JAN MYCKA

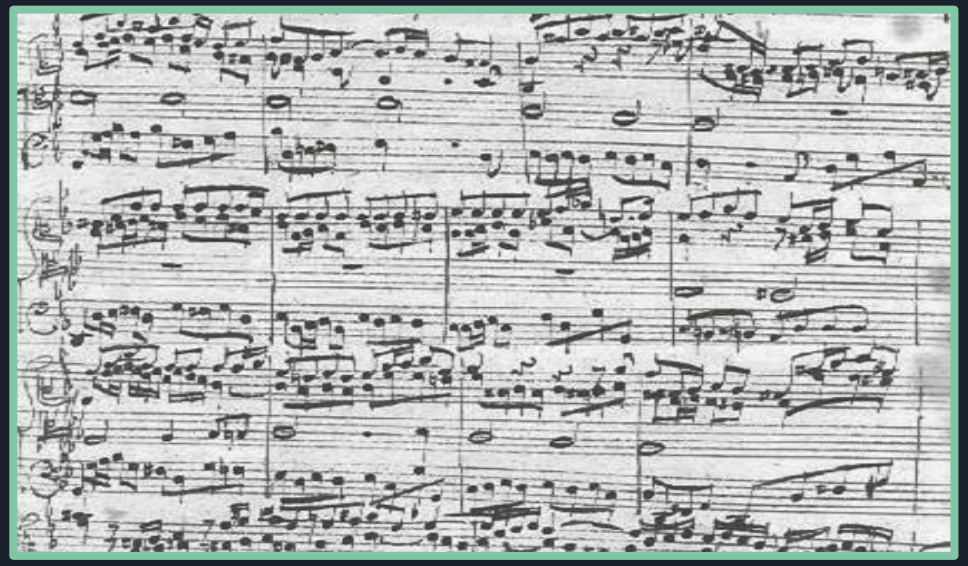
AUTOMATYCZNE GENEROWANIE MUZYKI EKSPRESYJNEJ

State-of-the-art

Istotność ekspresji w muzyce

Oore S., Simon I., Dieleman S., Eck D., Simonyan K.: *This Time with Feeling: Learning Expressive Musical Performance*

Kompozycja → Wykonanie

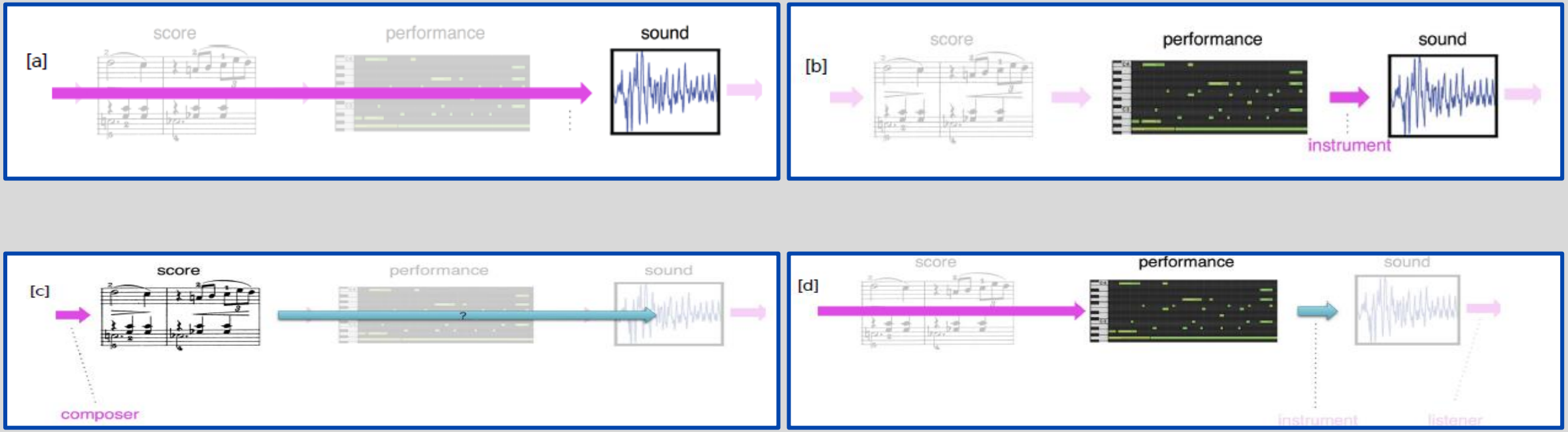


Źródło: https://leadingtone.tumblr.com/post/4006499072/christus-der-uns-selig-macht-chorale-prelude-bwv

Źródło: https://www.theguardian.com/music/2015/may/11/berlin-philharmonic-imon-rattle-orchestra-vote-chief-conductor

Istotność ekspresji w muzyce

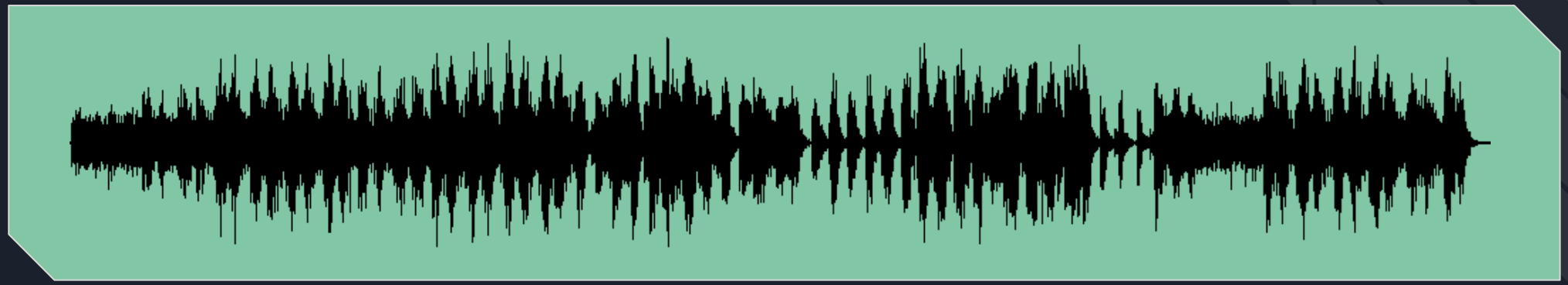
Oore S., Simon I., Dieleman S., Eck D., Simonyan K.: *This Time with Feeling: Learning Expressive Musical Performance*



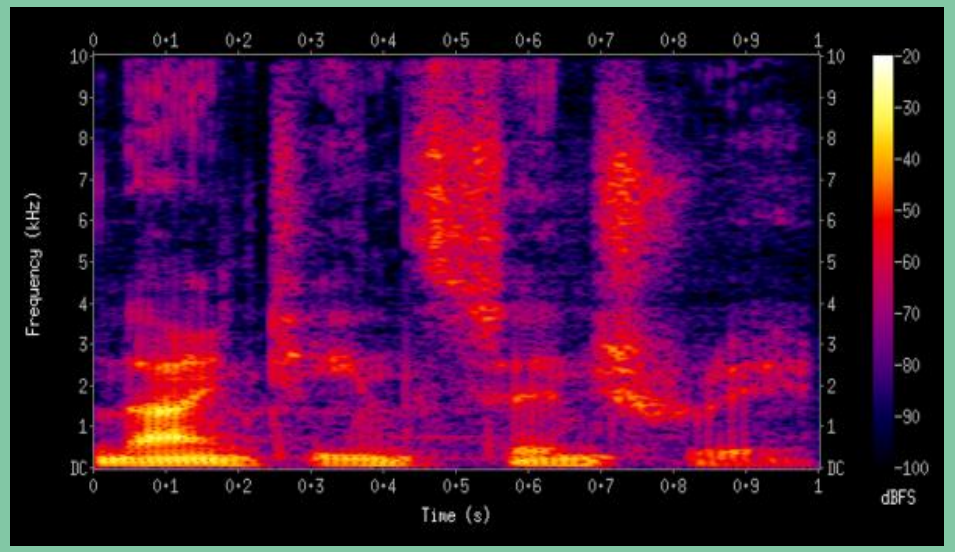
REPREZENTACJA EKSPRESYJNYCH DANYCH MUZYCZNYCH

