

Article

Not peer-reviewed version

Mel-Frequency Cepstral Coefficients and Recording Studio Features for the Analysis of Producer-Driven Music

[Tim Ziemer](#)*

Posted Date: 10 March 2026

doi: 10.20944/preprints202603.0656.v1

Keywords: music information retrieval; house music; techno music; electronic dance music; self-organizing maps; systematic musicology; recording studio technology; music production



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Mel-Frequency Cepstral Coefficients and Recording Studio Features for the Analysis of Producer-Driven Music

Tim Ziemer

University of Hamburg; tim.ziemer@uni-hamburg.de

Abstract

Computational methods for big data music research mostly come from the field of music information retrieval. Through feature extraction and machine learning, many practical tasks have been automated, like genre recognition and playlist generation. However, for musicological purposes, conventional features do not provide enough insight into the music production process. In this study, we evaluate how well Mel-frequency cepstral coefficients and recording studio features reveal aspects of early house and techno music from the United States of America and Germany. The explorative study is an exemplary case-study where music production plays an essential role. Further studies may reveal how much the findings transfer to other producer-driven music, like hip hop and electronic dance music.

Keywords: music information retrieval; house music; techno music; electronic dance music; self-organizing maps; systematic musicology; recording studio technology; music production

1. Introduction

In the field of Music Information Retrieval (MIR), big data audio analysis and retrieval are carried out to solve practical tasks, like automatic genre recognition or playlist generation. However, it has been observed that machines recognize patterns and transform input data to output data in ways that neither reflect how musicians make music, nor how listeners perceive it [1].

For the study of music, what is important is not the execution of a certain task but the identification of causal relationships. From explanatory features, we can derive how musicians made the music or how listeners perceive it. However, ubiquitous low-level features have been criticized as approximating stochastics inherent in a dataset instead of finding causal relationships between input and output [2–4]. They reveal no close relationship with the process of music production nor music perception. Thus, they are difficult to interpret and not explanatory.

For producer-driven music, like hip hop (cf. [5], p. 692) and dance music (cf. [6], p. 82), the dominance of the producer over artists has been attested (cf. [7]). For this type of music, it makes sense to analyze the sound qualities rather than traditional compositional aspects like melody and harmony.

Two respective sets of features have been proposed: Mel-Frequency Cepstral Coefficients (MFCCs) and recording studio features. The first are considered to relate to human timbre perception, the latter to the process of music production. Our aim is to explore a house and techno dataset with both feature sets to evaluate their strengths and weaknesses in explaining aspects of produced music in a case-study. The research is not hypothesis-driven but explorative.

The remainder of the paper is structured as follows: First, the two utilized feature sets are explained. Then, they are discussed based on previous work that utilized them. In Section 3, the analyzed dataset and the research method are described.

1.1. Feature Sets

The two feature sets are briefly described in this section. Details can be found in the cited literature.

1.1.1. Mel Frequency Cepstral Coefficients

Mel-Frequency Cepstral Coefficients (MFCCs) are ubiquitous in MIR [8]. They are extracted from an audio snippet in 3 steps: 1. a short time Fourier transform produces a linear frequency spectrum. 2. The frequencies are smeared over a Mel frequency scale. 3. The result is transformed to cepstral domain via a Discrete Cosine Transform (DCT). The outcome is a sort of spectrum of a spectrum, so together, the cepstral coefficients describe the (smeared) spectral distribution on the nonlinear Mel scale, which approximates human pitch perception of pure tones. Often, the first 13 MFCCs are considered.

The idea behind these transformations is that the MFCCs describe the characteristics of a resonator independent of the oscillator. This way, vowels and consonants are recognized independent of a speaker's pitch, which is why MFCCs work well in speech recognition task.

1.1.2. Recording Studio Features

In the recording studio, music producers and audio engineers complement close listening by consulting audio monitoring tools. Such tools include the peak meter, Root Mean Square (RMS) meter, Volume Unit (VU) meter, crest factor meter, the phase scope and channel correlation meter. Details and example graphics can be found in [9]. These tools are often applied on single tracks and on the master output, and sometimes in third-octave bands. In addition, the Beats Per Minute (BPM) of a song are set, not analyzed, by the music producer. The results are presented in Section 4 and discussed in Section 5. The conclusion section, Section 6, summarizes the results and gives an outlook on future work.

