Audio Engineering Society

Convention Paper 8960

Presented at the 135th Convention
2013 October 17–20 New York, USA

Spectral Characteristics of Popular Commercial Recordings 1950-2010

Pedro Duarte Pestana[1,3], Zheng Ma[2], Joshua D. Reiss[2], Álvaro Barbosa[1] and Dawn A. A. Black[2]

[1] *CITAR - Catholic University of Oporto, Rua Diogo Botelho, 1327, 4169-005 Oporto, Portugal*

[2] *C4DM - Queen Mary University of London, Mile End Road, London E1 4NS, United Kingdom*

[3] *also member of ILID - Lusíada University of Portugal and CEAUL, Centro de Estatística e Aplicacoes - UL*

Correspondence should be addressed to Pedro Duarte Pestana (pedro.duarte.pestana@gmail.com)

ABSTRACT

In this work, the long-term spectral contours of a large dataset of popular commercial recordings were analyzed. The aim was to analyze overall trends, as well as yearly and genre-specific ones. A novel method for averaging spectral distributions is proposed, which yields results that are prone to comparison. With it, we found out that there is a consistent leaning towards a target equalization curve that stems from practices in the music industry, but also to some extent mimics natural, acoustic spectra of ensembles.

For as long as spectral analysis has been a viable tool in the commercial sectors, audio engineers have looked at integrated spectral responses as possible answers for audio quality. Michael Paul Stavrou [1] states that, while at Abbey Road, he lost endless afternoons hopelessly chasing the illusive hit song characteristic in technical parameters and Neil Dorfsman [2] acknowledges that, while many sound engineers would not admit to doing it, he feels that most of them use spectral analysis and comparison to previous work or other commercial work as a standard tool during mixing. In the mixing context, "achieving frequency balance (also referred to as tonal balance) is a prime challenge in most mixes" [3]. Bob Katz [4] proposes that the tonal balance of a symphony orchestra is the ideal reference for the spectral distribution of music. Yet there is no consistent academic study that tackles the question of how generally similar is the spectral response of critically acclaimed tracks, nor has anyone analyzed the surrounding factors upon which it depends.

The seminal work in spectrum analysis of musical signals is [5] (in which live signals are used), and it pioneered the 1/3 octave filter bank analysis process that influenced most early studies of the same type. The musical signals were of individual instruments and ensembles in live rooms. McKnight [6] took a similar approach in the realm of pre-recorded material but was looking for technical correction measures in the distribution format and used a small dataset. The earliest study that is closest to ours is Bauer's [7], where the author looked for the average statistical distribution of a small classical dataset. Moller [8] is the only analysis that tries to track down the yearly evolution of spectra. The BBC [9] researched the spectral content of pop music, using custom recordings made for the purpose of the test and [10] focused on the effect of the Compact Disc media on the spectral contour of recordings. Recently, [11] and [12] returned to the subject with a broader dataset, but their analyses focused more on dynamics and panning than frequency response, and their dataset does not follow any objective criteria of popularity. No study relies on a detailed FFT approach as we do, often choosing instead the coarser and more error prone Real Time Analysis (RTA) filter bank approach; nor has any of the aforementioned works tackled a really large representative dataset that follows the idea of commercial popularity, and thus a *'best-practices'* approach.

For our analysis to be consistent with general public preference, we must run it on a dataset that includes the most commercially relevant songs of the time period of interest. We chose to select songs that had been number ones in either the US or the UK charts, found primarily from [13, 14] and Wikipedia. The anglo-saxon bias was considered acceptable as most of the western world's music industry has a very strong anglo-saxon influence. The list of all the aforementioned singles can be found at [15], a document which also indicates the songs we were able to use. Our dataset is comprised of about half the singles that have been number one over the last 60 years, with a good representation of both genre and year of production (as there were no pilot tests that would allow an estimation of the ideal sample size, we tried, as is customary, to get the largest possible number of observations). All the songs in our dataset are uncompressed and, while we tried to find un-remastered versions, it was not always possible. This means that we are giving extra prominence to current standards of production and the differences we present should be even greater than that which our data suggests. Table 1 shows the number of songs we had available, divided by decade.

