THE MIR PERSPECTIVE ON THE EVOLUTION OF DYNAMICS IN MAINSTREAM MUSIC

Emmanuel Deruty
Sony Computer Science Laboratory / Akoustic Arts
Paris, France
emmanuel.deruty@gmail.com

François Pachet
Sony Computer Science Laboratory
Paris, France

ABSTRACT

Understanding the evolution of mainstream music is of high interest for the music production industry. In this context, we argue that a MIR perspective may be used to highlight, in particular, relations between dynamics and various properties of mainstream music. We illustrate this claim with two results obtained from a diachronic analysis performed on 7200 tracks released between 1967 and 2014. This analysis suggests that 1) the so-called “loudness war” has peaked in 2007, and 2) its influence has been important enough to override the impact of genre on dynamics. In other words, dynamics in mainstream music are primarily related to a track’s year of release, rather than to its genre.

1. INTRODUCTION

Mainstream popular music is in constant evolution. There may be more differences than common points between progressive rock albums from the 1970’s such as Pink Floyd’s best-selling “Dark Side of the Moon” and contemporary rap albums such as Nicki Minaj’s platinum-certified “Roman Reloaded”. Studies tracking down the yearly evolution of signal descriptors are useful to characterize this diversity.

In 1982, Moller [1] established that recent recordings feature a larger dynamic excursion than older ones. More recently, Tardieu [2] studied the evolution of stereo, dynamic and spectral features on pop/rock songs, and showed that decade classification accuracies using spectral and dynamic features are equal. Pestana [3] focused on spectral features and found that while spectra are dependent on genre, they also follow the yearly evolution of production standards. Serrà [4] performed a systematic analysis of more than 400,000 tracks and concludes that popular music “show[s] no considerable changes in more than fifty years” other than becoming louder, a result challenged by Mauch [5]. Deruty [6] focused on the changes in loudness and dynamics over the same period, and provided a characterization of the phenomenon referred to as the “loudness war”. The loudness war, or loudness race, is a trend in popular music production that affects mainstream music dynamics [7]. It has been described as a contest between bands and record companies, in which music is engineered to be louder than the competition’s [8, pp. 237–292]. Starting at the end of the 80’s [4], [6], [9], its effects have been spectacular enough to reach the general media [10]–[11]. A distinction is made between dynamics occurring on different time-scales. The large-scale variations are known as macrodynamics, whereas the faster ones variations are referred to as microdynamics [12]–[13]. The loudness war favors high loudness tracks with reduced microdynamics [4], [6], [9], although some authors claim it has also reduced macrodynamics [14]. Efforts have been made to reverse the trend, through measurement protocols [15]–[16], integrated loudness-leveling engines such as iTunes’ Sound Check [17], or public communications [18]–[19].

In this paper, we perform a diachronic analysis on 7200 mainstream tracks released between 1967 and 2014, and present two results. First, we show that the evolution towards louder and less dynamic content peaked in 2007, and then started to decrease. If this trend continues, pre-loudness war values for most descriptors of music dynamics may be observed sometimes between 2017 and 2026. Second, we demonstrate that the loudness war’s impact supersedes the influence of music genre on dynamics. In mainstream music, a piece’s dynamics are more typical of a given year than they are of a given genre.

2. METHOD

2.1 Music corpus

The music corpus we rely on is a revision and extension of the corpus used in [6]. It includes 7200 tracks released between 1967 and 2014, 150 tracks per year. Track selection is based on Besteveralbums.com, a review aggregator. For each year, we choose the albums with the best ratings. If a given artist is the author of more than three well-rated albums, we select the artist’s complete discography. While this method does not lead to a random sampling, it ensures that the corpus is based on music that is popular. We choose to start the corpus at the end of the sixties because these years can be considered as the advent of the contemporary pop/rock era, characterized by the creative use of the recording studio [8, p. 157] along with mass media availability [20].