2. Previous Work

Both feature sets have been utilized and evaluated before. [10] criticize that MFCCs, originally designed for speech analysis, have been utilized for MIR tasks without justification. However, listening tests with synthesized sounds in [11] indicated that each MFCC has a close, linear relationship with the perception of sound color (not timbre, see [12] – for a comprehensive distinction, see [13]). They conclude that because small changes of a sound with respect to a single MFCC result in small dissimilarity ratings in listening tests, while large changes result in large dissimilarity ratings. How the results can be transferred to music pieces with a number of simultaneous complex timbres remains unclear. [3] acknowledge that MFCCs consider aspects of auditory perception, like using the psychoacoustic Mel scale. At the same time, they criticize that many parameter decisions are justified through computer science and not through perception, like the smearing and the DCT. And interpreting MFCCs is difficult, because the set of MFCCs describe frequency distribution, but the single MFCC has no perceptual meaning. [14] showed that the loudness war in popular music could be explained through recording studio features like RMS and crest factor. [15] showed that the vector scope data alone, i.e., PhaseSpace and ChannelCorrelation, can be sufficient for a genre recognition task. [16,17] put the phase scope in relation to both the music production process and the perception of spaciousness. [9] extracted recording studio features from over 1,200 tracks. Using a random forest classifier, they could predict which out of 10 famous producers played which track with an accuracy of 63%, not taking bpm into account. Some features, like the VU meter and RMS, correlate with each other.

In [18], we could show how recording studio features describe the divergent paths of US and German house and techno music from the viewpoint of music production and mixing.

In [19], we analyze music by three hip hop producers. We found that each producer had a unique sound profile, which was more distinct in terms of the phase scope feature than MFCCs. Yet, in a genre classification task, [20] found that MFCCs have a higher discriminative power than BPM and RMS (which belong to the set of recording studio features). Likewise, [21] perform well in a genre classification task based exclusively on MFCCs.

Overall, MFCCs and recording studio features have proven their value. The strength of MFCCs is, supposedly, that they have high discriminative power. The strength of recording studio features seems

to lie in their close relationship with music mixing. In this study, we compare how the two feature sets perform in music analysis tasks.

Our terminology is based on [12], i.e., we clearly distinguish between perceptual and physical sound aspects.

3. Method

This section describes the analyzed dataset, followed by the two analysis methods and their purposes.

3.1. The Dataset

The HOTGAME dataset [22], contains recording studio features extracted from about 9,000 early German and US house and techno tracks from 1984 to 1994. For the present study, we extracted Mel Frequency Cepstral Coefficients (MFCCs) from the same tracks. Analyses showed that some features correlate with one another, so only a subset of non-correlating features is utilized in this study. For the recording studio features, these are the 5 features Phase Scope, Channel Correlation, Crest Factor, RMS, and BPM. For the MFCC set, these are the 4 Coefficients MFCC2, 9, 12, and 13.

Each track (i.e. music piece) is represented by the median value of the extracted features, which form the input vector for a Machine Learning (ML) method.

3.2. Random Forest

For a quantitative comparison, tracks are assigned to 9 German house/techno styles (Breakbeat, Downbeat, Eurodance, Happy Hardcore, Hardcore, Hardtrance, House, Tekkno, Trance) and 9 US styles (Acid House, Chicago House, Deep Downbeat, House, Detroit Techno (1st Generation), Detroit Techno (2nd Generation), Garage House, Hardcore, Hip House). Using a Random Forest classifier (RF), we compare the classification performance of both feature sets. This approach quantifies the discriminative power of the feature set.

RF is a supervised classifier that uses a decision tree for a subset of training data. Here, the decision is what magnitude range of each feature represents the class (in our case the style) best. Using many subsets, many trees form a forest. For the classification, all trees classify an input track. The commonest answer is the final classification result. This approach is comparable to the "Ask the Audience" lifeline in "Who Wants to Be a Millionaire". RFs are frequently applied in music analysis task, like [9,23,24].

3.3. Self-Organizing Maps

Using Self-Organizing Maps (SOMs), we explore how much insight both sets of features provide for the distinction of the two nations, qualitatively through visual inspection. This approach helps to understand if and in what respects German and US tracks are different.

