"Hardness" as a semantic audio descriptor for music using automatic feature extraction

Isabella Czedik-Eysenberg[1], Denis Knauf[2] and Christoph Reuter[1]

Abstract: The quality of "hardness" in music is an attribute that is most commonly associated with genres like metal or hard rock. However, other examples of music raise the question of whether there is a genre-independent general dimension of "hardness" that can be obtained from the signal automatically based on psychoacoustical features. In listening experiments 40 subjects were asked to rate 62 music excerpts according to their hardness. Using MATLAB toolboxes, a set of features covering spectral and temporal sound properties was obtained from the stimuli and investigated in terms of their correlation with the subjective ratings. By means of multiple linear regression analysis a model for musical hardness was constructed which shows a correlation of r = 0.86 with the experimental results. This proposes musical hardness as a useful high level descriptor for analysing collections of music. In ongoing experiments the fitness of this model is being further evaluated.

Keywords: Music, Semantic Audio Feature Extraction, Hardness, Heaviness, Metal, High Level Descriptor.

1 Background

Studies concerning "hard" (or "heavy") music - in most cases heavy metal - often focus on its sociological aspects or psychological role, for example investigating a connection between listening preferences and aggressive behavior amongst adolescents [We05], personality and modulation of emotions [Ge11] or its subcultural environment in general [We91] [Wa93] [Gr90]. These studies only briefly touch on the question of the actual sound properties that mark the examined music as "hard".

When reviewing articles about the production of corresponding music and other descriptive texts, a number of characteristic features can be found. Among them are strongly distorted guitars [Wa93] [Be99], a high intensity of low or high frequency ranges respectively [Re08] [BF05] [My12], high loudness in connection with a low dynamic range [We91] [Wa93], in particular a flat dynamic envelope caused by sound distortion [BF05], pronounced percussive sounds [Gr90], a distinct noise character of the vocal timbre [WBG11], ambiguous tonality with harmonic dissonances [Be99] as well as a particularly fast or slow tempo [WBG11].

Although this quality of music is often equated with the genre metal [Re08], other musical

[1] University of Vienna, Institute of Musicology, Spitalgasse 2-4, Hof 9, 1090 Vienna, Austria, isabella.czedik-eysenberg@univie.ac.at, christoph.reuter@univie.ac.at
[2] Technical University of Vienna, Software Engineering, denis@denkn.at

doi:10.18420/in2017_06

styles are also sometimes considered as "hard" like e.g. hard rock, hardcore techno, punk or "New German Hardness". This raises the question of whether there are common psychoacoustical features affecting the perception of hardness in music.

Therefore the goal is to use a computational approach in order to identify acoustic signal properties connected to this perceptive dimension. By examining sound features across different genres, a general model for musical hardness is aimed for. The concept of designing a high-level music descriptor on the basis of low-level audio features (cf. e.g. [BEL03]) offers a way of describing and extracting a semantically meaningful dimension that can be used in automatic music analysis, similarity measurement, and other musicological applications. Finally, the subjective character of the hardness ratings is investigated by determining whether and to what extent the ratings of listeners with a preference for hard music differ from those of other subjects.

2 Method

2.1 Listening test

An online listening test was conducted in which 40 subjects took part. Sixty-two music excerpts with a length of 10 seconds from a set of different genres were used as stimuli. A detailed listing of the used music pieces and their corresponding genres is provided as additional material on a webpage [CKR17]. Participants were asked to rate each of them on a seven-level scale according to their perceived hardness.

Next to basic descriptive data (age, gender) the listening habits and personal preference for hard music were queried in a general questionaire. The 40 participants were 18 to 59 years old (mean age = 31.08) and consisted of 15 female and 25 male subjects.

Finally, participants were also asked to provide a written description of the features that are relevant for the perception of musical hardness from their point of view. The test was carried out in a specifically developed web environment [Pr17] using Ruby on Rails.

2.2 Sound analysis

Sounds were analyzed in MATLAB using the TSM toolbox [DM14], the MIRtoolbox [LT07] and the Loudness Toolbox by Genesis [Ge17]. The performed signal analysis included a large set of spectral as well as temporal and other features (see [CKR17]), e.g. frequency distribution measures, dynamic envelope parameters and ranges, loudness values, timbral features like roughness and inharmonicity as well as information about percussive and harmonic signal components. All time-dependent features were aggregated by averaging over time, except were otherwise stated. These features were examined in terms of their correlation with the results of the listening test.

