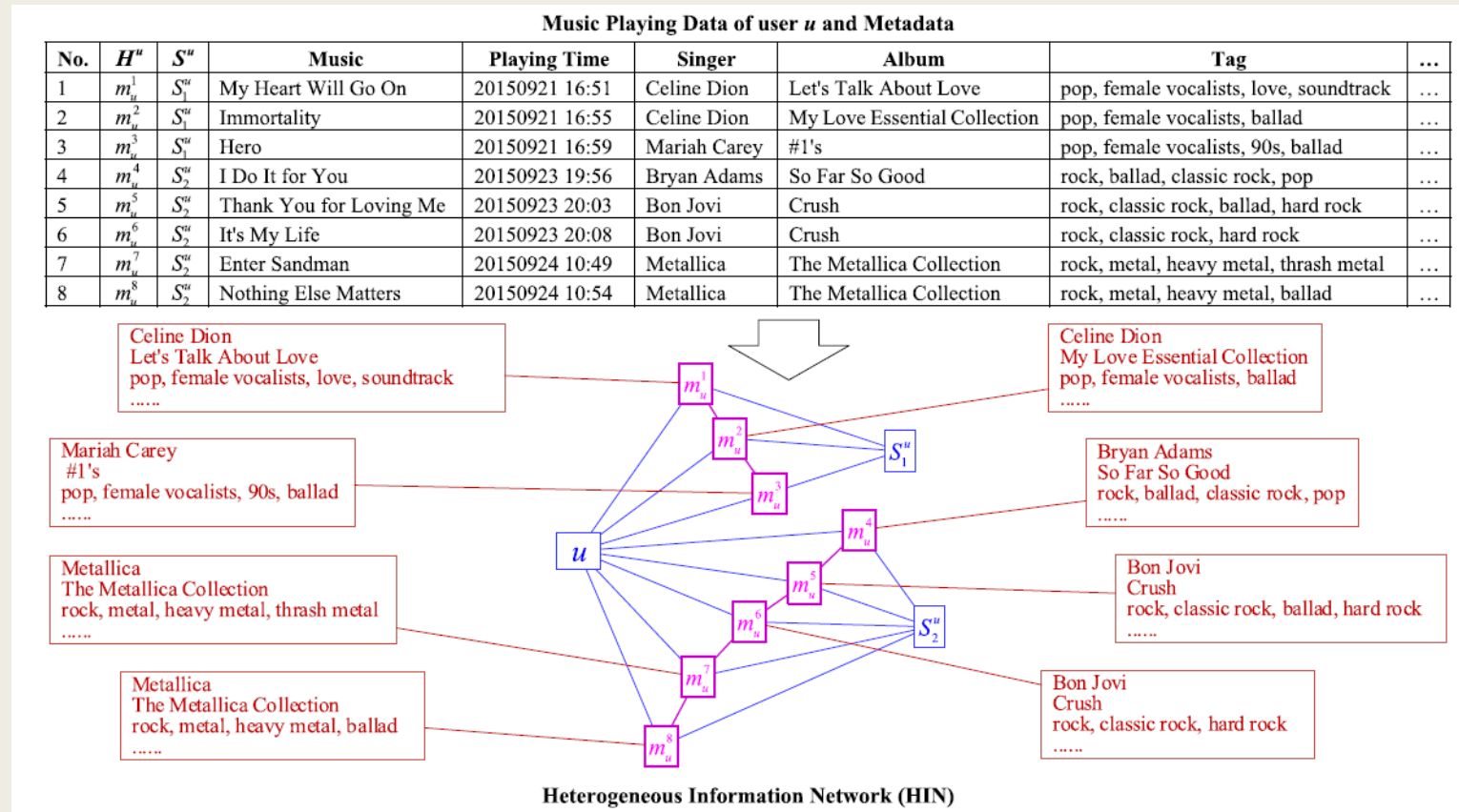
- Rekomendacja oparta o zawartość utworu i kontekst
- Kontekst – sieć HIN (*Heterogeneous Information Network*)
- Dwa typy zanurzenia dla wierzchołka – strukturalne i tekstowe