Zapis fali dźwiękowej (format *wave*)

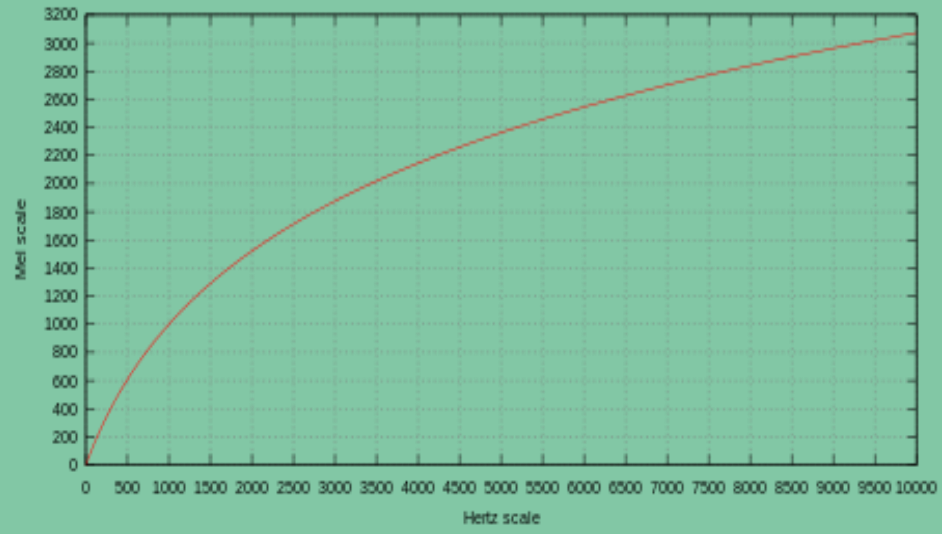
- Wszystkie możliwe informacje o wykonaniu
- Mało przystępna postać danych



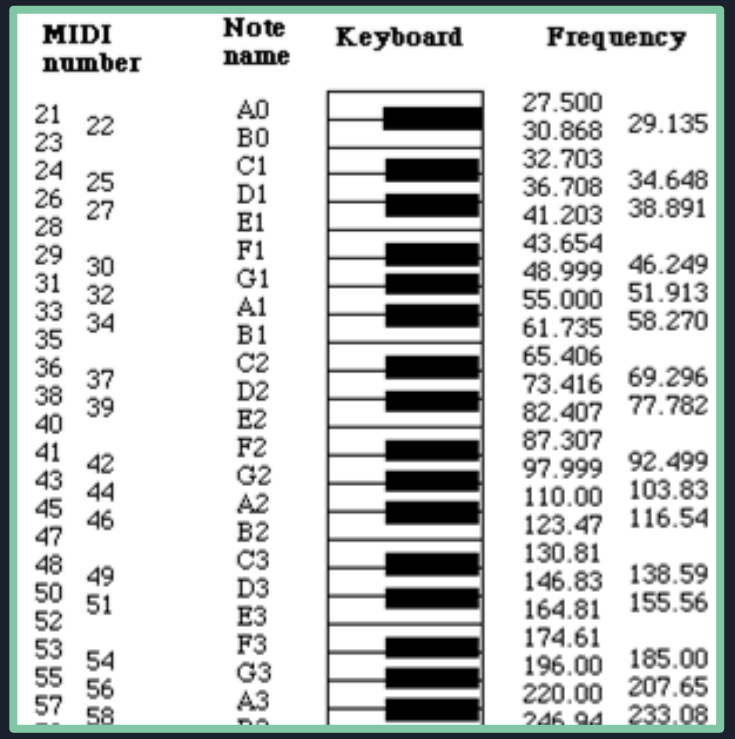
Spektrogram



Skala *mel*



Standard MIDI (format *mid*)



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

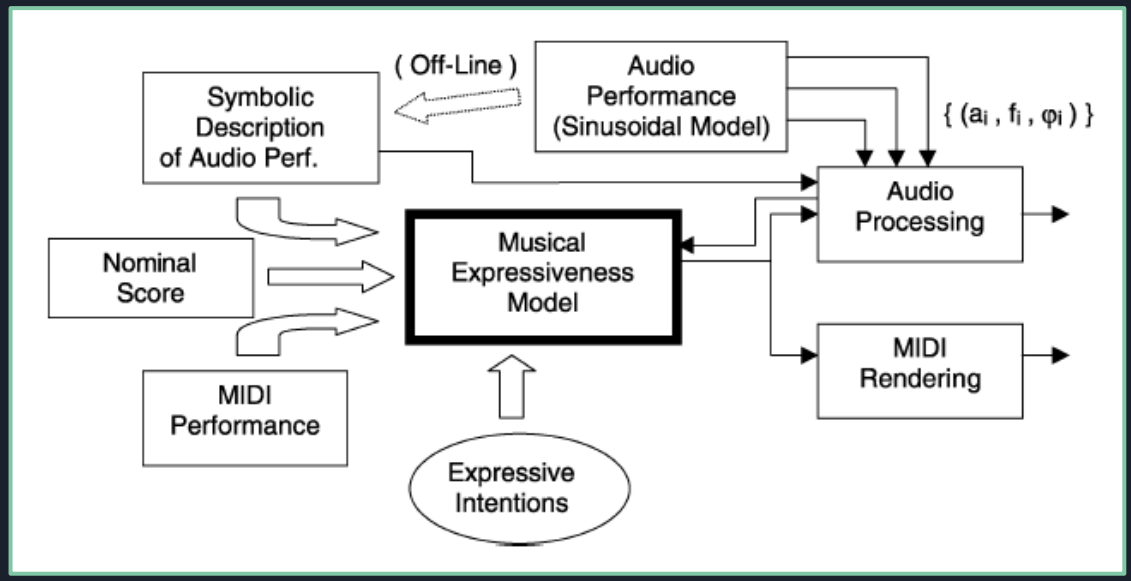
- MIDI – *Musical Instrument Digital Interface* (ang. cyfrowy interfejs instrumentów muzycznych)
- Sekwencja zdarzeń
- Opis ekspresji przez długość trwania nuty i głośność

PODEJŚCIA STATE-OF-THE-ART

Modyfikacja istniejącego wykonania

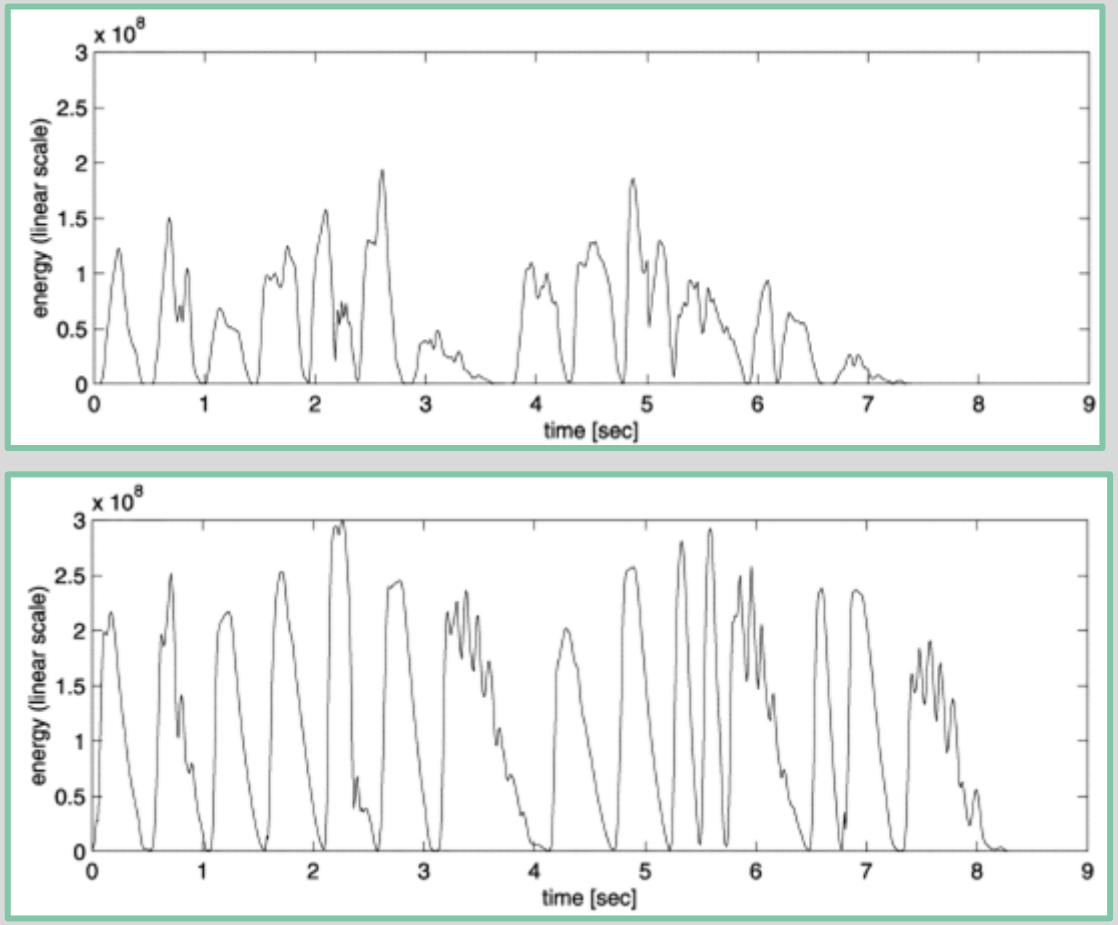
Canazza S., De Poli G., Drioli C., Roda A., Vidolin A.: *Modeling and Control of Expressiveness in Music Performance*. Proceedings of the IEEE (2004)