3 Results

3.1 Identified sound features

After summarizing synonymous categories, the descriptions given by the test participants were ranked according to their frequency of mention (see Tab. 1).

Feature	Frequency of mentions
High tempo	26
Characteristic singing style	17
Not very melodical	15
High loudness	14
Presence of percussive instruments	14
Dominant bass	13
Distortion	9
Specific guitar riffs	8
Noise-like	6
Temporal density of sound	6

Tab. 1: Characteristic features of hard music according to written descriptions by the test participants

Based on that, a series of signal parameters, some of which are associated with the mentioned psychoacoustic features, were extracted. For a full listing of all examined features with explanations see [CKR17]. Subsequently, correlations between these descriptor values and the mean ratings of each musical stimulus were analyzed. Tab. 2 summarizes the strongest connections.

Sound feature	r	p
Percussive Energy (E_{perc})	0.81	< 0.001
Spectral Flux (Median) (SF)	0.80	< 0.001
Roughness (R)	0.75	< 0.001
Number of Onsets (NoO)	0.68	< 0.001
High Frequency Ratio (HFR)	0.59	< 0.001
Loudness (Sone) (L)	0.54	< 0.001
Low Centroid Rate (LCR)	-0.52	< 0.001
2-4 kHz Energy ($E_{2\text{-}4kHz}$)	0.51	< 0.001
Envelope Flatness (EF)	0.50	< 0.001
Low Frequency Ratio (LFR)	0.48	< 0.001
Inharmonicity (I)	0.25	0.0484

Tab. 2: Correlation between hardness ratings and acoustic signal properties, sorted by the level of correlation. (For a full listing of features see [CKR17])

A feature that proves to be particularly promising in this context is the percussive energy, which describes the intensity of the percussive components of the signal. For extracting this parameter the procedure for harmonic-percussive separation as implemented in the TSM toolbox [DM14] using median filtering according to Fitzgerald [Fi10] was applied to the audio signal as a first step. This approach relies on the observation that percussive components tend to manifest as vertical structures in a spectrogram [DM15, Fi10, On08]. After separation, the RMS energy of the obtained percussive signal part was calculated. The resulting value is highly correlated with the perceived hardness of the musical stimuli. This is in line with descriptions that mention intense percussion sounds as a central element of hard music (e.g. [Re08] or "pounding percussion" in [Gr90]). These quickly changing percussive components in the spectrogram also reflect in a strong correlation with the spectral flux.

Additionally, for hard music, high as well as low frequency bands appear to be more pronounced compared with the medium frequency range (see high frequency ratio and low frequency ratio in Tab. 2). In order to obtain the high frequency ratio and low frequency ratio, the signal parts below 100 Hz and above 1000 Hz respectively were extracted using low/high pass filters and compared with the 250 to 400 Hz band in respect to their RMS energy. This correlation is in accordance with the results of Berger and Fales, who showed that the perception of hardness in guitar timbres increases with their content of high frequency energy [BF05]. Of note is the correlation with the spectral energy in the area between 2 and 4 kHz, as this frequency band has been shown to be associated with the perception of unpleasant sounds (like e.g. scraping on a blackboard) due to outer ear resonances [ROM14].

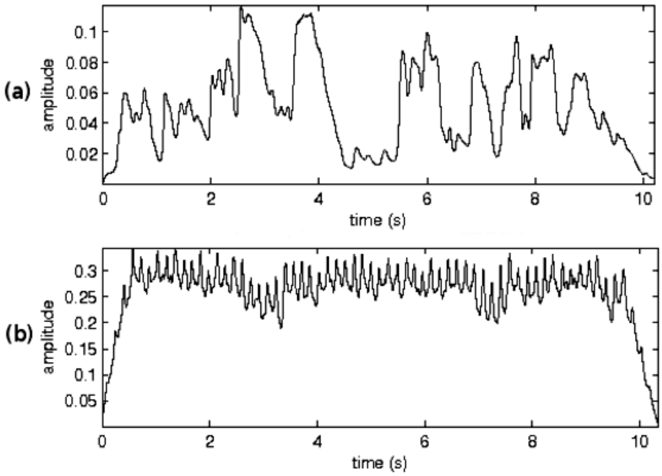

Fig. 1: Dynamic envelope of the (a) example that was rated least hard ("Cat Stevens - Sad Lisa") and the (b) example rated the hardest ("Marduk - Slay the Nazarene")

Among the most frequently mentioned features by participants of the listening test were "high tempo" as well as "temporal density of sound". The significance of these attributes manifests in a correlation with the number of onsets. Onsets were detected based on the temporal envelope using the MIRtoolbox [LT07]. Then, the number of onsets in relation to the length of the music excerpt was determined. The denser the onsets, the higher hardness ratings could be observed.