SOMs are a type of neural network with just one layer, the output layer. SOMs represent the distribution of tracks not in its original, high-dimensional feature vector space, but on a 2-dimensional map that preserves the original topology of the data distribution. Details on SOMs can be found in the inventor's monograph [25]. The output layer is trained in an unsupervised manner. Each item is presented to the output layer and changes its condition such that similar tracks (according to the features) are placed close to each other and distinct tracks are further away. In what respect tracks are (dis-)similar can be seen in the Component Planes. These show each feature's magnitude over the map. Including all component planes in one has recently been proven possible by means of sonification rather than visualization [26]. SOMs are frequently used for data exploration and analysis in musicology, see e.g. [8,19,27].

Both ML methods are not the most sophisticated approaches, but in contrast to deep learning, they relate closely to the input features and no drastic, nonlinear transforms.

4. Results

The results section is divided in the Random Forest and the Self-Organizing Map subsection.

4.1. Random Forest Classifier

The average training and test accuracy (\pm standard deviation) for the MFCCs and the recording studio feature set are listed in Table 1. In all cases, the training accuracy was 0.98 ± 0.02 to 0.99 ± 0.02 , and the test accuracy is higher than chance ($1/9 = 0.11$). The classification accuracy based on MFCCs is 0.22 for German and 0.28 for us styles. When adding bpm as a feature, the accuracy rises to 0.51 and 0.39, respectively. This highlights the discriminative power of bpm in our dataset, so it is fair to consider results with and without the bpm feature.

Table 1. Mean test accuracy \pm standard deviation for 9 German (G) and US (U) house and techno styles for a RF model based on MFCCs and Recording studio features.

	MFCCs	MFCCs+bpm	Rec	Rec+bpm
G	0.22 ± 0.02	0.51 ± 0.02	0.19 ± 0.02	0.52 ± 0.02
U	0.28 ± 0.02	0.39 ± 0.02	0.25 ± 0.01	0.37 ± 0.02

One could argue that the strength of MFCCs is their collective discriminative power. Therefore, another RF is trained with MFCC1–7, 10 and 11 (i.e., excluding only those that exhibit a very high correlation with either of the 9, e.g., MFCC10 and 11 correlate with MFCC12 and 13). The results are summarized in Table 2. German styles are classified with an accuracy of 0.54 and US styles with an accuracy of 0.39, i.e., as good as the 4 MFCCs plus bpm. When adding bpm as a feature, the accuracy rises to 0.57 for German styles and 0.59 for US styles.

Table 2. Mean test accuracy \pm standard deviation for 9 German (G) and US (U) house and techno styles for a RF model based on 9 MFCCs and 9 Recording studio features.

	MFCCs	MFCCs+bpm	Rec	Rec+bpm
G	0.54 ± 0.02	0.57 ± 0.02	0.36 ± 0.02	0.56 ± 0.02
U	0.39 ± 0.02	0.59 ± 0.02	0.30 ± 0.02	0.43 ± 0.02

The classification accuracy based on the recording studio features PhaseSpace, ChannelCorrelation, RMS, and CrestFactor (without bpm) with 0.19 for both nations is similar to the accuracy based on 4 MFCCs. When replacing RMS with bpm, the accuracy rises to 0.52 and 0.37. Obviously, bpm is a strong feature. Apart from that, both feature sets serve similarly well for a style classification task. It seems that the US styles are less distinct than the German styles.

Since the performance of 9 MFCCs has been analyzed as a collection, a collection of 9 recording studio features has been analyzed, too. The collection contains the features after applying band-pass filters to analyze the low (20–150 Hz), mid (150–2,000 Hz) and high (2,000–10,000 Hz) frequencies. The result is an accuracy of 0.36 for German and 0.30 for US styles without considering bpm, and 56 and 43 when including bpm. So indeed, the MFCCs as a collection have a higher discriminative power than a collection of recording studio features.

4.2. Self-Organizing Map

Figure 1 shows the distribution of German and US tracks on a Self-Organizing Map (SOM). Tracks from both nations distribute over the complete SOM. There is a small region in the upper-right corner where the population of US tracks is higher, and a small region on the upper-left, where you can find more German tracks. Component planes show that the upper-right corner exhibits a high MFCC2, 9, and 12, and a low MFCC13, while the upper-left corner has a high MFCC13 value.