- 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

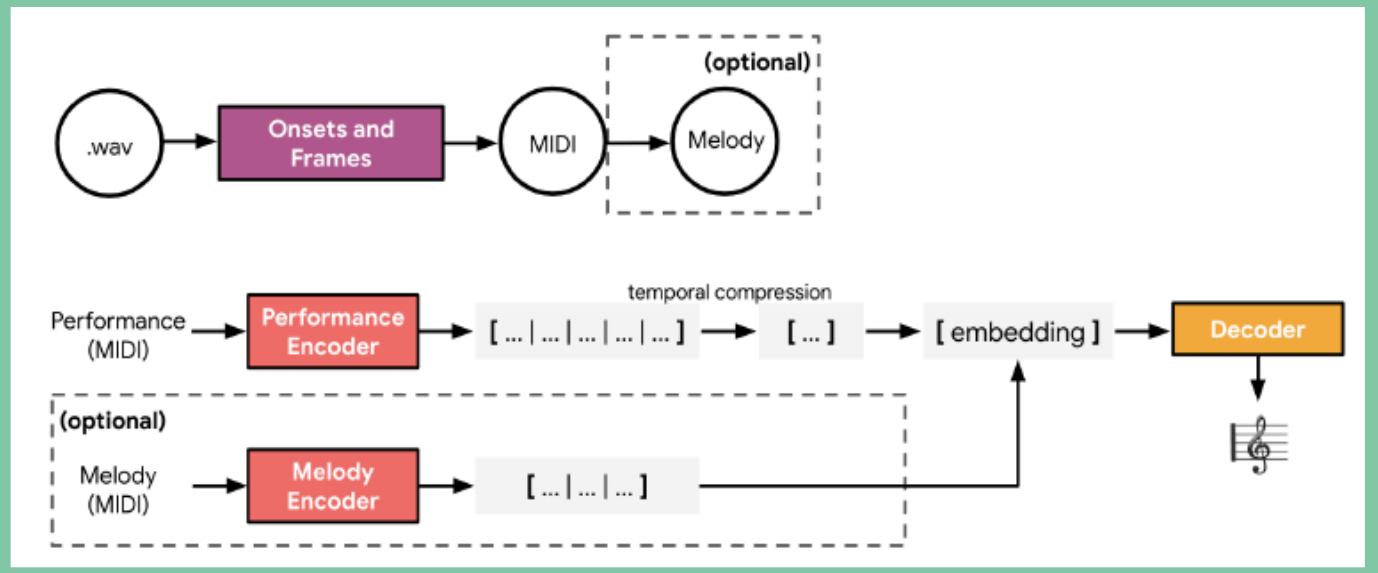
Canazza S., De Poli G., Drioli C., Roda A., Vidolin A.: *Modeling and Control of Expressiveness in Music Performance*. Proceedings of the IEEE (2004)



Energia wykonania neutralnego (góra) i wykonania *ciężkiego* (dół)

Generowanie na podstawie MIDI

Choi K., Hawthrone C., Simon I., Dinculescau M., Engel J.: *Encoding musical style with Transformer Autoencoders*



- Model *encoder-decoder*

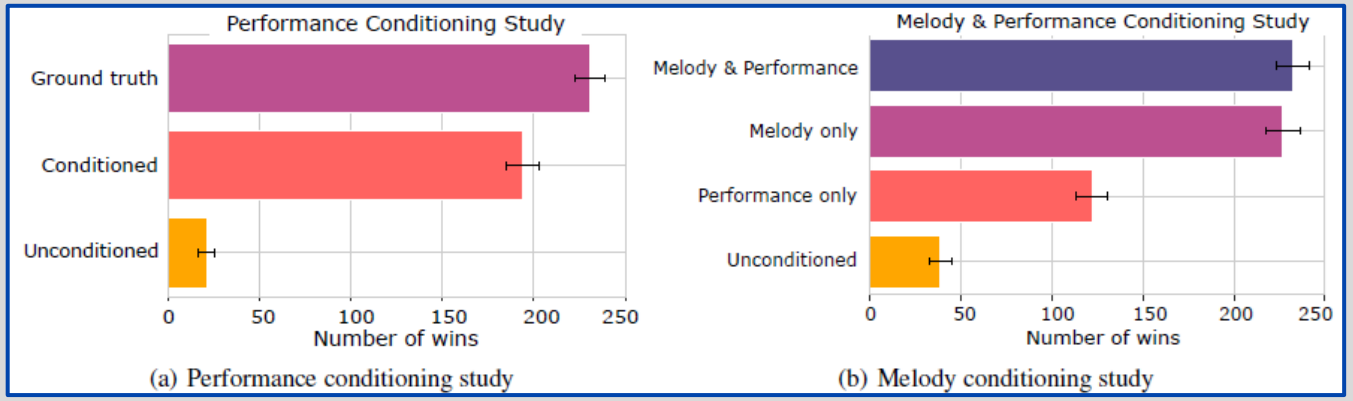
- Opcjonalne korzystanie z reprezentacji melodii

- Wszystkie części oparte o Transformery

Generowanie na podstawie MIDI

Choi K., Hawthrone C., Simon I., Dinculescau M., Engel J.: *Encoding musical style with Transformer Autoencoders*

Model variation	MAESTRO	YouTube
Melody-only Transformer with rel. attention (Huang et al., 2019b)	1.786	1.302
Melody & performance autoencoder with rel. attention, sum (ours)	**1.706**	1.275
Melody & performance autoencoder with rel. attention, concat (ours)	1.713	**1.237**
Melody & performance autoencoder with rel. attention, tile (ours)	1.709	1.248