2.2 Signal descriptors

We use the signal descriptors defined in [6]. The track’s physical level is measured using the RMS power of the signal after normalization. Track loudness evaluation is performed using the EBU3341 integrated loudness [21], which has been shown to be as robust as more complex measures such as detailed perceptual models [22]. Microdynamics are measured using a variation of the crest factor, as defined in [6]. For macro-dynamics, we rely on the EBU3342 Loudness Range [23], which is, to our knowledge, the only normative descriptor to quantify dynamics in a musical sense (*piano*, *forte*…) [9]. We evaluate the overall amount of dynamic processing using the Peak to RMS Regression Coefficient (PRRC). PRRC values below 1 indicate usage of dynamic compression, values above 1 usage of dynamic expansion [6]. Finally, we estimate the amount of limiting applied to the tracks using the High Level Sample Density (HLSD) [6]. HLSD can be linked to the practice of limiting [6], which is suspected to have a decisive impact on mainstream music production during the last 30 years [8, pp. 237-292], [9], [14], [24]–[27]. Using relations between limiting and HLSD as shown in [6], we indeed find that a significant amount of limiting (> 3dB) seems to have been applied on 33% of all tracks from our corpus, and on 65% of tracks released after 1994.

For each descriptor, we provide a projection based on the current trend by fitting the descriptor's smoothed median values using a second-degree polynomial, starting from the year for which the loudness war is observed to peak. As illustrated in Figure 1 (black dot at the right of the graphs), estimation of the return to pre-loudness war values is obtained using the crossing of the projected values with the median of the pre-1990 descriptor values.

2.3 Genre labels

Following [28]–[30], we draw the music genre labels from AllMusic, a website that provides "unoptimized expert annotated ground truth dataset for music genre classification" [30] in the form of a database of commercial music annotated in terms of "genres", "meta-styles" and "styles". Whereas AllMusic provides only 21 "genres", album information also comes with 905 "styles" and "meta-styles" that can be interpreted as sub-genres to refine the major genre labels. In this paper, while relying on the "styles" provided by Allmusic, we designate them as "genres", "a conventional category that identifies pieces of music as belonging to a shared tradition or set of conventions" [31]. Under this terminology, the 7500 tracks from the corpus correspond to 272 distinct mainstream music genres, each track being associated with a mean of 4 genres, the minimum being 1 and the maximum 11. Conversely, each genre is represented with a mean of 110 tracks, the minimum being 3 and the maximum 2482. Issues linked to the pertinence of the results regarding this diversity of representation are discussed in Section 4.3.

3. DIACHRONIC STUDY OF DYNAMICS

Figure 1 illustrates the descriptors' behavior over time. The boxes' upper and lower limits indicate the 25th and 75th percentiles of the distribution. The darker box indicates the peak of the loudness war for the descriptor, i.e. the year for which the median value is maximal. The small horizontal lines inside the boxes indicate the median. The outer whiskers stand for the 5th and 95th percentiles. The solid, thick black curve is the smoothed median, on which the projection is based. The projection itself is represented by a dashed gray line. The thin horizontal line indicates the median pre-1990 descriptor values.