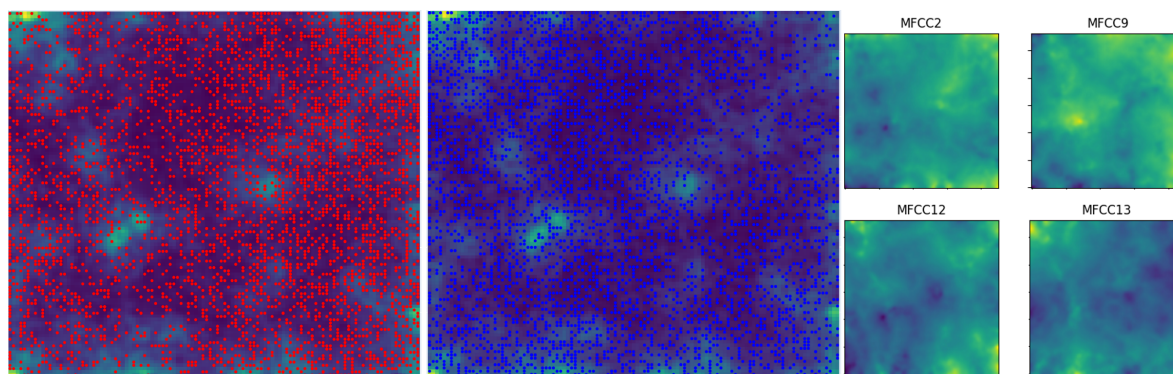


Figure 1. US (red) and German (blue) tracks on a SOM, and the component planes of MFCC2, MFCC9, MFCC12, and MFCC13.

Adding the bpm feature tears the nations apart. In Figure 2, US tracks densely occupy the complete map, except for a semicircular region on the right, where many German tracks can be found. Again, the component planes highlight the larger MFCC13-value in the German region, but also a high bpm value. However, the largest bpm values are on the upper-left, where German and US tracks are found.

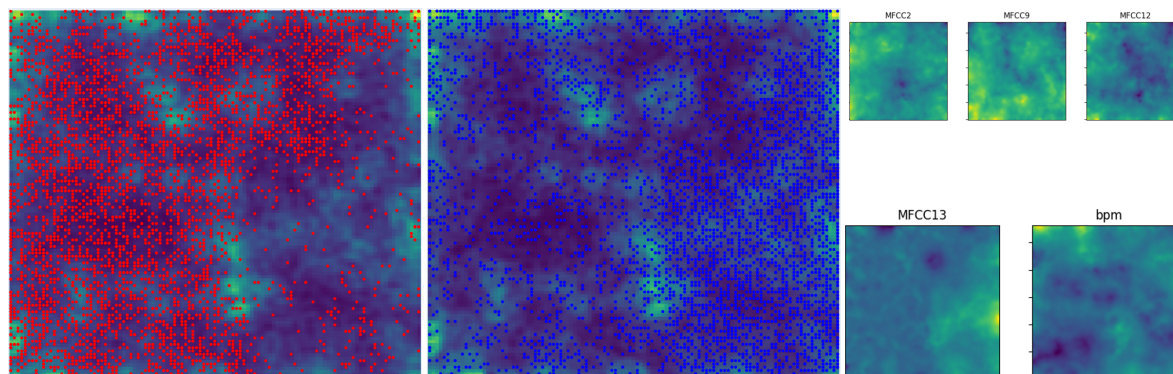


Figure 2. US (red) and German (blue) tracks on a SOM, and the component planes of MFCC2, MFCC9, MFCC12, MFCC13, and bpm.

As for the RF, Figure 3 shows the distribution of US and German tracks when the SOM is trained with the 9 MFCCs. Here, the US tracks exhibit the densest population on the right edge, where the German tracks are more sparsely populated. This region has low MFCC4 and 6 values and a high MFCC11 value.