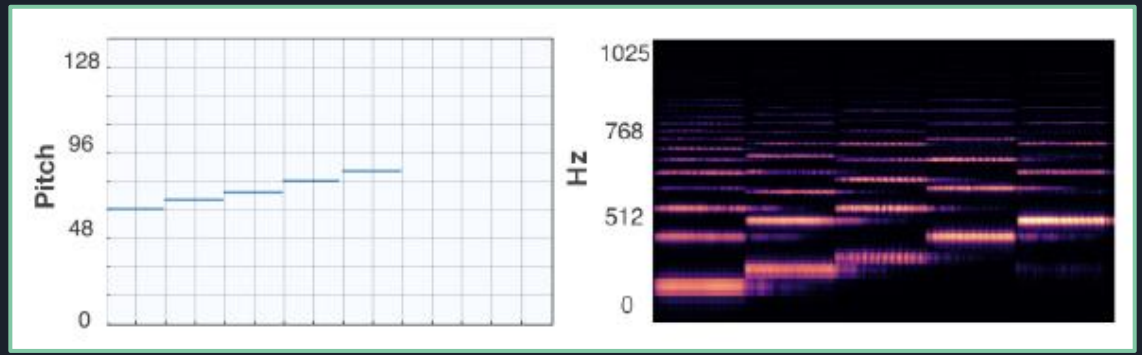


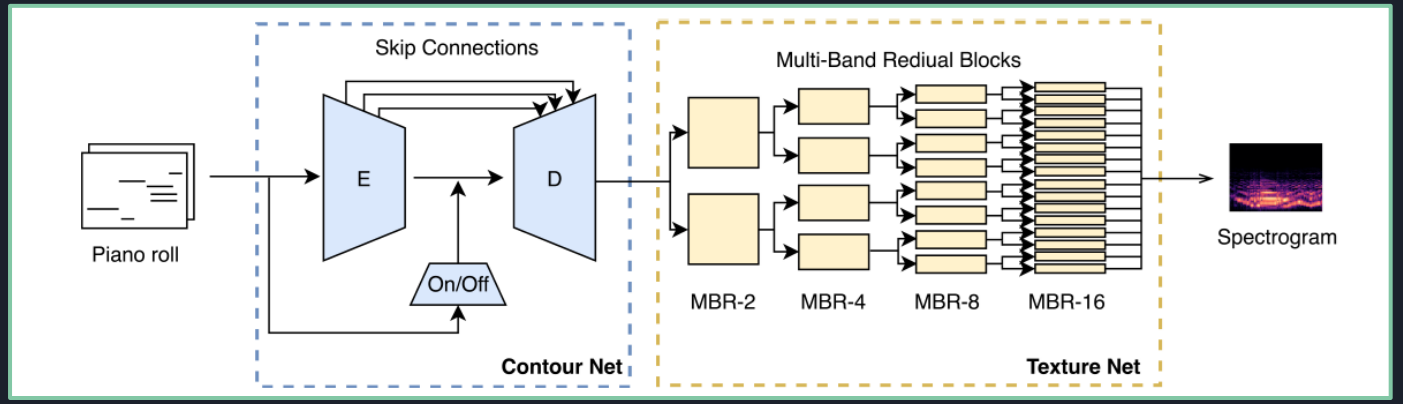
(a) Performance conditioning study (b) Melody conditioning study

Generowanie spektrogramu

Wang B., Yang Y.: *PerformanceNet: Score-to-Audio music generation with multi-band residua network*

- Wstępne tworzenie spektrogramu
- Uszczegóławianie (zwiększenie rozdzielczości)

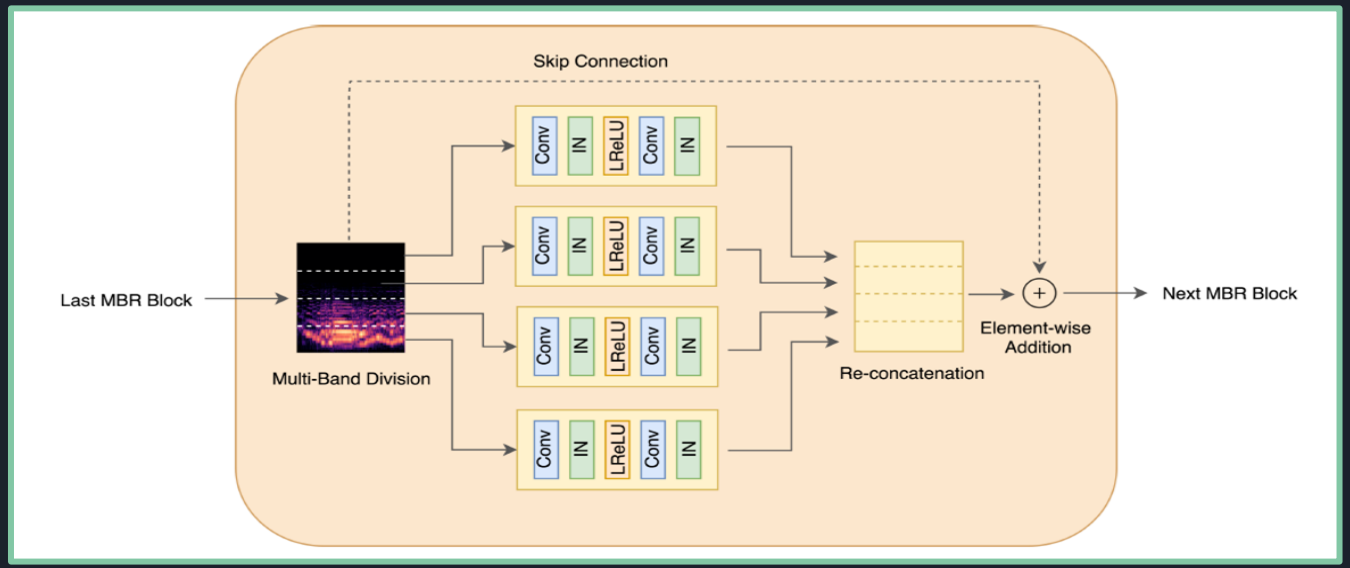




Generowanie spektrogramu

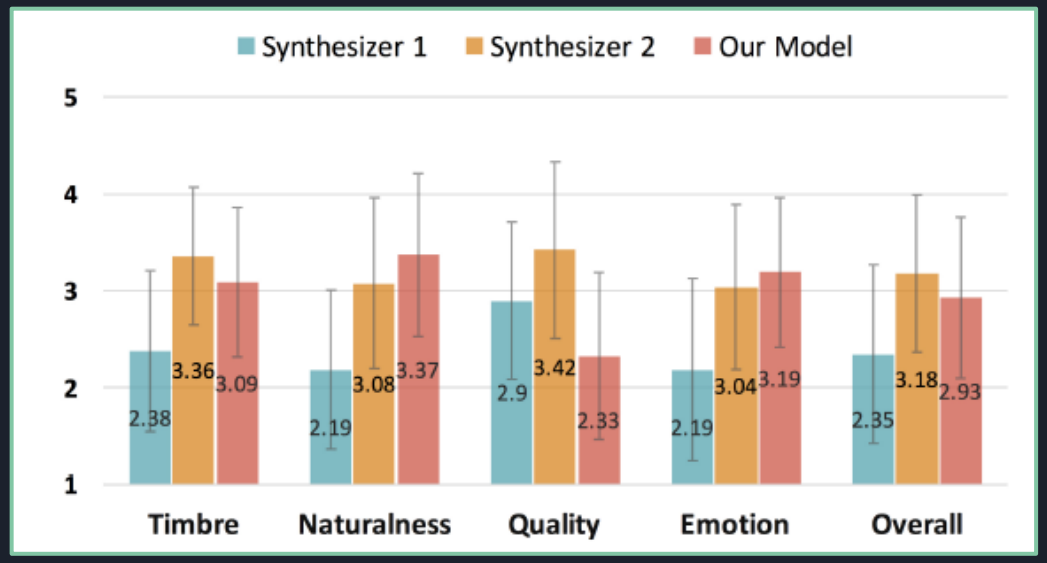
Wang B., Yang Y.: *PerformanceNet: Score-to-Audio music generation with multi-band residua network*

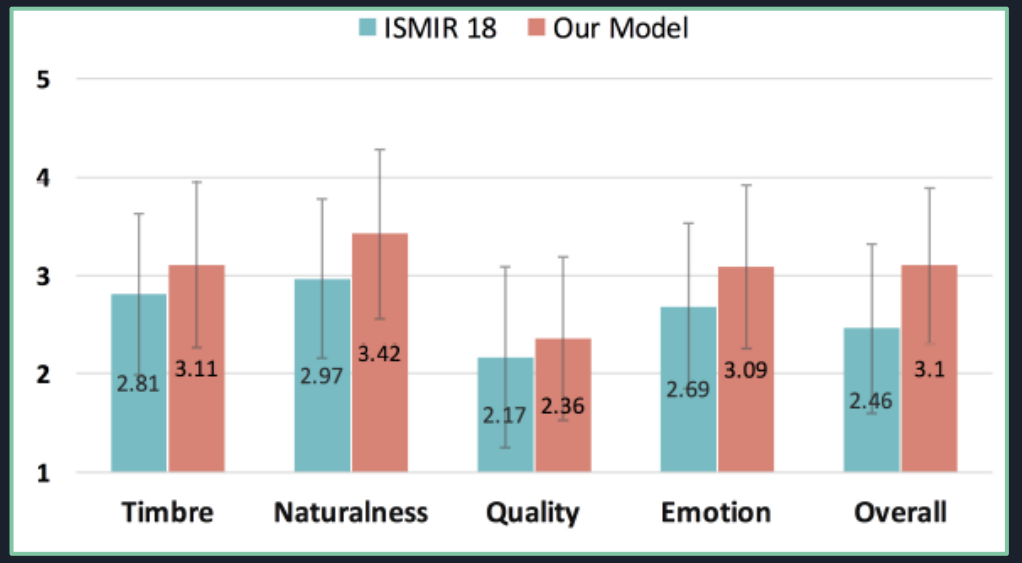
- Podział spektrogramu względem częstotliwości
- Uczenie różnych charakterystyk dla różnych częstotliwości