Years	Number of Songs
50s	71
60s	156
70s.	129
80s	193
90s	96
After 2000	127
Total	772

Table 1: Number of songs per decade in the dataset.

In Section 2 we will look at the overall average of all the songs in our collection. In Section 3 and 4, the data will be broken down by year and genre respectively and some additional low-level features are introduced to better characterize the differences we are unveiling. Section 5 presents an overview of the present research and presents some viable future directions and applications. The aforementioned accompanying website [15] includes more detailed plots, discussion of remastering, and extended numerical data for the results that have been found in this research.

1. OVERALL AVERAGE SPECTRUM OF COMMERCIAL RECORDINGS

Our main analysis focused on the monaural (left+right channel over two), average long-term spectrum of the aforementioned dataset. In order for spectra to be comparable, we first make sure that all songs are sampled at the same frequency (44.1 kHz being the obvious candidate for us, as most works stemmed from CD copies), and that we apply the same window length (4096 samples) to all content, so that the frequency resolution is consistent ($\approx 10\ Hz$). Let:

$$\begin{gathered} X(k,\tau) = \sum_{n=\tau\cdot w_{len}}^{(\tau+1)w_{len}-1} x(n)\, e^{-j2\pi k \frac{n}{N}}, \\ k = \left\{0,1,...,2^{12}-1\right\}, \tau = \left\{0,1,...,\left\lfloor \frac{x_{len}}{w_{len}} \right\rfloor\right\} \end{gathered}, \tag{1}$$

where k is the frequency bin and τ the time window number. x_{len} and w_{len} are the song and window lengths, respectively. This will yield a $K \times \mathrm{T}$ matrix ($K = w_{len}, \mathrm{T} = \lfloor x_{len}/w_{len} \rfloor$), and we then consider the integrated spectral response to be the mean magnitude over τ:

$$\overline{X}(k) = \frac{\sum_{\tau} |X(k,\tau)|}{\left\lfloor \frac{x_{len}}{w_{len}} \right\rfloor + 1}. \quad (2)$$

Equation (2) loses the 1 in the denominator whenever $\bmod(x_{len}, w_{len}) = 0$.

It is still necessary to tackle the problem of different spectral distributions having potentially different overall power values. Strict normalization is not the answer, as spurious radical peaks in the frequency distribution might cause overall lower power levels, and the comparison would yield results that showed a variability that was greater than the real variability (one could take, as an example, a comparison between a white noise spectrum and one that adds a single sinusoid at 1000 Hz to the same white noise — if the sinusoid is greater in magnitude, a normalization process would bring all other bins in the second spectrum down and lead us to conclude that the spectra were very different, while in actuality they are not). There are several available solutions, but we opted to scale all spectral distributions so that the bin sum would be 1, followed by averaging the cumulative distribution function. This means normalizing according to:

$$\widetilde{X}(k) = \frac{\overline{X}(k)}{\sum_{k} \overline{X}(k)}, \quad (3)$$

and accumulating over the bins:

$$X_c(k) = \sum_{i=0}^{k} \widetilde{X}(i). \quad (4)$$

We then compute a mean calculation of each point in the cumulative distribution $(\overline{X_c(k)})$. The average spectrum is found by computing the differences between adjacent bins, and multiplying by the average magnitude of all songs. This is basically an inversion of the process described above, and it is shown in the following equation:

$$\bar{X}_{AV}(k) = \frac{\sum_{k} \bar{X}(k)}{S} \left(\overline{X_c(k)} - \overline{X_c(k-1)} \right). \quad (5)$$

with S the total number of songs. The result of averaging the spectra of all songs in the dataset is shown in Figure 1, along with a plot that overlaps all the individual distributions. The trend seen in the average spectrum is consistent with what can be observed for the individual distributions and the 95% confidence intervals indicated are so narrow that they are not perceptible on the shown scale. The average standard deviation for the normalized cumulative values is 0.044, which is a well behaved value across frequency bins (though averaging 2048 standard deviations drowns out the larger values in the low-end frequency region). All the subsequent analysis follows this averaging scheme.