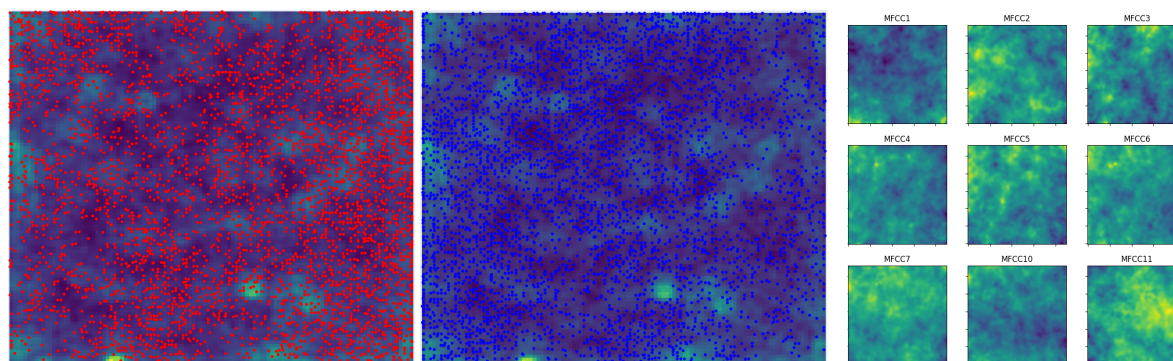


Figure 3. US (red) and German (blue) tracks on a SOM, and the component planes of MFCC2-7, MFCC10, and 11.

Figure 4 shows the distribution of German (blue) and US (red) tracks on a SOM based on recording studio features Phase Space, ChannelCorrelation, RMS and CrestFactor. The American tracks occupy almost the whole map, but a triangular region on the lower-right and the upper-left are sparsely

populated. The opposite is observable for the German tracks. The U-matrices show that US region has a higher PhaseSpace value and lower CrestFactor, indicating a louder mix with more dynamic compression.

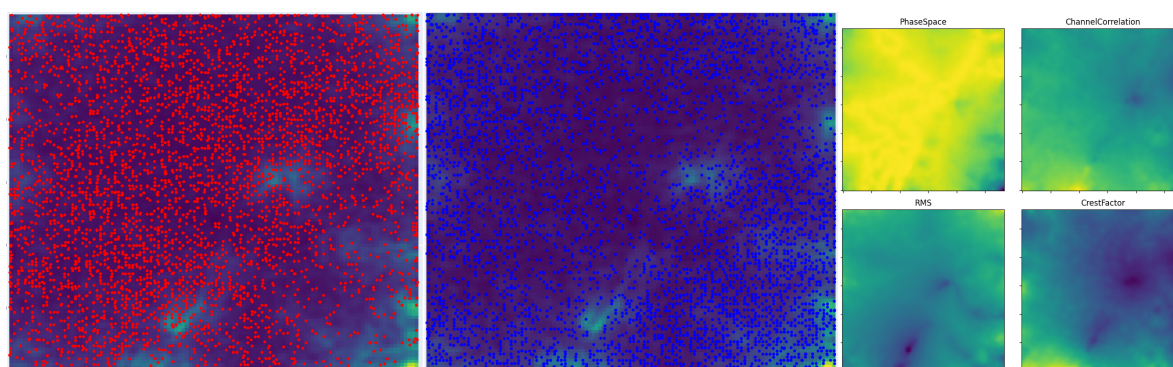


Figure 4. US (red) and German (blue) tracks on a SOM, and the component planes of PhaseSpace, ChannelCorrelation, RMS, and CrestFactor.

When replacing the RMS with bpm, the distinction between the two becomes even stronger, as observable in Figure 5: A “C”-shaped US region lies in the middle, while most German tracks lie on the right-hand side, in the lower-left corner and in a small upper-left corner. The bpm component plane shows that the tempo is often higher and partly lower than the US tracks. On this SOM, it is clearly visible that the CrestFactor is also higher in the German regions.

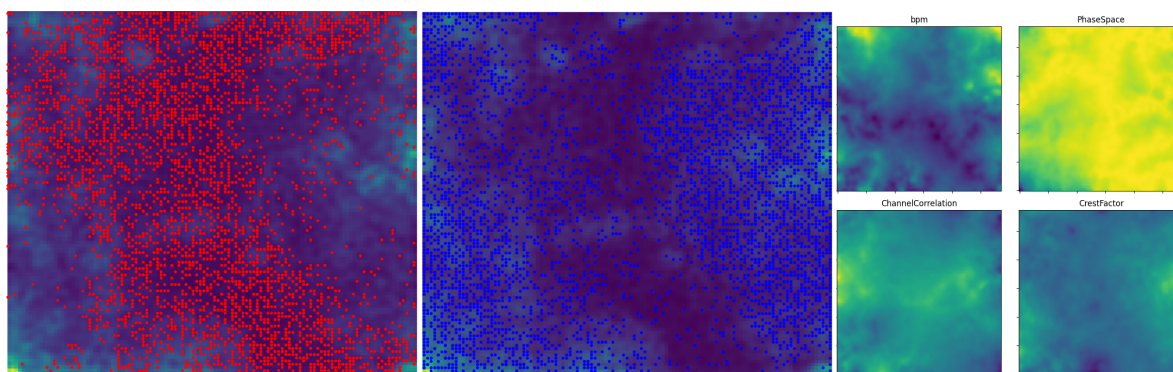


Figure 5. German (blue) and US (red) tracks on a SOM, and the component planes of bpm, PhaseSpace, ChannelCorrelation, and CrestFactor.