Generowanie spektrogramu

Wang B., Yang Y.: *PerformanceNet: Score-to-Audio music generation with multi-band residua network*





Generowanie *wave* - inspiracja

van den Oord A., Dieleman S., Heiga Z., Simonyan K., Vinyals O., Graves A., Kalchbrenner N., Senior A., Kavukcuoglu K.: *WaveNet: A generative model for raw audio*

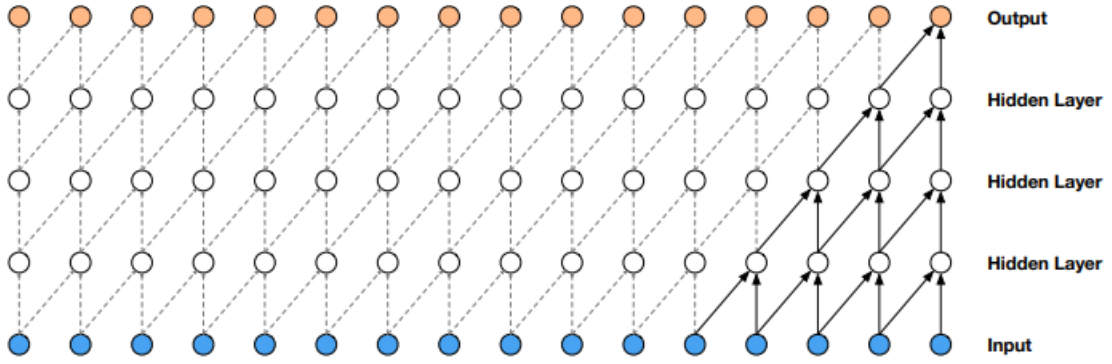


Figure 2: Visualization of a stack of causal convolutional layers.

- Model text-to-speech
- Poszerzenie kontekstu

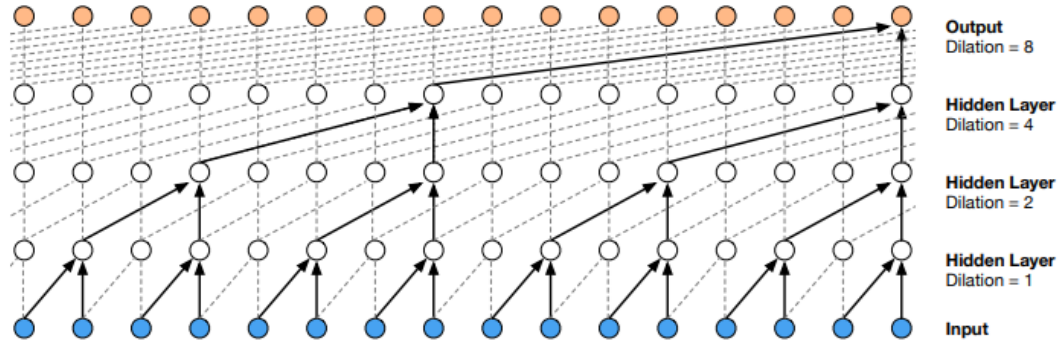
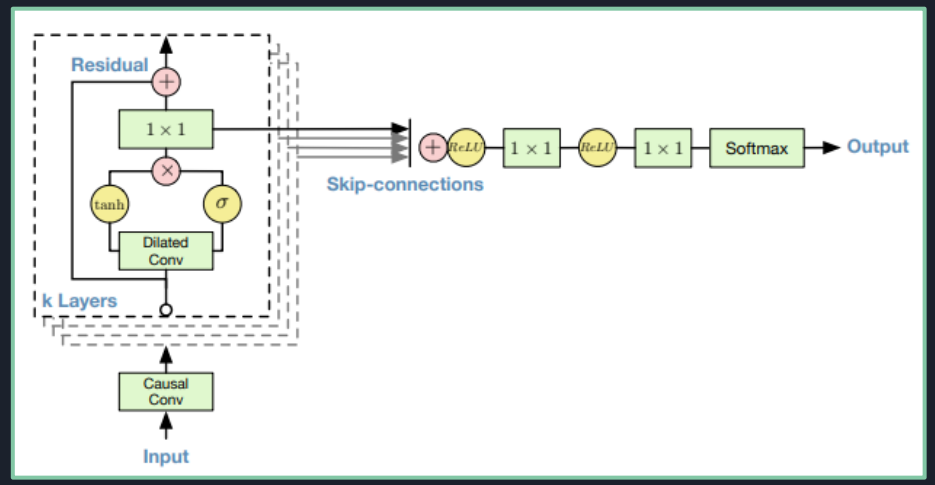
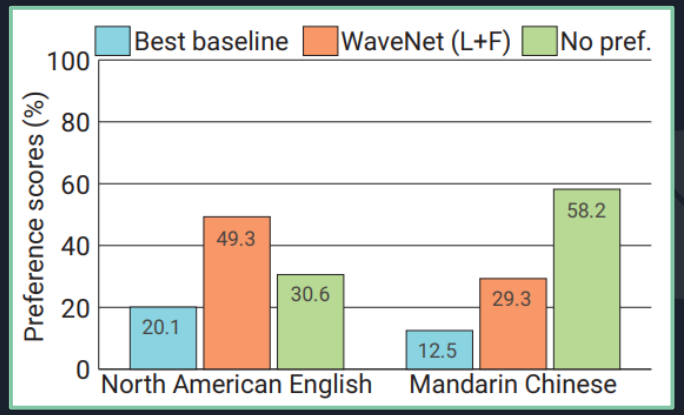


Figure 3: Visualization of a stack of *dilated* causal convolutional layers.

Generowanie *wave* - inspiracja

van den Oord A., Dieleman S., Heiga Z., Simonyan K., Vinyals O., Graves A., Kalchbrenner N., Senior A., Kavukcuoglu K.: *WaveNet: A generative model for raw audio*