Another relevant feature is the loudness (Sone). Here it was measured according to the ANSI 2007 norm by Moore et al. [MGB97] as implemented in the Loudness Toolbox by Genesis [Ge17]. This confirms reports of high loudness in hard music (e.g. [Wa93] and [We91]).

As seen in the comparison between the music excerpts with the lowest and highest hardness rating respectively (Fig. 1), not only a higher fundamental amplitude can be observed for hard music, but also a generally flatter dynamic envelope (as previously described in [BF05]) characterized by dense peaks and relatively small amplitude drops. This constant (percussive) event density and the accompanying more uniform signal properties also manifest in correlations with the number of onsets, the envelope flatness and the low centroid rate (the percentage of frames showing a less than average spectral centroid value).

3.2 Construction of High Level Descriptor

Considering intercorrelations between the extracted descriptors (see Tab. 3), a combination of percussive energy (Eperc), the intensity of the signal components between 2 and 4 kHz (E2-4kHz) and the low centroid rate (LCR) was chosen as an effective set for describing the overall hardness (Fig. 2). For choosing these three parameters, a sequential feature selection approach was applied: Starting by adding the feature showing the highest correlation with the ratings, all other features were tested for partial correlations given the already added controlling variable. The feature with the strongest partial correlation was then added to the set. This process was performed iteratively until no significant (partial) correlation could be found any more.

Feature	E_{perc}	SF	R	NoO	HFR	L	LCR	$E_{2\text{-}4kHz}$	EF	LFR	I
E_{perc}	1	0.98	0.92	0.73	0.55	0.62	-0.47	0.37	0.43	0.49	0.30
SF	0.98	1	0.90	0.74	0.53	0.64	-0.50	0.34	0.41	0.55	0.25
R	0.92	0.90	1	0.63	0.69	0.52	-0.33	0.51	0.39	0.40	0.21
NoO	0.73	0.74	0.63	1	0.34	0.55	-0.39	0.34	0.58	0.23	0.11
HFR	0.55	0.53	0.69	0.34	1	0.32	-0.25	0.77	0.28	0.38	0.08
L	0.62	0.64	0.52	0.55	0.32	1	-0.36	0.45	0.42	0.22	0.23
LCR	-0.47	-0.50	-0.33	-0.39	-0.25	-0.36	1	-0.18	-0.31	-0.50	-0.26
$E_{2\text{-}4kHz}$	0.37	0.34	0.51	0.34	0.77	0.45	-0.18	1	0.45	-0.05	0.16
EF	0.43	0.41	0.39	0.58	0.28	0.42	-0.31	0.45	1	-0.01	0.31
LFR	0.49	0.55	0.40	0.23	0.38	0.22	-0.50	-0.05	-0.01	1	0.04
I	0.30	0.25	0.21	0.11	0.08	0.23	-0.26	0.16	0.31	0.04	1

Tab. 3: Intercorrelations between descriptors in Tab. 2

By means of multiple linear regression analysis a model (1) was constructed from those three features.

$$\text{Hardness} \approx 2.67 + 33.8 * E_{perc} + 6.37 * E_{2\text{-}4kHz} - 4.65\ LCR \quad (1)$$

An automatically extractable high level descriptor based on this regression model shows a strong correlation ($r = 0.86$, $p < 0.01$, R^2 adjusted = 0.723) with the subjective hardness ratings (see Fig. 3).