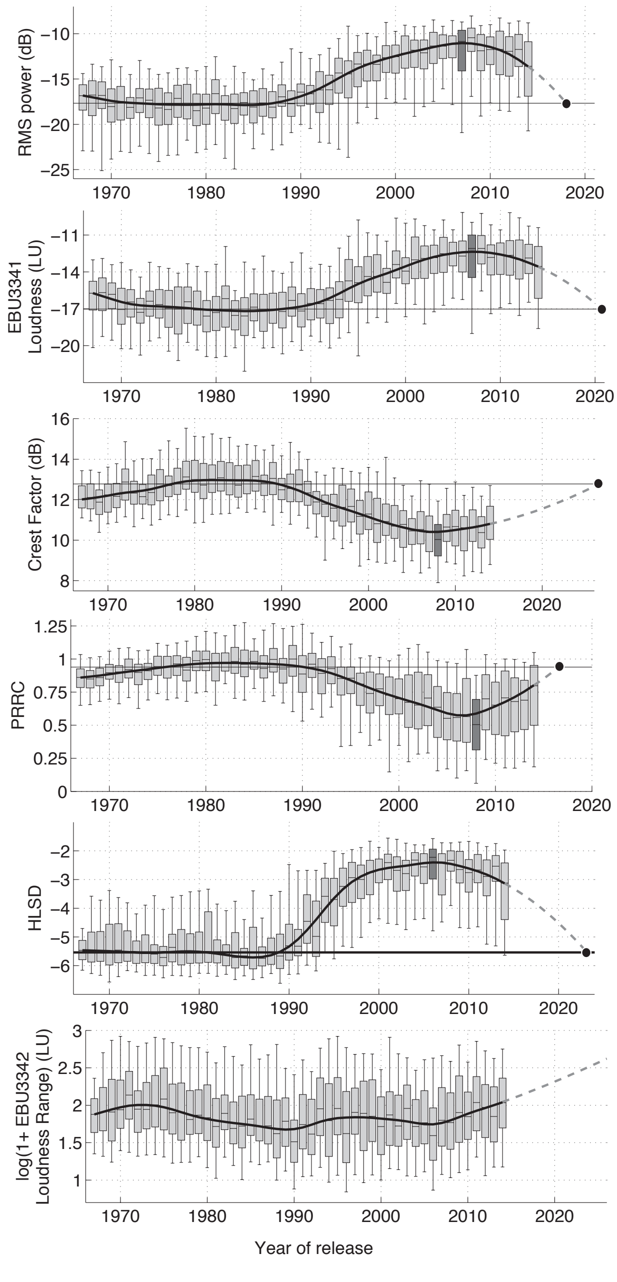

Figure 1. Descriptor evolution over the years. From top to bottom RMS power, EBU3341 integrated loudness, crest factor, PRRC, HLSD and EBU3342 Loudness Range.

The loudness war may be characterized by a change towards previously unobserved descriptor values that starts around 1990 and indicates the use of more dynamic compression [6]. Table 1 summarizes the loudness war timeline depending on the descriptor. It took ca. 15 years for the loudness war to peak. The return to pre-loudness war values could take between 10 and 20 years. Figure 1 shows that macrodynamics are not affected by the loudness war. No significant change of values starting around 1990 and pointing toward more dynamic compression can be observed. The loudness war has increased music level and micro-dynamics, but has not decreased macro-dynamics.

Descriptor	Corresponding phenomenon	Peak	Estimated return to pre-loudness war values
RMS power	Physical level	2007	2018
EBU3341	Loudness [21]	2007	2020
Crest factor	Microdynamics [6], [12]	2008	2026
PRRC	Overall amount of dynamic processing [6]	2008	2017
HLSD	Amount of limiting [6]	2006	2023

Table 1. Loudness war timeline summary.

Since 2006, macrodynamics have increased consistently, and are higher in 2014 than they have ever been during the time-span covered by the corpus. This increase can be put in relation with a demand for more dynamics combined with the confusion that's often made between micro- and macrodynamics [6], [10]–[11], [14], [32]. Musicians and producers may be trying to counter the effects of the loudness war by raising macrodynamics, whereas raising microdynamics would be more productive in that respect. However, examination of Figure 1 shows that macrodynamics follow relatively shorter trends than other descriptors, and a reversal of the present tendency towards less macrodynamics could be witnessed as soon as 2015.

4. DYNAMICS AND MAINSTREAM GENRES

4.1 Dependency of dynamics on genres and trends

In this section, we show that dynamics of mainstream music are more typical of a given year than they are of a given genre. Figure 2 illustrates the distribution of RMS power values depending of the music genre of the track. On first approach, it suggests that music genre and RMS power are related. However, as illustrated in Figure 1, RMS power is also related to the year of the album release. Figure 3 provides more details, by illustrating RMS power evolution for the four most represented genres in the corpus (Alternative Pop/Rock, Alternative/Indie Rock, Album Rock and Contemporary Pop/Rock). It indicates that genres follow the year's trend in terms of RMS power. This phenomenon, previously mentioned in [32], suggests that RMS values may be primarily related to the year of the track release, rather than to its genre. We use two methods to confirm the tendency: a standard ANOVA and a variance evaluation.