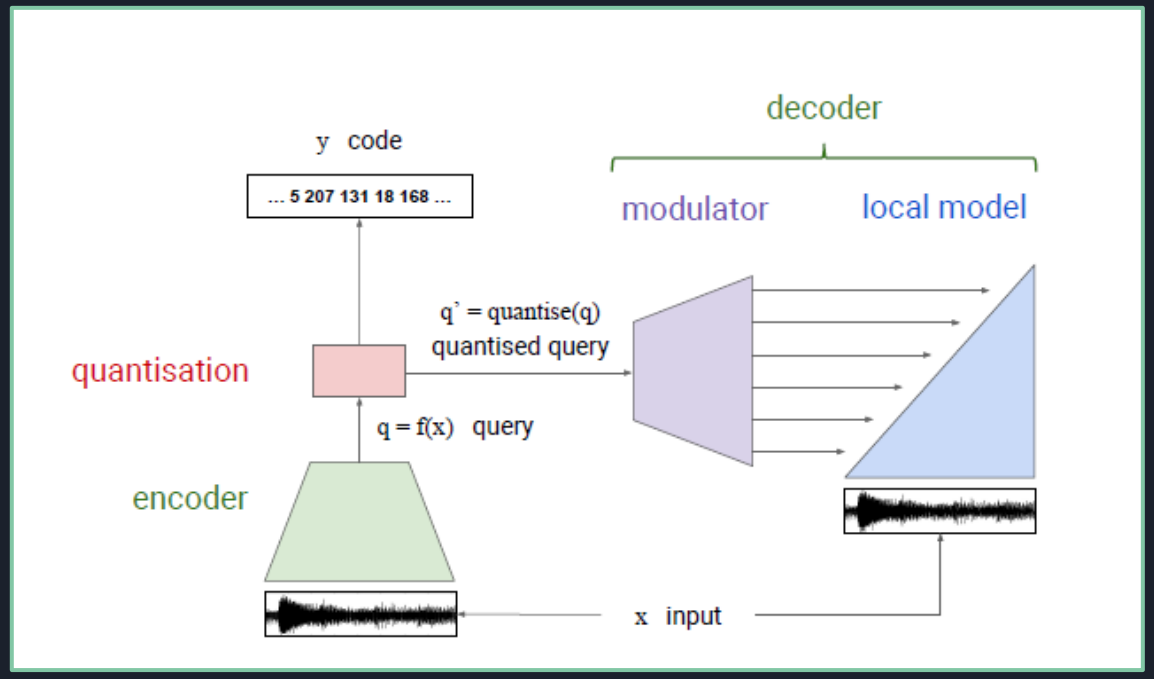




- Wyżej oceniany niż pozostałe modele text-to-speech

Generowanie *wave*

Dieleman S., van den Oord A., Simonyan K.: *The challenge of realistic music generation: modeling raw audio at scale*



- Model *encoder-decoder*
- Wszystkie części oparte o WaveNet
- *Bottleneck* tworzony przy pomocy dyskretyzacji danych

Generowanie *wave*

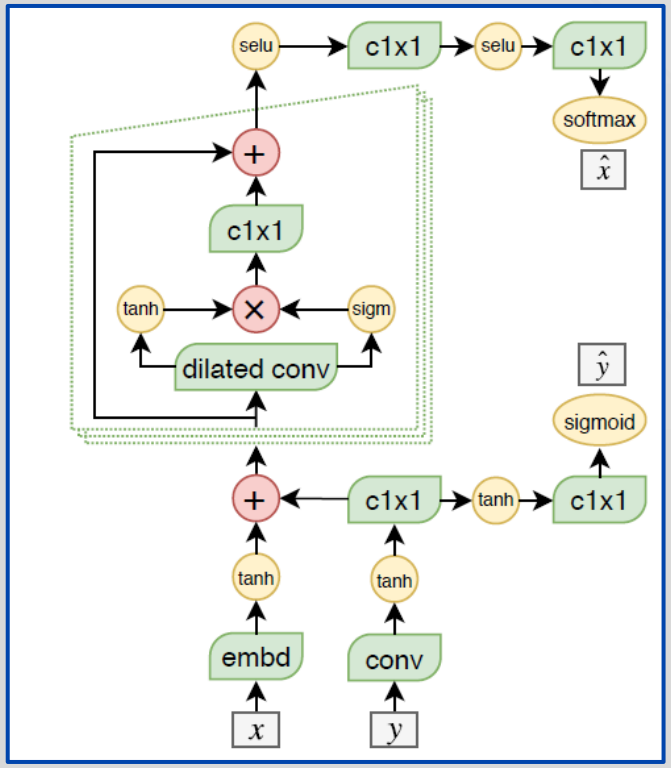
Dieleman S., van den Oord A., Simonyan K.: *The challenge of realistic music generation: modeling raw audio at scale*

Model	Num. levels	RF	Human evaluation Fidelity	Musicality
Large WaveNet	1	384 ms	3.82 ± 0.18	2.43 ± 0.14
Very large WaveNet	1	768 ms	3.82 ± 0.20	2.89 ± 0.17
Thin WaveNet with large RF	1	3072 ms	2.43 ± 0.17	1.71 ± 0.18
hop-8 VQ-VAE + large WaveNet	2	3072 ms	3.79 ± 0.16	3.04 ± 0.16
hop-64 VQ-VAE + large WaveNet	2	24576 ms	3.54 ± 0.18	3.07 ± 0.17
VQ-VAE + PBT-VQ-VAE + large WaveNet	3	24576 ms	3.71 ± 0.18	4.04 ± 0.14
VQ-VAE + AMAE + large WaveNet	3	24576 ms	3.93 ± 0.18	3.46 ± 0.15

Generowanie *wave*, transfer stylu

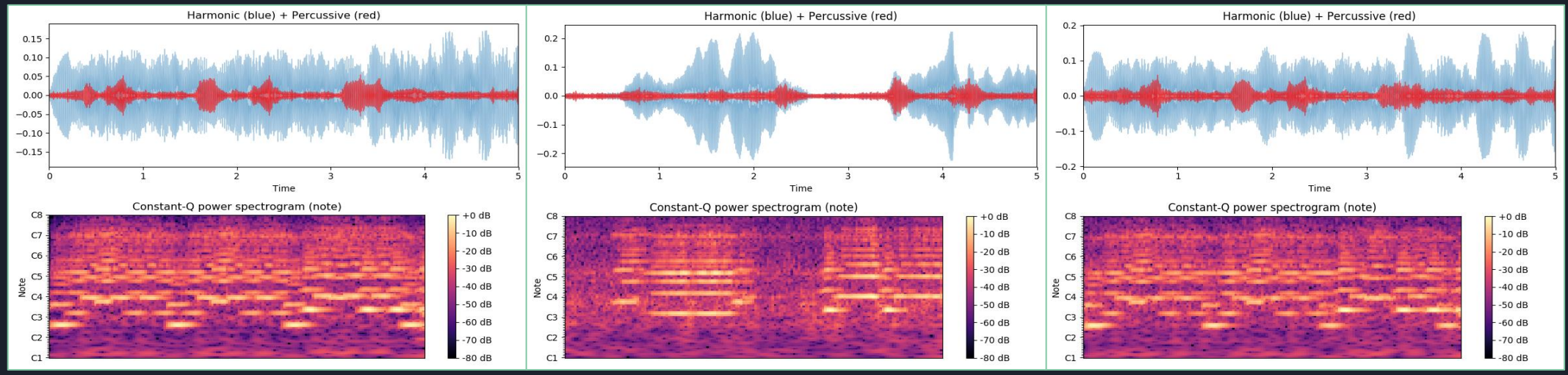
Schimbinschi F., Walder C., Erfani S. M., Bailey J.: *SynthNet: Learning to synthesize music end-to-end*

- Model oparty o WaveNet
- Korzysta zarówno z MIDI jak i *waveform*



Generowanie *wave*, transfer stylu

Schimbinschi F., Walder C., Erfani S. M., Bailey J.: *SynthNet: Learning to synthesize music end-to-end*

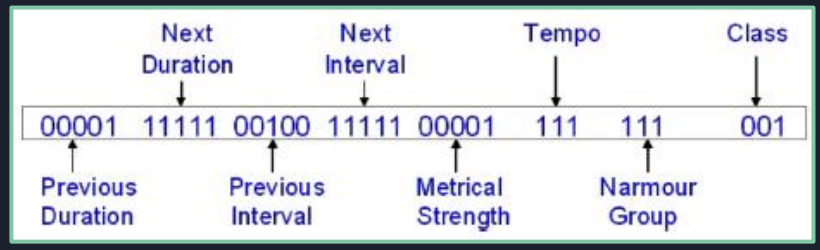


Experiment	Piano	Glockenspiel	Guitar	Cello	Trumpet	Flute	Square	All
WaveNet L26	2.22±0.25	2.48±0.23	2.18±0.25	2.37±0.28	2.18±0.29	2.37±0.22	2.30±0.09	2.30±0.10
DeepVoice L26	2.55±0.32	1.85±0.23	2.30±0.39	2.62±0.27	2.28±0.32	2.20±0.25	1.87±0.03	2.24±0.11
SynthNet L24	4.75±0.14	4.45±0.17	4.30±0.19	4.50±0.15	4.25±0.18	4.15±0.21	4.10±0.16	**4.36±0.07**