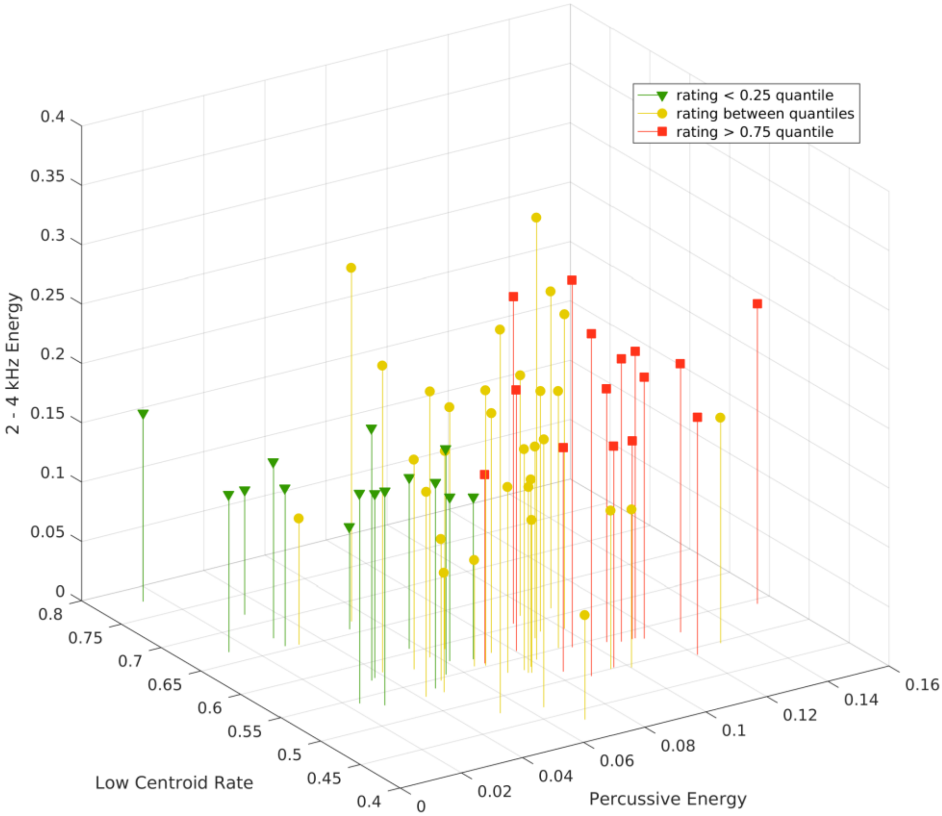

Fig. 2: The features percussive energy, low centroid rate and 2-4 kHz energy combined show a strong connection to the hardness ratings of the test subjects

In a different approach, when considering not only automatically extracted features, but also the singing style (which was manually classified into the categories "singing", "speaking/rap", "screaming", "growling", "combination" and "none") an even slightly higher correlation with the hardness ratings could be obtained ($r = 0.88$, $p < 0.01$, R^2 adjusted = 0.759).

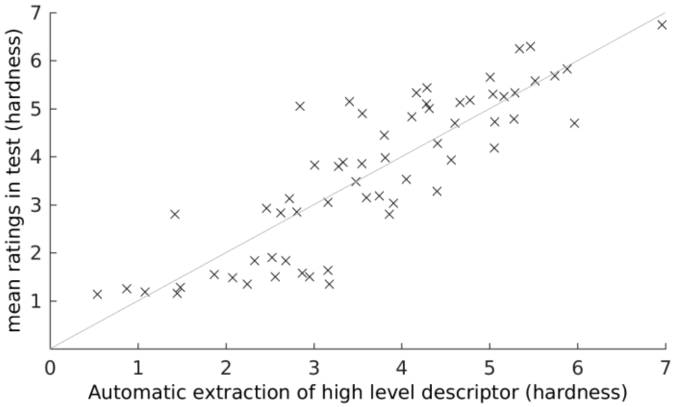

Fig. 3: Correlation between resulting hardness model and listening test ratings (r = 0.86, p < 0.01)

3.3 Musical genres

Hardness ratings varied with the style of the studied music excerpts with examples from the genres Black/Death Metal, Techno/Hardcore, Metal in general and New German Hardness receiving the highest average ratings (see Fig. 4).