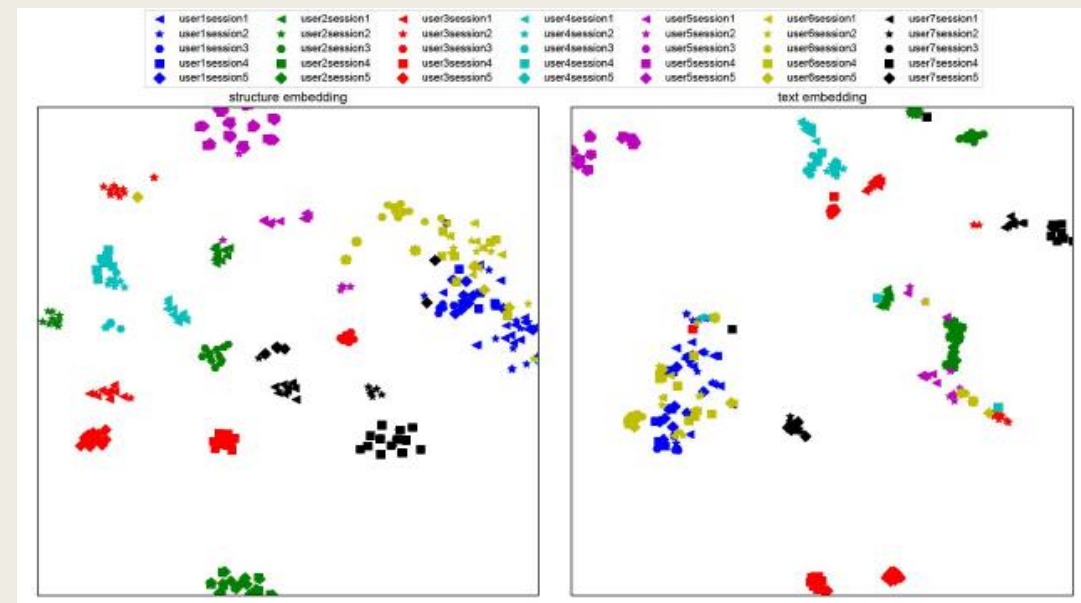
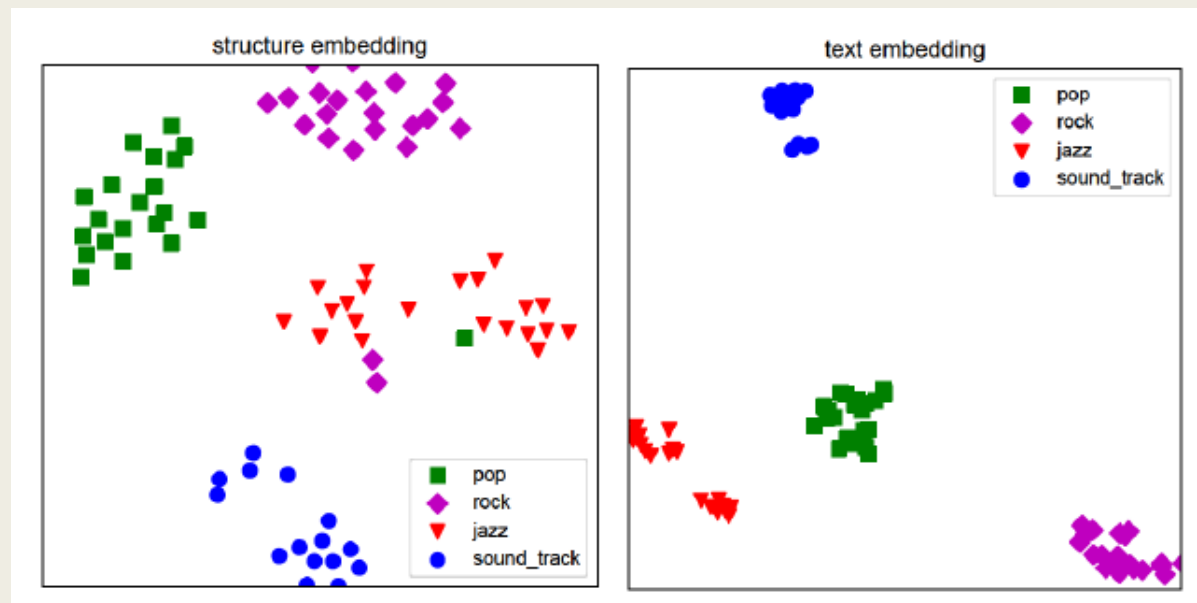
# Rekomendacja – kontekst i zawartość