Figure 6 shows a SOM based on 9 recording studio features. Here, the features are not only extracted from the broadband signal, but also from the low-, mid- and high-frequency region. Only those features that do not correlate with each other have been chosen. One can clearly see a separation, with mostly German tracks at the bottom and in a circular region at the upper-left. So clearly, it is not (only) the bpm feature that separates the music from the two nations. It is the music mixing style that is represented by recording studio features. The bpm is a strong feature, but when enough recording studio features are considered, they can segregate the two nations even without taking bpm into account.

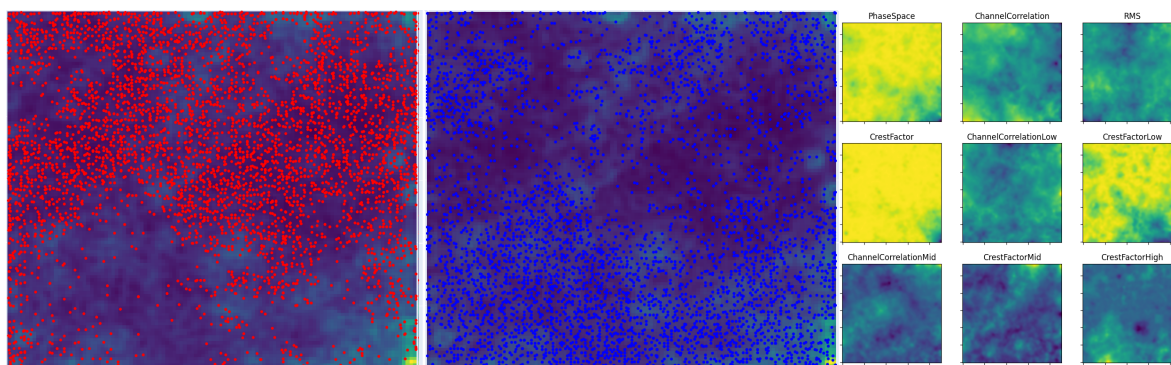


Figure 6. German (blue) and US (red) tracks on a SOM, and the component planes of PhaseSpace, ChannelCorrelation, RMS and CrestFactor of the complete bandwidth and filtered lows, mids and heights.

5. Discussion

Clearly, the BPM is a strong feature for the analysis of early house and techno music. Including it improves the style classification accuracy and the distinction of the nations a lot.

Results of the RF classification indicate US styles are less distinct than German styles regarding both MFCCs and recording studio features. The classification accuracy is much lower than in many genre recognition studies, like [20]. However, the objective was to compare the discriminative power of the two feature sets, so standardized machine learning was used instead of an optimized approach for the classification task at hand. Moreover, it has been shown that style (or subgenre) classification is more discerning than genre classification, as the differences are more subtle (cf. [23,24]). MFCCs have a slightly higher discriminative power than recording studio features. The assumption MFCCs have a higher discriminative as a collection proves true. They perform similarly well as the recording studio features including bpm, and much better than recording studio features without bpm.

The inspection of the SOMs shows that MFCC13 seems to play a role for the distinction of both nations. But the regions on the SOM mostly overlap. The situation improves only slightly when considering 9 MFCCs. The recording studio features show a stronger separation of both nations. Here, the PhaseSpace and CrestFactor seem important. The results are in accordance with [19] who found that hip hop producers have a certain sound profile, which is quite unique in terms of recording studio features, but with much overlap in terms of MFCCs. A SOM based on 9 recording studio features shows that they segregate US and German music well, even without inclusion of bpm.

For both feature sets, adding the bpm improves the discrimination of both nations a lot, as the German tracks are mostly faster.

