

KLASYFIKACJA EMOCJI W MUZYCE FILMOWEJ Z WYKORZYSTANIEM TESTÓW SUBIEKTYWNYCH

T. Ciborowski¹, S. Reginis¹, A. Kurowski^{1,2}, D. Weber^{1,2} i B. Kostek^{1,2}

¹Katedra Systemów Multimedialnych, Wydział Elektroniki, Telekomunikacji i Informatyki, Politechnika Gdańska

²Laboratorium Akustyki Fonicznej, Wydział Elektroniki, Telekomunikacji i Informatyki, Politechnika Gdańska

s165501@student.pg.edu.pl

Streszczenie

Celem referatu było przedstawienie testów odsłuchowych, w których zadaniem osób ankietowanych było przypisanie danego fragmentu muzycznego do odpowiedniej klasy emocji. Kolejne kroki eksperymentu obejmowały wybór muzyki filmowej do testów (baza Epidemic Sound), przygotowanie założeń ankiety oraz modelu emocji wykorzystywanych w testach odsłuchowych, jak również konstrukcję ankiety. Ankieta została zrealizowana za pomocą formularzy Google. W ramach analizy wyników testów odsłuchowych w pierwszej kolejności sprawdzone zostały korelacje pomiędzy parami klas emocji przypisanych do katalogów utworów w bazie Epidemic Sound. Następnie zmapowano uzyskane wyniki ankiety na wykorzystany model, który zawiera emocje odniesienia (autorski model). W referacie zawarte też zostały histogramy zawierające przykłady list rankingowych przypisań kolorów do emocji. Przedstawiony również został zarys dalszych eksperymentów.

1. Wstęp

Muzyka filmowa stanowi ważny fragment przekazu artystycznego towarzyszącemu obrazowi. Jej celem jest podbudowanie lub wywołanie emocji związanych z daną sceną, ale przede wszystkim staje się wizytówką filmu, często określając jego gatunek. Klasyfikacja emocji w muzyce filmowej – w uproszczeniu – obejmuje zagadnienia związane z analizą muzyki, kolorystyki obrazu, narracji filmu oraz przekazu emocji, które mają towarzyszyć projekcji filmu. Celem pracy jest przedstawienie eksperymentów związanych z przeprowadzeniem testów subiektywnych określających emocje w muzyce filmowej. W pierwszej kolejności przygotowano założenia ankiety internetowej, która została następnie wykorzystana w testach subiektywnych. Ankiety skonstruowano z wykorzystaniem formularzy Google (z wykorzystaniem narzędzia "allocate monster" do dystrybucji). Opracowana ankieta jest dostępna pod adresem: <https://allocate.monster/RVCTOXP>.

1.1. Modele emocji

Modele emocji wykorzystywane w algorytmach klasyfikacji można podzielić na dwie grupy: kategoryjne (ang. *categorical*) i wymiarowe (ang. *dimensional*). Podejście kategoryjne zakłada, że w modelu znajdują się dyskretne deskryptory emocji, które mogą być podzielone na klasy. Modele wymiarowe definiują przestrzeń emocjonalną za pomocą osi, a każda emocja traktowana jest jako punkt w tej przestrzeni [1], [2].

Do modeli kategoryjnych dla przykładu można zaliczyć listę 66 deskryptorów podzielonych na 8 grup - są one wynikiem jednych z pierwszych badań nad ekspresyjnością w muzyce przeprowadzonych przez K. Hevner [3]. Innym przykładem może być model wykorzystywany w międzynarodowym konkursie MIREX, który zawiera 29 określeń podzielonych na 5 klastrów. Został on zaproponowany przez X. Hu oraz J. S. Downie [4]. Modele kategoryjne badane były również w kontekście muzyki (i etykiet w j. polskim) w pracach M. Plewy oraz B. Kostek [5], [6].

Model emocji Russela [7] jest przykładem modelu wymiarowego. Zakłada on, że emocje rozmieszczone są na płaszczyźnie, która zdefiniowana jest za pomocą dwóch wymiarów - walencji oraz pobudzenia. Model emocji Thayera [8] prezentuje się w podobny sposób, ale przestrzeń dwuwymiarowa aktywacji opisana jest w nim za pomocą dwóch rodzajów pobudzeń - energetycznego oraz napięciowego.