Wang D., Zhang X., Yu D., Xu G., Deng S.: *CAME: Content- and Context-Aware Music Embedding for Recommendation*. IEEE Transactions on Neural Networks and Information Systems (2021)



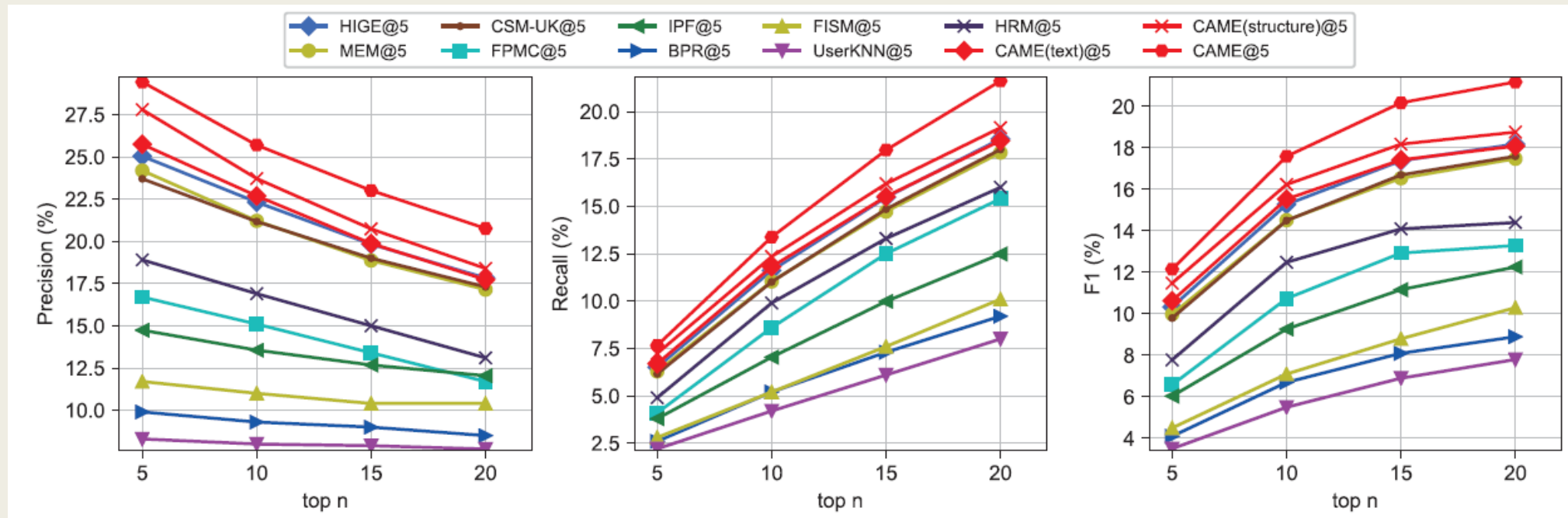
# Rekomendacja – kontekst i zawartość

Wang D., Zhang X., Yu D., Xu G., Deng S.: *CAME: Content- and Context-Aware Music Embedding for Recommendation*. IEEE Transactions on Neural Networks and Information Systems (2021)



# Rekomendacja – kontekst i zawartość

Wang D., Zhang X., Yu D., Xu G., Deng S.: *CAME: Content- and Context-Aware Music Embedding for Recommendation*. IEEE Transactions on Neural Networks and Information Systems (2021)



# PYTANIA I DYSKUSJA





DZIĘKUJĘ ZA UWAGĘ