Tworzenie reguł na podstawie *wave*

Ramirez R., Maestre E., Serra X.: *A rule-based evolutionary approach to music performance modeling*

```
GeneticSeqCovAlg(Class, Fitness, Threshold, p, r, m, Examples)
  Pos = examples which belong to Class
  Neg = examples which do not belong to Class
  Learned_rules = {}
  While Pos do
      P = generate p hypotheses at random
      For each hypothesis h in P,
         Compute fitness(h)
      While max(fitness(h))<Threshold and #generations<400 do
         Create a new generation Pnew
         P = Pnew
         For each h in P,
            Compute fitness(h)
      NewRule = the hypothesis in P that has the highest fitness
      Rpos = members of Pos covered by NewRule
      Compute PredictedValue(Rpos)
      NumericNewRule = NewRule with Class replaced by
                                              Regression(Rpos)
      Learned_rules = Learned_rules + NumericNewrule
      Pos = Pos - Rpos
  Return Learned_rules
```



- Zbiór reguł opisujący zależności między dźwiękami
- Funkcja dopasowania $\frac{tP^{\alpha}}{tP+fP}$

Tworzenie reguł na podstawie *wave*

Ramirez R., Maestre E., Serra X.: *A rule-based evolutionary approach to music performance modeling*

Attribute	Bit Meaning
Previous duration	*Bit* 1: much shorter than current note *Bit* 2: shorter than current note *Bit* 3: same than current note *Bit* 4: longer than current note *Bit* 5: much longer than current note
Next duration	*Bit* 1: much shorter than current note *Bit* 2: shorter than current note *Bit* 3: same than current note *Bit* 4: longer than current note *Bit* 5: much longer than current note
Previous interval	*Bit* 1: much lower than current note *Bit* 2: lower than current note *Bit* 3: same than current note *Bit* 4: higher than current note *Bit* 5: much higher than current note
Next interval	*Bit* 1: much lower than current note *Bit* 2: lower than current note *Bit* 3: same than current note *Bit* 4: higher than current note *Bit* 5: much higher than current note

Metrical strength	*Bit* 1: very weak *Bit* 2: weak *Bit* 3: medium *Bit* 4: strong *Bit* 5: very strong
Tempo	*Bit* 1: slow *Bit* 2: nominal *Bit* 3: fast
Narmour group	000 = "P" group 001 = "D" group 010 = "ID" group 011 = "IP" group 100 = "VP" group 101 = "R" group 110 = "IR" group 111 = "VR" group
Class	*Bit* 1: shorten/delay/soft/ornamentation *Bit* 2: same-duration/same-onset/medium/none *Bit* 3: lengthen/advance/loud/none

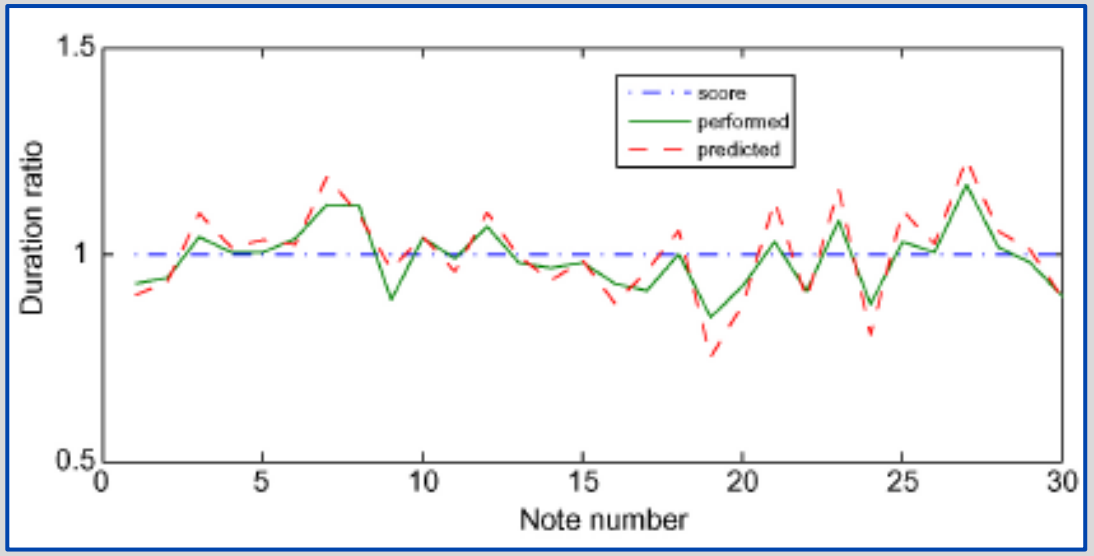
"if the previous note duration is much longer and its pitch is the same and it is in a very strong metrical position and the current note appears in Narmour group R then lengthen the duration of the current note"

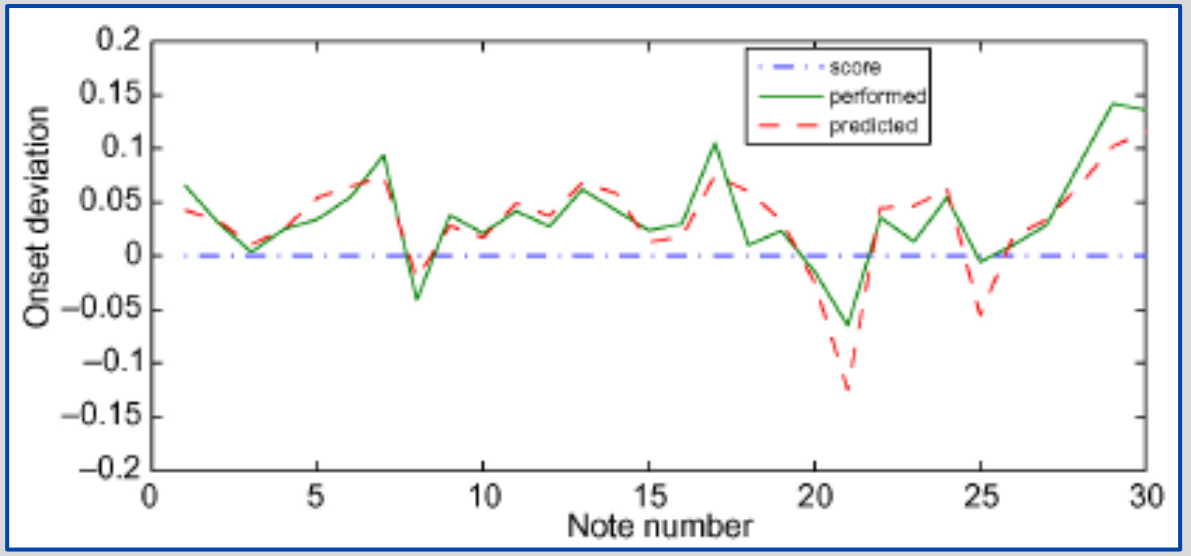
is coded as the binary string

00001 11111 00100 11111 00001 111 110 001.

Tworzenie reguł na podstawie *wave*

Ramirez R., Maestre E., Serra X.: *A rule-based evolutionary approach to music performance modeling*