1.2. Rozpoznawanie emocji w muzyce

W literaturze można zauważyć zróżnicowane podejścia do zadania rozpoznawania emocji w muzyce. Istnieje też obszar badań (*Music Emotion Recognition*, MER) związany z klasyfikacją emocji w muzyce i nastroju, który muzyka może wywoływać [6], [9]. Wybór metody klasyfikacji jest zależny w dużej mierze od przyjętych założeń - emocje można bowiem rozpoznawać w kilkusekundowym fragmencie utworu muzycznego lub analizować go jako całość. Ważnym aspektem wymagającym uwzględnienia jest również sposób reprezentacji sygnału fonicznego, na podstawie którego opiera się będzie praca algorytmu. Utwory mogą być przedstawione w formie surowej (próbki dźwięku postaci czasowej [wavenet]), parametrycznej (np. wektor parametrów mel-cepstralnych czy parametrów wykorzystujących standard MPEG 7) [10], [11], [12], [13], [14], [15] lub też obecnie - w przypadku głębokiego uczenia - reprezentacji graficznej (spektrogram, chromagram) lub parametrów uzyskanych przez strukturę głęboką [16], [17], [18], [19], [20]. Kolejnym, ale bardzo istotnym elementem rozpoznawania emocji w muzyce, jest wybrany model emocji [9], [16], [21], [22]. Percepcja muzyki jest bardzo subiektywnym i indywidualnym procesem, co sprawia, że problem klasyfikacji emocji nie należy do trywialnych, zwłaszcza, że należy brać pod uwagę emocje, które zostały intencjonalnie zawarte przez kompozytora muzyki, jak również te, które towarzyszą odbiorcy muzyki i obrazu.

2. Eksperymenty

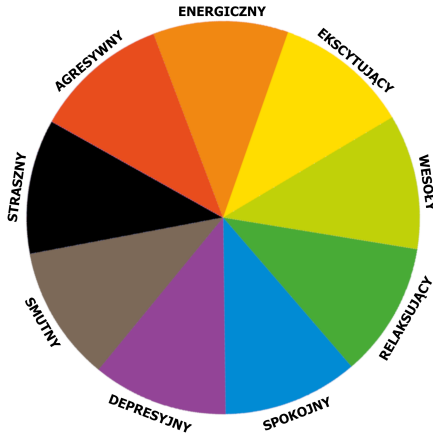
2.1. Testy subiektywne określające emocje w muzyce filmowej

W testach subiektywnych określających emocje zawarte w danym fragmencie muzyki wzięło udział 176 osób. Zadaniem słuchaczy było przypisanie danego fragmentu muzycznego do odpowiedniej klasy emocji. Założono, że w ankiecie internetowej (rys. 1) znajdują się takie klasy, jak: *dark, epic, floating, glamorous, mysterious, hopeful, laid back, euphoric, smooth, sneaking, dreamy, sentimental, sad, romantic, happy, sexy, angry, relaxed, suspense*, zgodnie z katalogami zawartymi w bazie Epidemic Sound (epidemicsound.com). Ponieważ wykorzystywane klasy są w niektórych przypadkach bliskoznaczne, dlatego w pierwszej kolejności sprawdzono korelacje pomiędzy parami klas emocji. Niektóre przypadki na podstawie analizy korelacyjnej oraz testu chi-kwadrat (wyniki statystycznie istotne) wskazały na bliskie powiązania pojęć. Dotyczy to m.in. etykiet: *dark-mysterious, dreamy-laid back, floating-relaxed*.

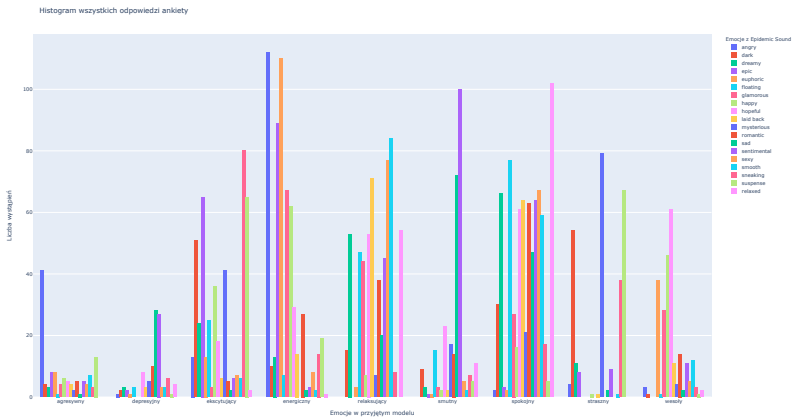
Rysunek 1 Strona wejściowa skonstruowanej ankiety internetowej oraz pytania dotyczące odsłuchiwane go fragmentu muzyki filmowej