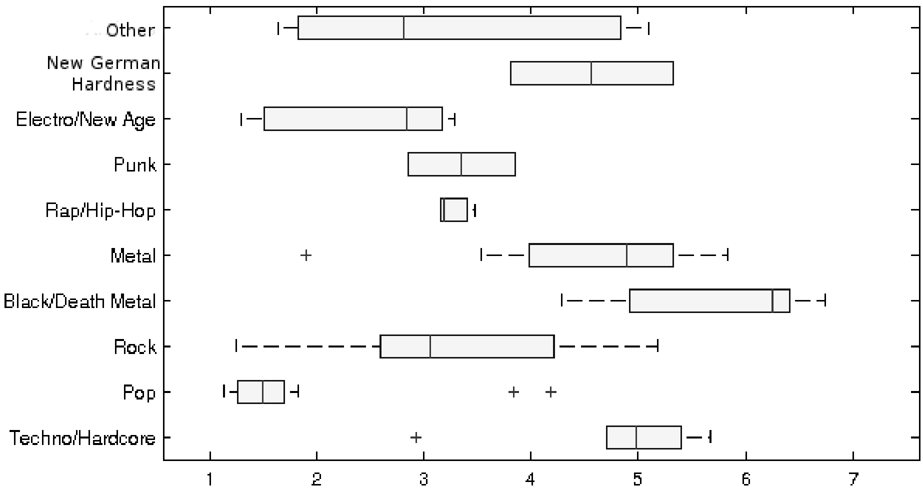

Fig. 4: Hardness ratings by musical genre

While the relevant psychoacoustic features were largely congruent across most genres, for Techno/Hardcore the flatness of the dynamic envelope alone showed a much stronger correlation ($r = 0.9$, $p < 0.01$) with the hardness ratings than in the case of other genres. A possible explanation could be dynamic compression used in the production of electronic music. However, as only eight examples from the genre Techno/Hardcore were being examined, it cannot be ruled out that this effect might be caused by outliers. A possible connection should be further investigated using a larger sample from this genre.

3.4 Influence of subjective preferences

Test participants with a positive preference for hard music were significantly more often male ($p < 0.05$ according to Fisher's exact test) and younger than those with negative preference ($p < 0.01$ according to two-sample t-test). Overall, those listeners of hard music gave significantly lower average hardness ratings than those with an aversion did ($p < 0.05$ according to t-test).

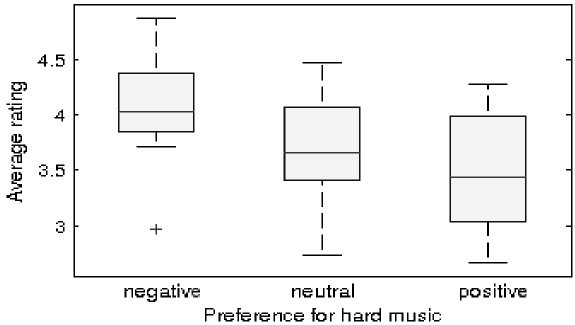

Fig. 5: Average hardness ratings by preference

Remarkably, participants with a negative preference for hard music gave significantly higher ratings for examples from the genre Techno/Hardcore than listeners of hard music did ($p < 0.01$ according to t-test), while a similar effect could not be found for example for Black/Death Metal ($p = 0.14$).

4 Discussion and outlook

Here the concept of musical hardness was examined using an audio feature extraction approach. This presents an example of how unbiased analysis by computational methods can be applied to gain insight and confirm knowledge in the study of musicological questions. Among these attributes were the presence of percussive instruments, that manifests in a high intensity of the corresponding signal components (as measured by percussive energy and spectral flux). Moreover, a flat temporal envelope in connection with a high temporal density (envelope flatness, number of onsets) and a generally high loudness with an emphasis on low as well as high frequency areas (loudness, low / high frequency ratio, 2-4 kHz energy) could be observed to correlate with hardness in music.

While manually classified attributes like e.g. the singing style can improve the performance of the model, future advances in music information retrieval will presumptively enable further approximation even when relying only on automatically extracted features. However, next to individual differences in the perception of hardness (see chapter 3.1), it has to be considered, that many aspects like compositional characteristics (e.g. specific heavy metal riffs), textual content of the lyrics and musical metadata are not taken into account with this approach and will - at least partially - remain as a semantic gap in the full explanation of the concept of musical hardness.

Nevertheless, it was possible to construct a general high level descriptor for musical hardness based only on automatically extractable signal features which shows a strong correlation ($r = 0.86$, $p < 0.01$) with the subjective hardness ratings. In an ongoing study, the resulting hardness model is being further evaluated in terms of its predictive power and particularly its interrelation with musical genres. Overall, the results already propose musical hardness as a useful high level descriptor in the context of analyzing and searching music collections, music recommendation systems and similar tasks.

Acknowledgements

Thank you to Jason Bosch, Daniel Back, Angelika Czedik-Eysenberg, Saleh Siddiq and Jörg Mühlhans for providing valuable suggestions and feedback on the manuscript.

Bibliography

[Be99] Berger, H.: Metal, Rock and Jazz: Perception and the Phenomenology of Musical Experience. Hanover, N.H.: Wesleyan University Press/University Press of New England, 1999.