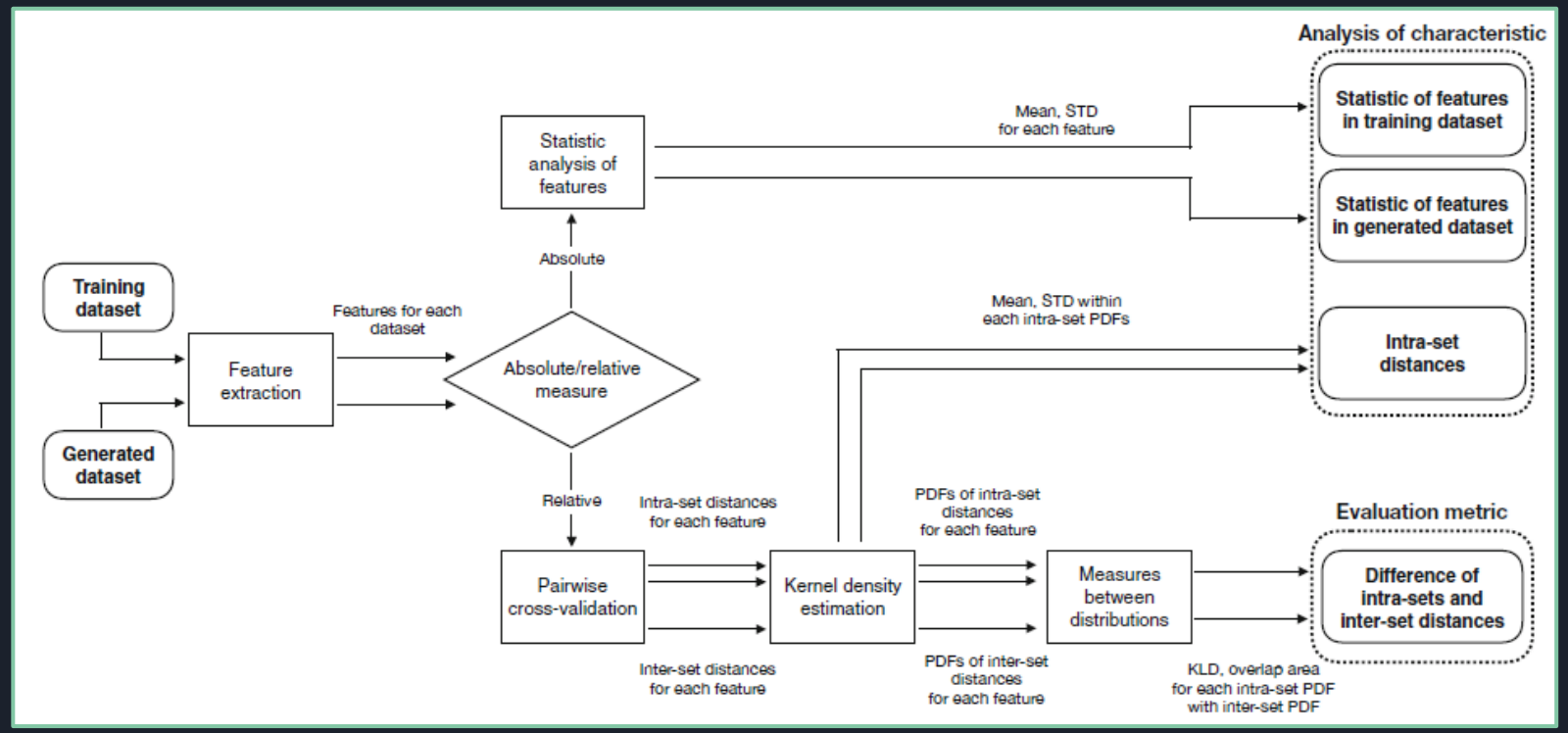


Problemy i wyzwania

- Jakość generowanej muzyki
- Ocena generowanej muzyki

Obiektywna ocena

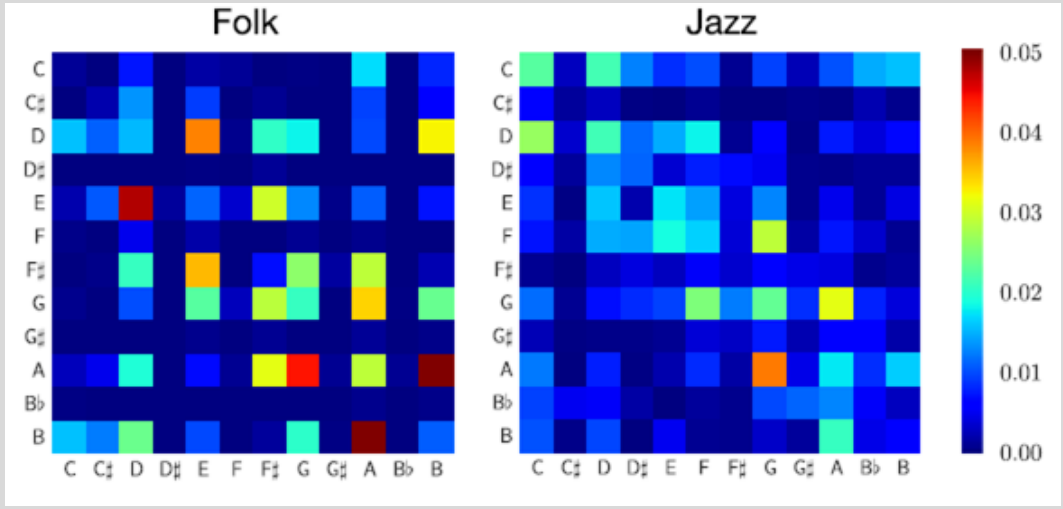
Yang L., Lerch A.: *On the evaluation of generative models in music*



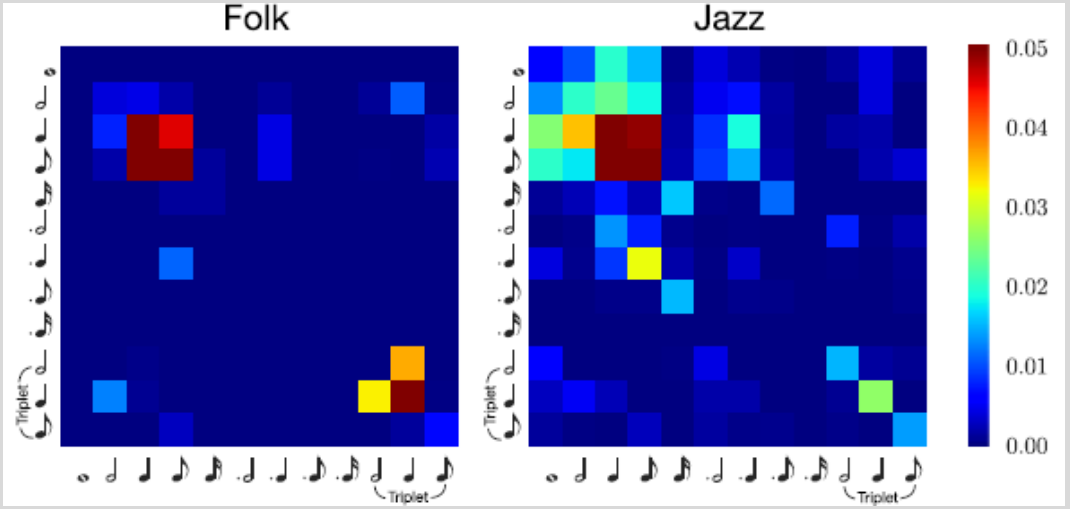
- Miara obiektywna i względna
- Oparta o właściwości statystyczne

Obiektywna ocena - właściwości

Yang L., Lerch A.: *On the evaluation of generative models in music*



Macierz przejść dźwięków



Macierz przejść długości dźwięków

Problem przydatności

Sturm B. L., Ben-Tal O., Monaghan U., Collins N., Herremans D., Chew E., Hadjeres G., Deruty E., Pachet F.: *Machine learning research that matters for music creation: A case study*

Pieces for organ: The Glas Herry Comment & X:7153 by *folk-rnn + DeepBach* (2017)	**Richard Salmon** *organ*
Traditional Irish Sets (with *folk-rnn* tunes in *italics*) • **Jigs** (The Cuil Aodha, The Dusty Windowsill, *The Glas Herry Comment*) • **Slow Reels** (Maghera Mountain, *X:2897*) • **Fast Reels** (The Rookery, *X:1068*, Toss The Feathers II)	**Daren Banarsë and Musicians**
March to the Mainframe, Interlude, The Humours of Time Pigeon by Bob L. Sturm + *folk-rnn* (2017)	**Ensemble x.y**
Ed SheerAI vs XenAkIs vs AIdele by Nick Collins (2017)	**Ensemble x.y**
3 morphed pieces from "A Little Notebook for Anna Magdalena" by J. S. Bach (1722) + MorpheuS (2017) **3 morphed pieces from "30 and 24 Pieces for Children"** by Kabalevsky (1937) + MorpheuS (2017)	**Elaine Chew** *piano*
Safe Houses by Úna Monaghan + *folk-rnn* (2017) ***The Choice*** by Úna Monaghan (2015) ***The Chinwag*** by Úna Monaghan (2015)	**Úna Monaghan** *Irish harp, concertina, electronics*
Pieces for organ: X:633 & The Drunken Pint by *folk-rnn + DeepBach* (2017)	**Richard Salmon** *organ*
Chicken Bits and Bits and Bobs by Bob L. Sturm + *folk-rnn* (2017)	**Ensemble x.y**
Bastard Tunes by Oded Ben-Tal + *folk-rnn* (2017)	**Ensemble x.y**

DZIĘKUJĘ ZA UWAGĘ