2.2. Analiza wyników

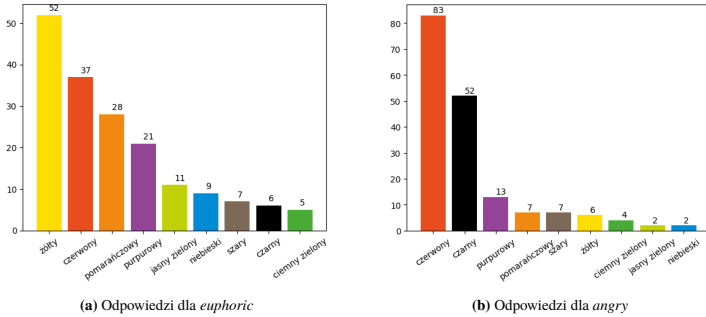
Wyniki uzyskane w ankiecie internetowej, w której zadaniem słuchaczy było przypisanie klasy emocji do danego utworu, przedstawiono na rys. 2. W dalszym kroku analiz zmapowano uzyskane wyniki na wykorzystany model, który zawiera emocje odniesienia (rys. 3): *energiczny, ekscytujący, wesół, relaksujący, spokojny, depresyjny, smutny, straszny, agresywny*. Histogramy przedstawione na rys. 4 zawierają przykłady list rankingowych przypisań emocji do kolorów.



Rysunek 2 Model emocji wykorzystywany w pracy



Rysunek 3 Wyniki uzyskane w ankiecie internetowej, w której zadaniem słuchaczy było przypisanie klasy emocji do danego utworu (zmapowanie klas z bazy Epidemic Sound na własny model)



Rysunek 4 Przykłady list rankingowych przypisania emocji do kolorów w postaci wykresów słupkowych

Na podstawie analiz histogramów można wyszczególnić pojęcia, które słuchacze wiążą ze sobą. Dla przykładu emocja *sad* budzi - w proporcji przypisać - następujące skojarzenia:

- *smutny* + *spokojny*: 39
- *depresyjny* + *smutny*: 36
- *relaksujący* + *spokojny*: 17

Z kolei, dla emocji *romantic* te typowania słuchaczy są następujące:

- *relaksujący* + *spokojny*: 34
- *smutny* + *spokojny*: 13
- *relaksujący* + *wesoły*: 13

3. Wnioski i dalsze prace

Na podstawie analiz można zauważyć, że pomimo iż występują rozbieżności w adnotacjach słuchaczy, większość typowań jest zgodna z oczekiwaniami. Głównym problemem - często wskazywanym w literaturze - jest trudność jednoznacznej definicji poszczególnych pojęć opisujących emocje. Niektóre z nich wydają się badanym osobom bliskoznaczne, zaś inne oddalone pojęciowo od siebie w sposób znaczący. Kolejnym etapem eksperymentów będzie klasyfikacja emocji z wykorzystaniem uczenia głębokiego. W tym celu zostanie wykorzystany algorytm sieci spłotowej (ang. CNN – *Convolutional Neural Network*). Ze względu na wymaganie związane z głębokim uczeniem, algorytm będzie wykorzystywał reprezentacje 2D sygnałów fonicznych. W danych wejściowych zostaną uwzględnione emocje z zaproponowanego modelu (rys. 2) oraz typowania słuchaczy.

Bibliografia

- [1] T. Eerola i J.K. Vuoskoski: „A comparison of the discrete and dimensional models of emotion in music”, w: *Psychology of Music* 39.1, 2011, s. 18–49, doi: [10.1177/0305733610362821](https://doi.org/10.1177/0305733610362821)
- [2] J. Grekow: *From Content-Based Music Emotion Recognition to Emotion Maps of Musical Pieces*, Warszawa, 2017
- [3] K. Hevner: „Experimental Studies of the Elements of Expression in Music”, w: *The American Journal of Psychology* 48.2, 1936, s. 246–268