[BEL03] Berenzweig, A., Ellis, D. P., & Lawrence, S.: Anchor space for classification and similarity measurement of music. In: Proc. Int. Conf. On Multimedia and Expo (ICME'03), Baltimore, Vol. 1, pp. I-29, 2003.

[BF05] Berger, H. & Fales, C.: Heaviness in the Perception of Heavy Metal Guitar Timbres: The Match of Perceptual and Acoustic Features over Time. Wired for Sound: Engineering and Technologies in Sonic Cultures. Middletown, CT: Wesleyan University Press, 2005.

[CKR17] Czedik-Eysenberg, I., Knauf, D., Reuter, C.: Hardness as a semantic audio descriptor - supplementary material. http://homepage.univie.ac.at/isabella.czedik-eysenberg/hardness/, accessed: 23/6/2017.

[DM14] Driedger, J. & Müller, M.: TSM Toolbox: MATLAB Implementations of Time-Scale Modification Algorithms. In: Int. Conf. on Digital Audio Effects, September 2014, pp. 249-256, 2014.

[DM15] Driedger, J., & Müller, M.: Harmonisch-Perkussiv-Rest Zerlegung von Musiksignalen. In: Proc. Deutsche Jahrestagung für Akustik (DAGA), pp. 1421-1424, 2015.

[Fi10] Fitzgerald, D.: Harmonic/percussive separation using median filtering. In: Proc. Int. Conf. on Digital Audio Effects (DAFX), Graz, Austria, pp. 246-253, 2010.

[Ge11] Georgi, R. von, Kraus, H., Cimbal, K., \& Schütz, M.: Persönlichkeit und Emotionsmodulation mittels Musik bei Heavy-Metal Fans. In (Auhagen, W., Bullerjahn C., Höge, H., ed.): Musikpsychologie. Jahrbuch der Deutschen Gesellschaft für Musikpsychologie, Bd. 21, pp. 90-118, 2011.

[Ge17] Genesis Loudness Toolbox, http://genesis-acoustics.com/en/loudness_online-32.html, accessed: 25/4/2017.

[Gr90] Gross, R. L.: Heavy metal music: A new subculture in American society. The Journal of Popular Culture, 24(1) pp. 119-130, 1990.

[LT07] Lartillot, O. & Toiviainen, P.: A Matlab toolbox for musical feature extraction from audio. In: Int. Conf. on Digital Audio Effects, September 2007, pp. 237-244, 2007.

[MGB97] Moore, B. C., Glasberg, B. R., & Baer, T.: A model for the prediction of thresholds, loudness, and partial loudness. Journal of the Audio Engineering Society, 45(4), pp. 224-240, 1997.

[My12] Mynett, M.: Achieving intelligibility whilst maintaining heaviness when producing contemporary metal music. Journal on the Art of Record Production 6, 2012.

[On08] Ono, N., Miyamoto, K., Kameoka, H. & Sagayama, S. A real-time equalizer of harmonic and percussive components in music signals. In: Proc. of the Int. Society for Music Information Retrieval Conference (ISMIR), Philadelphia, Pennsylvania, USA, pp. 139–144, 2008.

[Pr17] Protrabant Online-Versuchsumgebung, https://denkn.at/protrabant, accessed: 25/04/2017.

[Re08] Reyes, I.: Sound, Technology, and interpretation in Subcultures of Heavy Music Production. Dissertation - Pittsburgh University, 2008.

[ROM14] Reuter, C., Oehler, M. & Mühlhans, J.: Physiological and acoustical correlates of unpleasant sounds, In: Proc. Joint Conference ICMPC13-APSCOM5, August 4-8, 2014, Yonsei University, Seoul, Korea, pp. 97, 2014.

[Wa93] Walser, R.: Running with the devil: Power, gender, and madness in heavy metal music. Wesleyan University Press, 1993.

[WBG11] Wallach, J., Berger, H. M. & Greene, P. D.: Metal rules the globe: heavy metal music around the world. Duke University Press, pp. 180ff, 2011.

[We05] Weindl, D.: Musik & Aggression. Untersucht anhand des Musikgenres Heavy Metal. Peter Lang, Frankfurt am Main, 2005.

[We91] Weinstein, D.: Heavy Metal: A Cultural Sociology. New York, N.Y.: Maxwell Macmillan International, 1991.