The second method possesses the advantage of providing results formulated using the original descriptor's unit, and therefore being easier to interpret than the ANOVA's results. It involves the evaluation of the RMS distribution's variance for each genre and for each year, followed by the computation of the weighted arithmetic means of the variances, taking into account genre and year representativeness. The process is illustrated in Figure 4. The weighted mean variance for each year is 9.03dB, whereas the weighted mean variance for each style is 14.19dB. This shows that RMS values primarily originate from the track's year of release. In other words, particular physical levels are more typical of a given year than they are of a given genre. As shown in Table 2, this result is confirmed by the ANOVA's *F*-statistic. We repeat the experiment using the other descriptors described in Section 2.2. Results are similar. With the exception of the EBU3342 LRA, descriptors are clearly more related to the year's trend than to the piece's genre.

Descriptor	Mean variance for each year	Mean variance for each genre	ANOVA's *F*-statistic (years as classes)	ANOVA's *F*-statistic (genres as classes)
RMS power	9.03dB	14.2dB	107.7	6.4
EBU3341	4.57LU	7.41.LU	104.9	6.5
Crest factor	1.35dB	2.25dB	110.4	5.4
PRRC	0.04	0.06	77.5	4.9
HLSD	0.79	2.08	274.7	8.6
EBU3342	14.5LU	14.3LU	7.3	4.7

Table 2. Comparison of the weighted mean arithmetic means of the descriptor variances for each year and each genre, as well as comparison of the ANOVA's *F*-statistics, show that dynamics in mainstream music are primarily linked to the piece's year of release, rather than to its genre.