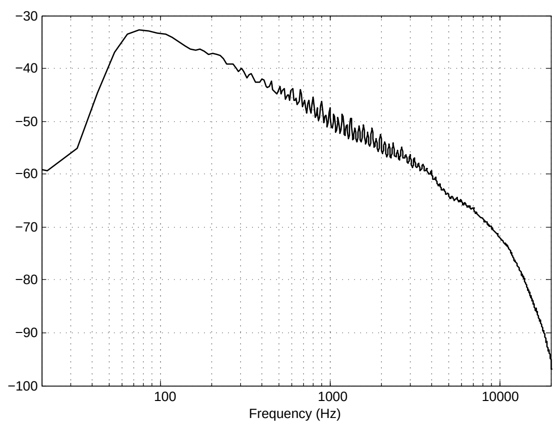

Fig. 1: Average spectrum of all available data.

2. YEARLY EVOLUTION OF SPECTRA AND SPECTRAL FEATURES

Figure 2 shows the average spectrum evolution through time, along with some decade-by-decade snapshots of revealing frequency ranges.

This overall picture allows us to understand that as well as the consistent net increase in magnitude, there is a tendency for an extended low frequency response with a lower resonant peak as we move to more recent years. The presence area (1.5 kHz to 4 kHz) also shows some differences, with the slope becoming shallower as the decades move forward, meaning we are more prone to a boost in definition that may arise from the progressive change from analogue to digital. The bottom part of the figure shows that the regions where apparent discrepancies

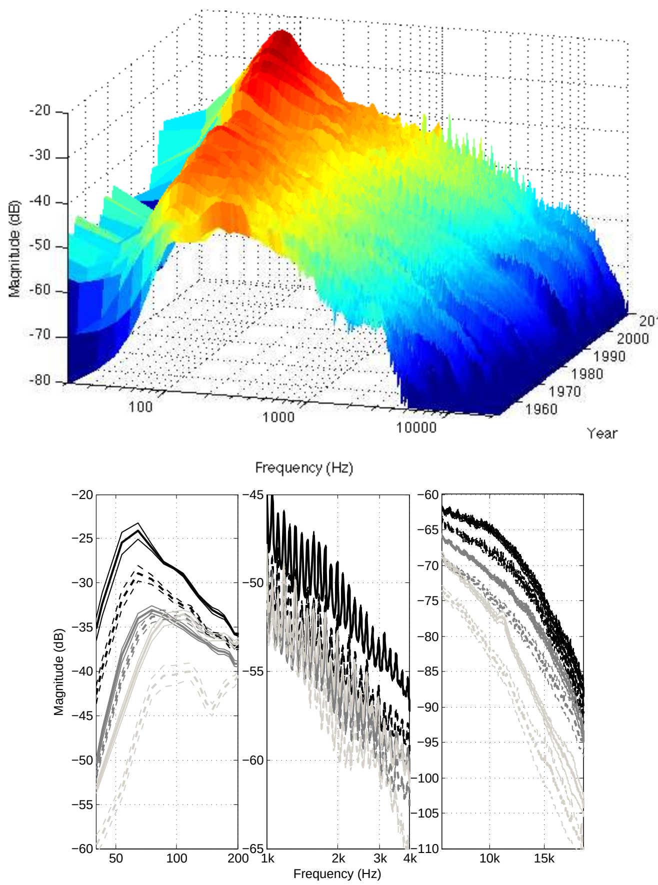

Fig. 2: Average spectra on a yearly base (top) and frequency region details per decade (bottom), from top to bottom: $40 - 200$ Hz, $1 - 4$ kHz and $7 - 20$ kHz. Darker colors represent later decades in the bottom plot.

are found are free from overlaps in the confidence intervals (shown as upper and lower bands, nearly overlapping), which can be interpreted as a rough indicator that group divergences are not due to trivial randomness, but to true underlying differences.