6. Conclusion

In this paper, we analyzed about 9,000 house and techno tracks based on Mel Frequency Cepstral Coefficients (MFCCs) and recording studio features.

Both feature sets are informative, so that a Random Forest (RF) classifier can transform them to classify styles with fair accuracy. Overall, classification based on MFCCs is a bit more accurate than based on recording studio features, if we exclude bpm. When bpm is acknowledged as a recording studio feature, the classification is much more accurate than classification based on MFCCs alone. But when we consider the collective discriminative power of 9 MFCCs, they perform similarly well as the recording studio features including bpm.

When inspecting the feature sets visually using Self-Organizing Maps (SOMs), both feature sets show rather different qualities: MFCCs reveal some differences between early US and German house/techno music, particularly concerning MFCC13. What this means in terms of sound color or music production is unclear. When training the SOM with 9 MFCC components, the two nations separate a bit more, due to MFCC4, 6 and 11. Recording studio features reveal differences between the nations: The higher PhaseSpace value and lower CrestFactor of US tracks indicates a louder mixing with more dynamic compression.

Overall, MFCCs do not seem superior to recording studio features for discriminating different house/techno styles or nations. But the benefit of recording studio features is that they relate to the music production and mixing process, which makes them more explanatory, or at least easier to interpret.

A more comprehensive examination of the data (e.g., a confusion matrix of the RF, and a SOM-analysis by year-of-production) is necessary to find out more about the relationships of the features and the music. Moreover, it is worth studying how the combination of both feature sets can improve the discriminative performance, and how much it degrades the interpretability of the results. More case studies with other data sets and different types of music are necessary to see whether the observations made can be generalized. Lastly, music analysis should always relate computational analyses with the music as a cultural phenomenon, i.e., the scene.

Acknowledgments: I thank Simon Linke, Rolf Bader, Michael Blaß and the students from my Techno seminar.

References

1. Morreale, F.; Martinez-Ramirez, M.A.; Masu, R.; Liao, W.; Mitsufuji, Y. Reductive, Exclusionary, Normalising: The Limits of Generative AI Music. *Transactions of the International Society for Music Information Retrieval* **2025**, *8*, 300–312. <https://doi.org/10.5334/tismir.256>.
2. Sturm, B.L. A simple method to determine if a music information retrieval system is a 'horse'. *IEEE. Trans. Multimedia* **2014**, *16*, 1636–1644.
3. Aucouturier, J.J.; Bigand, E. Mel Cepstrum & Ann Ova: The difficult dialog between MIR and music cognition. In Proceedings of the Proceedings of the International Society for Music Information Retrieval Conference, 10 2012, pp. 397–402.
4. Ziemer, T.; Yu, Y.; Tang, S. Using Psychoacoustic Models for Sound Analysis in Music. In Proceedings of the Proceedings of the 8th Annual Meeting of the Forum on Information Retrieval Evaluation, New York, NY, USA, 2016; FIRE '16, pp. 1–7. <https://doi.org/10.1145/3015157.3015158>.
5. Farinella, D.J. Rick Rubin. In *The Encyclopedia of Record Producers*; Olsen, E.; Verna, P.; Wolff, C., Eds.; Billboard Books, 1998.
6. Hawkins, S. Feel the beat come down: house music as rhetoric. In *Analyzing Popular Music*; Moore, A.F., Ed.; Cambridge University Press: Cambridge, 2009; chapter 5, pp. 80–102. <https://doi.org/10.1017/CBO9780511482014>.
7. Wilke, T. Disco. In *Handbuch Popkultur*; Hecken, T.; Kleiner, M.S., Eds.; J. B. Metzler: Stuttgart, 2017; pp. 67–72.
8. Knees, P.; Schedl, M. *Music Similarity and Retrieval. An Introduction to Audio- and Web-based Strategies*; Springer, 2016.
9. Ziemer, T.; Kiattipadungkul, P.; Karuchit, T. Acoustic features from the recording studio for Music Information Retrieval Tasks. *Proceedings of Meetings on Acoustics* **2020**, *42*, 035004, [<https://asa.scitation.org/doi/pdf/10.1121/2.0001363>].
10. Peeters, G. The Deep Learning Revolution in MIR: The Pros and Cons, the Needs and the Challenges. In Proceedings of the Perception, Representations, Image, Sound, Music; Kronland-Martinet, R.; Ystad, S.; Aramaki, M., Eds., Cham, 2021; pp. 3–30. https://doi.org/10.1007/978-3-030-70210-6_1.
11. Hiroko Terasawa, Jonathan Berger, S.M. In Search of a Perceptual Metric for Timbre: Dissimilarity Judgments among Synthetic Sounds with MFCC-Derived Spectral Envelopes. *J. Audio Eng. Soc.* **2012**, *60*, 674–685.
12. Ziemer, T. Sound Terminology in Sonification. *J. Audio Eng. Soc.* **2024**, *72*, 274–289.
13. Schneider, A. Perception of Timbre and Sound Color. In *Springer Handbook of Systematic Musicology*; Bader, R., Ed.; Springer: Berlin, Heidelberg, 2018; chapter 32, pp. 687–726.
14. Emmanuel, D.; Damien, T. About dynamic processing in mainstream music. *Journal of the Audio Engineering Society* **2014**, *62*, 42–55.
15. Ziemer, T. Goniometers are a Powerful Acoustic Feature for Music Information Retrieval Tasks. In Proceedings of the DAGA 2023 – 49. Jahrestagung für Akustik, Hamburg, Germany, 2023; pp. 934–937.
16. Stirnat, C.; Ziemer, T. Spaciousness in Music: The Tonmeister's Intention and the Listener's Perception. In Proceedings of the KLG 2017. klingt gut! 2017 – International Symposium on Sound, Hamburg, Germany, Jun 2017; pp. 42–51.