- [4] X. Hu i J. S. Downie: „Exploring Mood Metadata: Relationships with Genre, Artist and Usage Metadata.”, w: *Proceedings of the 8th International Conference on Music Information Retrieval, ISMIR 2007*, 2007, s. 67–72
- [5] M. Plewa i B. Kostek: „Music Mood Visualization Using Self-Organizing Maps”, w: *Archives of Acoustics* 40.4, 2015, s. 513–525, doi: [10.1515/aaa-2015-0051](https://doi.org/10.1515/aaa-2015-0051)
- [6] B. Kostek i M. Plewa: „Rough Sets Applied to Mood of Music Recognition”, w: *Proceedings of the 2016 Federated Conference on Computer Science and Information Systems*, red. M. Ganza, L. Maciaszek i M. Paprzycki, t. 8, ACSIS, 2016, s. 71–78, doi: [10.15439/2016F548](https://doi.org/10.15439/2016F548)
- [7] J. Russell: „A Circumplex Model of Affect”, w: *Journal of Personality and Social Psychology* 39, 1980, s. 1161–1178, doi: [10.1037/h0077714](https://doi.org/10.1037/h0077714)
- [8] R. E. Thayer: *The Biopsychology of Mood and Arousal*, 1989
- [9] M. Barthelet, G. Fazekas i M. Sandler: „Music Emotion Recognition: From Content- to Context-Based Models”, w: *From Sounds to Music and Emotions. CMMR 2012. Lecture Notes in Computer Science*, red. M. Aramaki, M. Barthelet, R. Kronland-Martinet i Ystad S, t. 7900, Berlin, Heidelberg: Springer, 2013, doi: [10.1007/978-3-642-41248-6_13](https://doi.org/10.1007/978-3-642-41248-6_13)
- [10] B. Kostek i M. Plewa: „Parametrization and Correlation Analysis Applied to Music Mood Classification”, w: *International Journal of Computational Intelligence Studies* 2.1, 2013
- [11] B. Kostek i M. Plewa: „Testing a Variety of Features for Music Mood Recognition”, w: *166th Meeting Acoustical Soc. of America* 2.1, 2013, s. 3994
- [12] O. Lartillot i P. Toiviainen: „Mir in matlab (II): A toolbox for musical feature extraction from audio”, w: *Proceedings of the 8th International Society for Music Information Retrieval Conference, ISMIR 2007*, 2007, s. 127–130
- [13] R. Panda, B. Rocha i R. P. Paiva: „Music emotion recognition with standard and melodic audio features”, w: *Appl. Artif. Intell.* 29, 2015, s. 313–334, doi: [10.1080/08839514.2015.1016389](https://doi.org/10.1080/08839514.2015.1016389)
- [14] Y. Song, S. Dixon i M. Pearce: „Evaluation of musical features for emotion classification”, w: *Proceedings of the 13th International Society for Music Information Retrieval Conference, ISMIR 2012*, 2012, s. 523–528
- [15] F. Zhang, H. Meng i M. Li: „Emotion extraction and recognition from music”, w: *2016 12th Int. Conf. Nat. Comput. Fuzzy Syst. Knowl. Discov. ICNC-FSKD 2016*, 2016, s. 1728–1733
- [16] S. Hizlisoy, S. Yildirim i Z. Tufekci: „Music emotion recognition using convolutional long short term memory deep neural networks”, w: *Engineering Science and Technology, an International Journal* 24.3, 2021, s. 760–767, issn: 2215-0986, doi: [10.1016/j.jestech.2020.10.009](https://doi.org/10.1016/j.jestech.2020.10.009)
- [17] Y.C. Lin, Y.H. Yang i H.H. Chen: „Exploiting Online Music Tags for Music Emotion Classification”, w: *ACM Trans. Multimedia Comput. Commun. Appl.* 7S.1, 2011, issn: 1551-6857, doi: [10.1145/2037676.2037683](https://doi.org/10.1145/2037676.2037683)
- [18] T. Liu, L. Han, L. Ma i D. Guo: „Audio-based deep music emotion recognition”, w: *AIP Conference Proceedings* 1967.1, 2018, doi: [10.1063/1.5039095](https://doi.org/10.1063/1.5039095)
- [19] R. Sarkar, S. Choudhury, S. Dutta, A. Roy i S.K. Saha: „Audio-based deep music emotion recognition”, w: *AIP Conference Proceedings* 1967.1, 2018, doi: [10.1063/1.5039095](https://doi.org/10.1063/1.5039095)
- [20] H. Liu, Y. Fang i Q. Huang: „Music Emotion Recognition Using a Variant of Recurrent Neural Network”, w: *Proceedings of the 2018 International Conference on Mathematics, Modeling, Simulation and Statistics Application (MMSSA 2018)*, Atlantis Press, 2019, s. 15–18, isbn: 978-94-6252-661-7, doi: [10.2991/mmssa-18.2019.4](https://doi.org/10.2991/mmssa-18.2019.4)
- [21] M. Plewa i B. Kostek: „Multidimensional Scaling Analysis Applied to Music Mood Recognition”, w: *134 Audio Engineering Society Convention*, 2013, preprint 8876
- [22] M. Plewa i B. Kostek: „Creating Mood Dictionary Associated with Music”, w: *132 Audio Engineering Society Convention*, 2012, preprint 8607