An interesting feature is the raggedness of the mid-distribution (detailed in Figure 3), and particularly its evolution. When we look at the comb-like shape of the line representing the most recent decade, we are seeing peaks in every note of the dodecaphonic scale in equal-tempered western tuning. Looking back in time we see that raggedness emphasizes some notes over others, which may well indicate predominance of certain tonalities over others. This is particularly clear during the 50s and 60s. While this is an interesting point, if we are concerned with equalization practices on the engineering and production side we should discard tonal features and concentrate on the broad spectral contour (see [15] for all data).

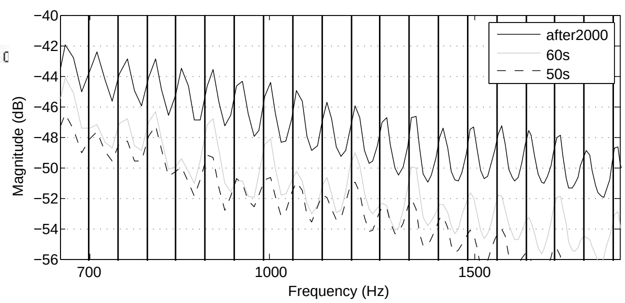

Fig. 3: Detail of the emphasis on tonal frequencies for the decades where the difference is more accentuated. Actual fundamental frequencies are shown as vertical black lines.

There are some additional spectral features whose evolution might be interesting to look at, detailed in Figure 4. Spectral centroid is defined following [16], spectral crest conforms to the formulation given in [17]. We simplified the spectral slope measure, in that it is simply the slope of the log-log regression of the data points between 100 Hz and 10000 Hz. Finally, the spectral peak is purely a measure of the log magnitude of the bin whose value represents the global maximum.

The average magnitude peak and overall magnitude are increasing, and the spectrum tends to become flatter (partly due to the increasing amount of compression, see [18]).

3. DIFFERENCES STEMMING FROM GENRE

Genre differences can also yield interesting results, and these are shown in Figure 5. We took our data from Wikipedia, with tags from EchoNest and LastFM. The extremely extended low end response of electronica and hip-hop is unmistakeable, whereas, as expected, r&b and jazz have a lighter bottom. The prominence of the top-end also yields differences in excess of 10 dB which are meaningful even in the light of the overall magnitude increase

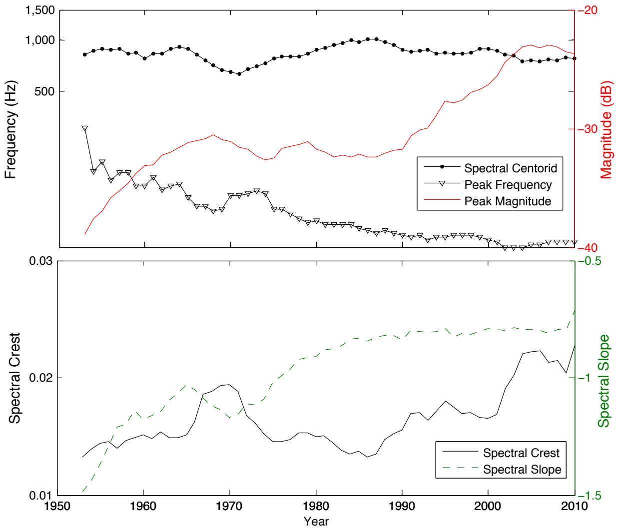

Fig. 4: Yearly evolution of low-level spectral descriptors.

of the brighter genres. The brightest mixes seem to be hip-hop ones, followed by electronic and disco. Here, however, this enhanced top end is negligible when considering that there is an overall enhancement (due to higher loudness specifications). On the dull side, folk and jazz genres suggest that there is a natural top-end decay on more acoustic endeavors, whereas electronic ones allow and benefit from bigger frequency extensions.