17. Ziemer, T. Source Width in Music Production. Methods in Stereo, Ambisonics, and Wave Field Synthesis. In *Studies in Musical Acoustics and Psychoacoustics*; Schneider, A., Ed.; Springer: Cham, 2017; pp. 299–340.
18. Ziemer, T.; Linke, S. From Imitation to Innovation: The Divergent Paths of Techno in Germany and the USA. *arXiv-Preprint* **2025**.
19. Ziemer, T.; Kudakov, N.; Reuter, C. Producer vs. Rapper: Who Dominates the Hip Hop Sound? *Journal of the Audio Engineering Society* **2024**, *73*, 54–62.
20. Baniya, B.K.; Lee, J.; Li, Z.N. Audio Feature Reduction and Analysis for Automatic Music Genre Classification. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, San Diego, CA, USA, 2014; pp. 457–462.
21. Lee, C.H.; Shih, J.L.; Yu, K.M.; Lin, H.S.; Wei, M.H. Fusion of Static and Transitional Information of Cepstral and Spectral Features for Music Genre Classification. In Proceedings of the IEEE Asia-Pacific Services Computing Conference, Washington, DC, NW, USA, 2008; pp. 751–756. <https://doi.org/10.1109/APSCC.2008.95751>.
22. Ziemer, T. HOTGAME: A Corpus of Early House and Techno Music from Germany and America. *Metrics* **2025**, *2*, 8. <https://doi.org/10.3390/metrics2020008>.
23. Caparrini, A.; Arroyo, J.; Pérez-Molina, L.; Sánchez-Hernández, J. Automatic subgenre classification in an electronic dance music taxonomy. *Journal of New Music Research* **2020**, *49*, 269–284. <https://doi.org/10.1080/09298215.2020.1761399>.
24. Popli, C.; Pai, A.; Thoday, V.; Tiwari, M. Electronic Dance Music Sub-genre Classification Using Machine Learning. In Proceedings of the Artificial Intelligence and Sustainable Computing; Pandit, M.; Gaur, M.K.; Rana, P.S.; Tiwari, A., Eds., Singapore, 2022; pp. 321–331. https://doi.org/10.1007/978-981-19-1653-3_25.
25. Kohonen, T. *Self-Organizing Maps*, 3 ed.; Springer: Berlin, Heidelberg, 2001. <https://doi.org/10.1007/978-3-642-56927-2>.
26. Linke, S.; Ziemer, T. SOMson – Sonification of Multidimensional Data in Kohonen Maps. In Proceedings of the International Conference on Auditory Display, Troy, NY, 2024; pp. 50–57.
27. Blaß, M.; Bader, R. Content-Based Music Retrieval and Visualization System for Ethnomusicological Music Archives. In *Computational Phonogram Archiving*; Bader, R., Ed.; Springer: Cham, 2019; pp. 145–173. https://doi.org/10.1007/978-3-030-02695-0_7.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.