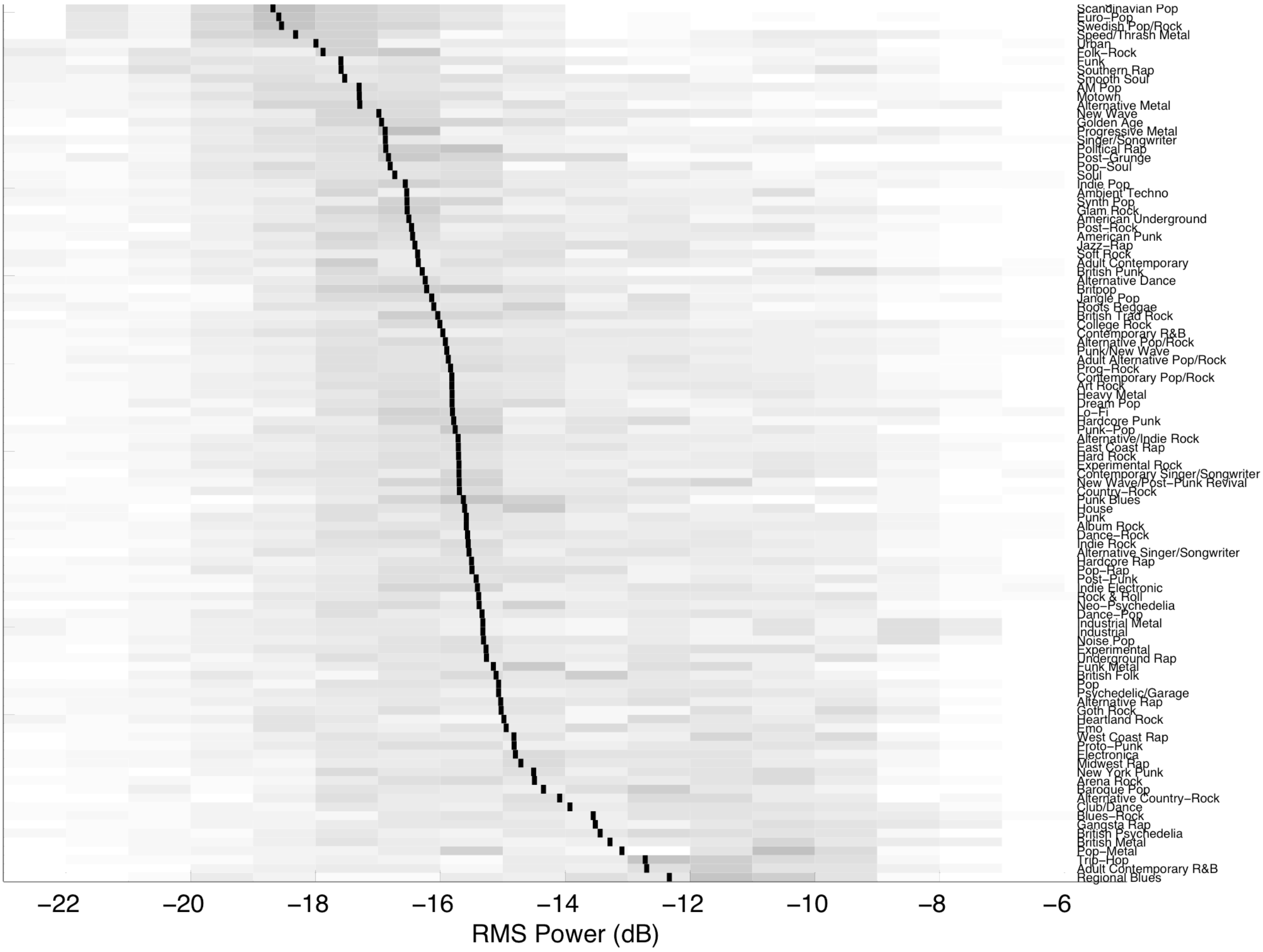

Figure 2. Distribution of RMS power values depending on the tracks' genres. Darker shades of gray indicate higher levels of distribution. The black rectangles indicate the median. This Figure is restricted to styles corresponding to more than 50 tracks.

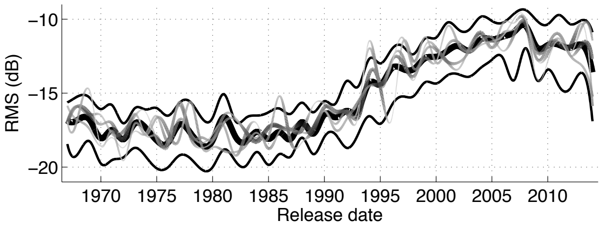

Figure 3. In gray, RMS power values corresponding to the music genres most represented in the corpus. Lighter gray sections indicate years with fewer tracks. The three black lines represent the 25th, 50th and 75th percentiles.

4.2 The particular cases of HLSD and LRA

As shown in Table 2, a particularly high dependence to trends is clear in the case of the HLSD, with an *F*-statistic being higher than in the case of the other descriptors. As seen in Section 2.2, it implies that the amount of limiting applied by audio engineers during mastering can be considered as independent from genre. Therefore, mainstream genres cannot be said to sound more or less "hot". This is an important information in the context of mainstream music mastering: it can help engineers choose and argue the output level with their client, which is often a critical debate [33]. On the other hand, dependency to trends is much lower in the case of the EBU3342 LRA. As a result, macrodynamics can be considered as relatively independent from both genre and year of release.

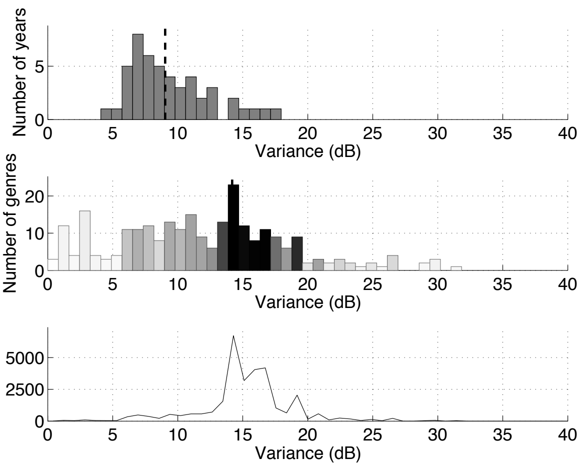

Figure 4. Distribution of RMS power variances. Top, by year. Middle, by genre. The dashed vertical line represents the weighted mean of the distribution. Bar hues indicates style representativeness. Bottom, style representativeness displayed quantitatively.