On the middle-part of the spectrum, it is interesting to observe that pop and rock seem to be more openly harmonic in nature (again, raggedness in the frequency response), with no preference of tonality. Hip-hop in contrast, seems to have less harmonicity, which may be due to the prominence of rhythmic elements. Note that there might be a bias induced by the number of songs in each genre. The domination of pop and rock in the charts may possibly enhance a more even distribution of tonal content, as there are more songs in more varied keys. We chose not to go into sub-genres, as the academic consensus is very low in terms of genre definition, let alone sub-genres. The genre divisions are much less clear-cut, and the only region with no confidence interval overlap is the low-end.

Table 2 shows the difference in the low-level descriptors mentioned above. These reinforce the observations above in that genre differences are significant in terms of spectra. However, genre-popularity shifts over time. Thus, hip-hop's more prominent loudness and extended bass response is evidently related to the fact that post-2000 songs share the same tendency.

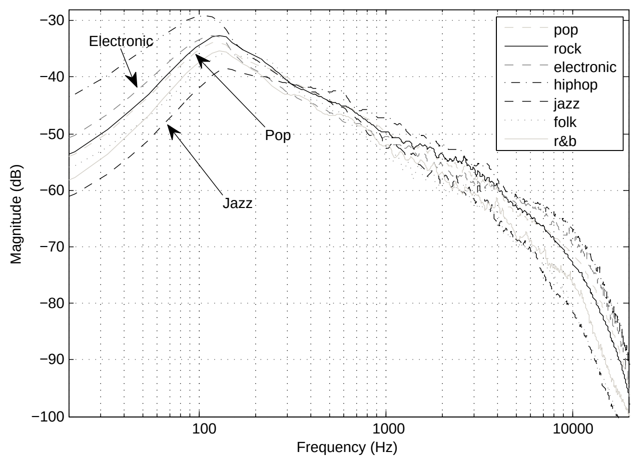

Fig. 5: Average spectra by genre for a selection of genres.

Genre	Spectral Centroid (Hz)	Spectral Crest	Spectral Slope	Peak Magnitude (dB)
Pop	868	0.0158	-0.9433	-30.58
Rock	858	0.0153	-0.9793	-30.66
Elect.	845	0.0194	-0.7461	-27.70
Hip-hop	662	0.0265	-0.8141	-22.52
Jazz	785	0.0141	-1.2929	-35.58
Folk	603	0.0191	-1.1824	-32.54
Disco	963	0.0148	-0.8042	-30.31
R&B	811	0.0149	-1.0336	-33.87
Soul	760	0.0157	-1.0303	-32.94

Table 2: Low-level spectral descriptors compiled by genre.

4. CONCLUSIONS AND FURTHER WORK

It seems that spectra of professionally produced commercial recordings show consistent trends, which can roughly be described as a linearly decaying distribution of around 5 dBper octave between 100 and 4000 Hz, becoming gradually steeper with higher frequencies, and a severe low-cut around 60 Hz. Apart from the artificial low-end boost, this is consistent with the contours of an acoustic ensemble found in previous works. We have also seen that spectra are dependent on genre and on the yearly evolution of production standards.

This knowledge could be useful for a more informed implementation of match-EQ type plug-ins or general equalization contours on playback devices, as well as for automated equalization work [19]. Our analysis was performed on a monaural version of the songs. It could be interesting to extend our approach to differences between the left and right channels in order to understand whether frequency balance on the stereo field is relevant. The comparison between the sum and difference channel spectra could also be revealing. The broad statistical analysis of successful commercial recordings shows a lot of promise for knowledge that could be useful for intelligent systems [20]. Overall Dynamic Range Profiling (DRP), or DRP per frequency band could also give us interesting insights, now that we have loudness and loudness range recommendations that are becoming widespread.