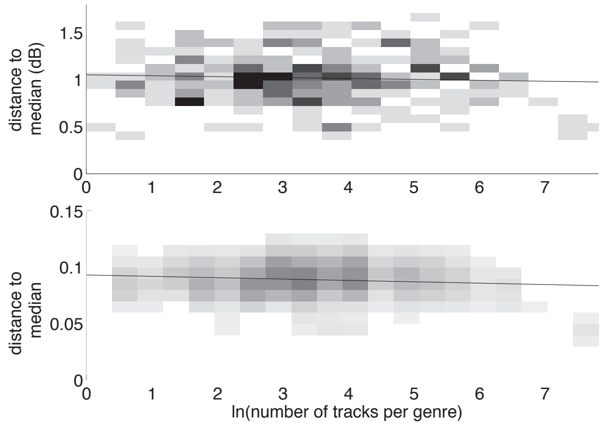

Figure 5. Top, distance between each genre and the all-genre median value for the RMS descriptor, against genre representation. Bottom, result of the same process using random values between 0 and 1. The horizontal line represents the linear regression.

4.3 Discussion

It may seem counter-intuitive to conclude that dynamics are more dependent on trends than they are on genres. Indeed, genres such as Euro-Pop exhibit high microdynamics and low overall loudness, whereas other genres such as Trip-Hop are associated with low microdynamics and high overall loudness. However, the Euro-Pop genre is most represented in the 1970s and 1980s [34], at a period when music was produced to feature high microdynamics and low overall loudness [6]. Trip-Hop is mainly a mid-1990s trend [35], a moment when low microdynamics and high overall loudness were common in music production [6]. Conversely, all genres that span several decades follow the trend of the year of production.

As mentioned in Section 2.3, not all genres are equally represented. This may bring the suspicion that dynamics are only dependent on the trends followed by the most represented genres, such as the subgenres of rock represented in Figure 3, but independent from the trends followed by most other genres, in which case our conclusion would not stand. To discard this suspicion, we evaluate the distance between each genre and the all-genre median value for the descriptors over the years. This distance is then matched against the genre's number of occurrences. Figure 5, top, illustrates the case of the RMS descriptor. A few well-represented genres are indeed closer to the median than most other genres. However, Figure 5, bottom, illustrates the same process using 1000 sets of 7500 random values in place of the 7500 RMS values. Both graphs are similar, and the few well-represented genres are closer to the median in both cases. Therefore, a particular dependency to a few genres is not a property of the present corpus. This discards the suspicion according to which the dependency to trends we found is only valid as far as a few genres are concerned.

5. CONCLUSION

Mainstream music dynamics are thought to be conditioned by genre, in terms of overall track loudness [36], microdynamics [9], [37], macrodynamics [15], [38] or amount of dynamic processing applied to music pieces during the production stage [7], [39, p. 121]. However, using a MIR perspective, we have shown that dynamics and overall loudness depend more on the track's year of release than on its genre. We have also found, as suspected by [40], that the loudness war has influenced all mainstream genres indiscriminately. A notable exception lies in macrodynamics as measured by the EBU3342 Loudness Range, which are more independent from both genre and year of release. In other words, dynamic range in the musical sense (*pianissimo* to *fortissimo*) is only marginally dependent on either mainstream genre or trend.

According to mastering engineer Bob Katz, the loudness wars were over in 2013 [41]. We have shown that the loudness war has peaked in 2007, and that a return to pre-loudness war dynamics may be reached in about ten years. As an exception, macrodynamics, which have not been significantly influenced by the loudness war, appear to increase since the loudness war's peak, and are currently reaching very high values.