5. ACKNOWLEDGEMENTS

The author P.D. Pestana was sponsored by national funds through the Fundação para a Ciência e a Tecnologia, Portugal, grant number SFRH/BD/65306/2009, and projects:
"PEst-OE/EAT/UI0622/2011"
"PEst-OE/MAT/UI2006/2011"
Support was also provided by EPSRC grant EP/K007491/1, 'Multisource audio-visual production from user-generated content".

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

6. REFERENCES

[1] M. P. Stavrou, "Personal Communication," 2012.

[2] N. Dorfsman, "Personal Communication," 2011.

[3] R. Izhaki, *Mixing Audio — Concepts, Practices and Tools*. Oxford: Elsevier Science & Technology, first ed., 2008.

[4] B. Katz, *Mastering Audio — the Art and the Science*. Oxford: Focal Press, 2007.

[5] L. Sivian, H. Dunn, and S. White, "Absolute Amplitudes and Spectra of Certain Musical Instruments and Orchestras," *Journal of the Acoustical Society of America*, vol. 2, pp. 330–371, 1931.

[6] J. G. McKnight, "The Distribution of Peak Energy in Recorded Music and Its Relation to Magnetic Recording Systems," *Journal of the Audio Engineering Society*, vol. 7, no. 2, pp. 65–71, 1959.

[7] B. Bauer, "Octave-band Spectral Distribution of Recorded Music," *Journal of the Audio Engineering Society*, vol. 18, no. 12, pp. 165–172, 1970.

[8] L. Moller, "How Much has Amplitude Distribution of Music Changed?," in *Proceedings of the 71st AES Convention*, (Montreaux), Audio Engineering Society, 1982.

[9] K. Randall, "BBC RD 1981/2: An Investigation into the Spectral Content of Pop Music," tech. rep., BBC, 1981.

[10] R. Greiner and J. Eggers, "The Spectral Amplitude Distribution of Selected Compact Discs," *Journal of the Audio Engineering Society*, vol. 37, no. 4, pp. 246–275, 1989.

[11] M. Mijić, D. Mašović, M. Petrović, and D. Šumarac Pavlović, "Statistical Properties of Music Signals," in *Proceedings of the 126th AES Convention*, (Munich), Audio Engineering Society, 2009.

[12] D. Tardieu, E. Deruty, C. Charbuillet, and G. Peeters, "Production Effect: Audio Features for Recording Techniques Description and Decade Prediction," in *Proceedings of the 14th International Conference on Digital Audio Effects*, (Paris), DAFx, 2011.

[13] OCC, "http://www.officialcharts.com/," 2012.

[14] Billboard, *Billboard Magazine Charts*. Last accessed: 2012/11/27, 2012.

[15] P. D. Pestana, "PhD Companion Website," 2012.

[16] G. Peeters, "A Large Set of Audio Features for Sound Description (Similarity and Classification) in the CUIDADO project," tech. rep., IRCAM, 2004.

[17] A. Lerch, *An Introduction to Audio Content Analysis*. Hoboken, New Jersey: John Wiley & Sons, first ed., 2012.

[18] E. Vickers, "The Loudness War: Background, Speculation and Recommendations," in *Proceedings of the 129th AES Convention*, (San Francisco), Audio Engineering Society, 2010.

[19] E. P. Gonzalez and J. D. Reiss, "Automatic Equalization of Multi-channel Audio Using Cross-adaptive Method," in *Proceedings of the 127th AES Convention*, (New York), Audio Engineering Society, 2009.

[20] J. D. Reiss, "Intelligent Systems for Mixing Multichannel Audio," in *17th International Conference on Digital Signal Processing (DSP)*, 2011.