This is useful knowledge in several situations. Many artists and producers ask sound engineers to increase loudness during mastering [33], arguing that the music genre to which their tracks belong is well suited to a "hot", loud and compressed sound. The present study provides objective data to challenge this claim. Loudness war activists argue for more important dynamics [32], [41]. We have shown that this concerns only microdynamics. Automatic mixing and mastering rely on constraints to be applied on initial audio content [42]–[44]. The present study has demonstrated that constraints relative to dynamics in mainstream music may be derived from trends rather than genres.

More generally, we suggest that the present method could be used for other audio descriptors, in order to establish their dependency to either diachronic trends, genre, or to any other musical dimension.

6. ACKNOWLEDGEMENTS

This research is supported by the project *Lrn2Cre8* which is funded by the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET grant number 610859.

7. REFERENCES

[1] L.G. Moller, "How Much Has Amplitude Distribution of Music Changed?," in *AES 71st Conv.*, Montreux, Switzerland, 1982.

[2] D. Tardieu *et al.*, "Production Effect: Audio Features For Recording Technique Description and Decade Prediction," in *14th Int. Conf. on Digital Audio Effects (DAFX 11)*, Paris, France, 2011.

[3] P.D. Pestana *et al.*, “Spectral Characteristics of Popular Commercial Recordings 1950-2010,” in *AES 135th Conv.*, New-York, NY, 2013.

[4] J. Serrà *et al.*, “Measuring the evolution of contemporary western popular music,” *Scientific Reports,* Jul. 2012.

[5] M. Mauch *et al.* (2015). *The evolution of popular music: USA 1960–2010* [Online]. Available: http://rsos.royalsocietypublishing.org/content/2/5/150081. Royal Society Open Science, DOI: 10.1098/rsos.150081.

[6] E. Deruty and D. Tardieu, “About Dynamic Processing in Mainstream Music,” *J. Audio Eng. Soc.* 62(1/2), pp. 42-55, Jan. 2014.

[7] E. Vickers, “The loudness war: Background, speculation and recommendations,” in *AES 129th Conv.*, San Francisco, CA, 2010.

[8] G. Milner, *Perfecting sound forever: an aural history of recorded music.* London, UK: Faber & Faber, 2009.

[9] E. Deruty, “’Dynamic range’ and the loudness war,” *Sound on Sound*, Sep. 2011.

[10] E. Smith, “Even heavy-metal fans complain that today’s music is too loud!!!,” *The Wall Street Journal*, Sep. 2008.

[11] K. Matersonn, “Loudness war stirs quiet revolution by audio engineers,” *Chicago Tribune*, Jan. 2008.

[12] E. Skovenborg, “Measures of Microdynamics,” in AES 137th Conv., Los Angeles, CA, 2014.

[13] B. Katz and R. Katz, *Mastering audio: the art and the science.* Focal Press, 2007

[14] S. Sreedhar. “The future of music,” *IEEE Spectrum*, Aug. 2007.

[15] E. Skovenborg, “Loudness Range (LRA), Design and Evaluation,” in *AES 132th Conv.*, Budapest, Hungary, 2012.

[16] European Broadcasting Union. (2015). *Loudness* [Online]. Available: https://tech.ebu.ch/loudness.

[17] T. Lund. (2013). *Audio for Mobile TV, iPad and iPod* [Online]. Available: http://www.tcelectronic.com/media/2040040/mobile-test-paper-2013.

[18] J. V. Serinus. (2012). *Winning the Loudness Wars* [Online]. Available: http://www.stereophile.com/content/winning-loudness-wars.

[19] R. Archer. (2014). *Have the Loudness Wars Peaked?* [Online]. Available: http://www.cepro.com/article/excuse_the_pun_have_the_loudness_wars_peaked/.

[20] *The Cambridge companion to recorded music.* Cambridge, UK: Cambridge University Press, 1999.

[21] European Broadcasting Union. (2011). *EBU - TECH 3341* [Online]. Available: https://tech.ebu.ch/docs/tech/tech3341.pdf.

[22] E. Skovenborg and S. Nielsen, “Evaluation of different loudness models with music and speech material,” in *AES 117th Conv.*, 2004, San Francisco, CA, 2004.

[23] European Broadcasting Union. (2011). *EBU-TECH3342* [Online]. Available: https://tech.ebu.ch/docs/tech/tech3342.pdf.

[24] E. Vickers, “The Non-flat and Continually Changing Frequency Response of Multiband Compressors,” in *AES 129th Conv.*, San Francisco, CA, 2010.

[25] P. Kraght, “Aliasing in Digital Clippers and Compressors,” *J. Audio Eng. Soc.* 48 (11), Nov. 2000.

[26] A. Travaglini, “Broadcast Loudness: Mixing, Monitoring and Control,” in *122nd AES Conv.*, Vienna, Austria, 2007.

[27] E. Skovenborg and T. Lund, “Loudness descriptors to characterize programs and music tracks,” in *AES 125th Conv.*, San Francisco, CA, 2008.

[28] J. Bergstra *et al.*, “Predicting genre labels for artists using FreeDB,” *Proc. of the 7th Int. Conf. on Music Information Retrieval*, 2006, pp. 85-88.

[29] Y. Hu and M. Ogihara, “Nextone Player: A Music Recommendation System Based On User Behavior,” *Proc. of the 12th Int. Conf. on Music Information Retrieval*, 2011, pp. 103-108.

[30] A. Schindler *et al.*, “Facilitating Comprehensive Benchmarking Experiments On The Million Song Dataset”, *Proc. of the 13th Int. Conf. on Music Information Retrieval,* 2012, pp. 469-474.

[31] J.Samson. (2001). *Grove Music Online: Genre* [Online]. Available: http://www.oxfordmusic online.com /subscriber/article/grove/music/40599.

[32] I. Shepherd. (2014). *What is the Loudness War?* [Online]. Available: http://dynamicrangeday.co.uk/about.

[33] T. Woodhead. (2015). *What is Mastering?* [Online]. Available: http://www. hippocraticmastering.com/whatismastering.html.

[34] AllMusic. (2015). *Euro-Pop* [Online]. Available: http://www.allmusic.com/style/euro-pop-ma0000004446.

[35] AllMusic. (2015). *Trip-Hop* [Online]. Available: http://www.allmusic.com/ style/trip-hop-ma0000002906.

[36] E. Skovenborg *et al.*, “Loudness Assessment of Music and Speech,” in *AES 116th Conv.*, Berlin, Germany, 2004.

[37] M. Walsh *et al.*, “Adaptive Dynamics Enhancement,” in *AES 130th Conv.*, London, UK, 2011.

[38] H. Robjohns, “The End Of The Loudness War?” *Sound on Sound*, Feb. 2014.

[39] B. Katz and R. Katz, *Mastering audio: the art and the science.* Focal Press, 2007.

[40] A.v. Ruschkowski, “Loudness war,” in *Systematic and comparative musicology : concepts, methods, findings*. Albrecht Schneider (ed.), pp. 213-230, 2008.

[41] B. Katz. (2013). *The Loudness War Has Been Won: Press Release* [Online]. Available: http://www.digido.com/forum/ announcement/id-6.html.

[42] Z. Ma *et al.,* “Implementation of an intelligent equalization tool using Yule-Walker for music mixing and mastering,” in *AES 134th Conv.*, Rome, Italy, 2013.

[43] E. Perez-Gonzales and J.D. Reiss, “Improved control for selective minimization of masking using interchannel dependancy effects,” *Proc. of the 11th Int. Conf. on Digital Audio Effects (DAFX-08)*, pp. 75-81, 2008.

[44] E. Deruty *et al.*, “Human Rock Mixes Exhibit Tight Relations Between Spectrum And Loudness.” *J. Audio Eng. Soc.*, 62 (10), Oct. 2014.