mathematics

Article

Partition Entropy as a Measure of Regularity of Music Scales

Rafael Cubarsi

Departament de Matemàtiques, Universitat Politècnica de Catalunya, 08034 Barcelona, Spain; rafael.cubarsi@upc.edu

Abstract: The entropy of the partition generated by an n-tone music scale is proposed to quantify its regularity. The normalized entropy relative to a regular partition and its complementary, here referred to as the bias, allow us to analyze various conditions of similarity between an arbitrary scale and a regular scale. Interesting particular cases are scales with limited bias because their tones are distributed along specific interval fractions of a regular partition. The most typical case in music concerns partitions associated with well-formed scales generated by a single tone h. These scales are maximal even sets that combine two elementary intervals. Then, the normalized entropy depends on each number of intervals as well as their relative size. When well-formed scales are refined, several nested families stand out with increasing regularity. It is proven that a scale of minimal bias, i.e., with less bias than those with fewer tones, is always a best rational approximation of $log_2 h$.

Keywords: metric entropy; convex sets; convex functions; continued fractions; best rational approximation; maximal even sets; well-formed scales; Pythagorean tuning

MSC: 94A17; 11A55; 39B62; 52A20; 11A07

check for updates

Citation: Cubarsi, R. Partition Entropy as a Measure of Regularity of Music Scales. *Mathematics* **2024**, *12*, 1658. https://doi.org/10.3390/math12111658

Academic Editors: M. Filomena Teodoro and Marina Alexandra Pedro Andrade

Received: 16 April 2024
Revised: 20 May 2024
Accepted: 23 May 2024
Published: 25 May 2024

1. Introduction

An n-tone music scale E_n determines a partition of the octave in n intervals. Regarding their regularity, scales can be of equal divisions of the octave, i.e., with n tones of equal temperament (n-TET scale), where discretization precision increases as the number of tones grows, or that of different divisions, where precision increases in two ways: as the size of the intervals decreases and as their regularity increases. In particular, regarding well-formed scales of one generator, hereinafter referred to as cyclic scales, if they are non-degenerate (of equal temperament), they are formed from two elementary intervals [1,2], one longer than the interval of an n-TET scale and the other of shorter width. These scales also fulfill the condition of *maximum evenness* [3,4]; that is, they present the most even distribution, which does not depend on the relative size of the intervals. In many cases, as the number of tones of a cyclic scale increases, the regularity of the intervals decreases, still satisfying the condition of a maximal even set. Therefore, for cyclic scales, it is not obvious how to quantify precision or the regularity of the partition as the number of tones increases. The present work had the purpose of analyzing this using partition entropy, as well as studying the bias from regular temperament in more general cases.

The regularity of the intervals of E_n can be quantified in several ways. A fairly common way is from their standard deviation. In the octave, for $i \in I = \{1, \dots, n\}$ the intervals A_i between consecutive tones, when they are considered a discrete random variable of equal probability such that $\sum_{i \in I} A_i = 1$, have expected value $t = \frac{1}{n}$, i.e., the size of the elementary interval of the n-TET scale. A measure of dispersion of these values about the mean is the standard deviation σ, defined from its square, the variance, $\sigma^2 = \frac{1}{n} \sum_{i \in I} (A_i - t)^2$. In this way, it is possible to compare different n-tone scales in terms of the average dispersion of their intervals relative to the n-TET scale. However, this procedure does not present interesting properties to deal with, for instance, successive refinements of cyclic scales.

Another way to measure the regularity of E_n may be based on the quadratic sum of its intervals. We consider this case in a more general way. Let us assume that $(\mathcal{M}, \mu)$ is a Lebesgue-measurable space such that $\mu(\mathcal{M}) = 1$ and $\alpha = \{A_i\}_{i \in I}$ is a finite measurable partition of $\mathcal{M}$; that is, $\mu(\cup_{i \in I} A_i) = 1$ and $\mu(A_i \cap A_j) = 0$ if $i \neq j$. (Equalities that involve measures are understood to be true except in a null set; that is, we should strictly write $\mu(\mathcal{M} \setminus \cup_{i \in I} A_i) = 0$, but we will not write this to simplify the notation.) Then, the sum of squares for the partition α relative to the measure μ is $S(\alpha) = \sum_{i \in I} \mu(A_i)^2$. The function $S(\alpha)$ is a convex function of $\mu(A_i)$, which, when constrained to $\sum_{i \in I} \mu(A_i) = 1$, has an absolute minimum for $\mu(A_i) = \frac{1}{n}, \forall i$. In this case, $S(\alpha) = \frac{1}{n}$ is a decreasing function of n. When the partition is refined, that is, if $\beta = \{B_j\}_{j \in J}$ is another partition of the octave, then its sum $\alpha \vee \beta = \{A_i \cap B_j\}_{i \in I, j \in J}$ is a refinement of them. In order to make successive refinements, it would be desirable to be able to express $S(\alpha \vee \beta) = S(\alpha) + F(\alpha, \beta)$ with a function F that is a linear combination of squares of $\mu(A_i \cap B_j)$, i.e., also a sum of squares of the refinement. But this is not possible because in F there will necessarily appear cross products.

Fortunately, the notions of regularity and fineness of a partition have, in several fields, a well-known way of being quantified, which is the partition entropy. The concept of entropy was introduced by Clasius in thermodynamics (in 1850), Boltzmann applied it to statistical mechanics (1877), Planck related it to probability theory (1906), Shannon applied it to information theory (1948), Jaynes used it as a measure of uncertainty (1957), and Kolmogorov extended it to deterministic systems (1958).

Certainly, entropy has been used in music almost from the beginning of information theory [5–7] by considering the musical language as a source that produces a sequence of symbols representing musical tones and by associating them with certain probabilities according to their frequency of appearance. In particular, a wide range of works have used entropy to identify music styles (e.g., [8–11]) although results may vary depending on the preanalytical assumptions made in order to treat the data (scale degrees, pitch class, octave equivalence, weighting by duration, key-signature dependency, modal bias, etc.). With the same purpose, cross-entropy (e.g., [12]) has been used to quantify stylistic similarity between two sequences [13]. An up-to-date review on entropy and other physical parameters applied to music is provided by Gündüz [14].

Nevertheless, it seems that entropy as a measure of regularity of music scales has been neglected. There is a significant difference between discrete state spaces and continuous state spaces. In most cases, entropy is considered in terms of information theory, where the natural application is to symbolic frequency analysis. Although Shannon did suggest a generalization to the continuum through infinitesimal equal measure states, becoming an integral form of his discrete theory, partition entropy measures an inverse density rather than a frequency over equal measure.

Therefore, in the current paper, it is proposed to use the normalized entropy as a measure of the regularity. Its application is illustrated in two cases: first, to study the similarity between an arbitrary scale and an n-TET scale; and second, to study the regularity of cyclic scales generated by a tone h and the relationship between their bias and the rational approximation of $\log_2 h$ that they provide.

2. Partition Entropy

Following the notation and terminology of Arnold and Avez [15], the entropy of the partition α is defined from a concave function $z(t)$, as shown below. (The entropy of the partition is also called metric entropy, and it consists of changing the metric of the theory of probabilities, i.e., the probability, by a generic metric, where the concepts relative to the partitions remain.)

$$H(\alpha) = \sum_{i \in I} z(\mu(A_i)) \; ; \quad z(t) = \begin{cases} -t \log_2 t, & 0 < t \leq 1 \\ 0, & t = 0 \end{cases} \tag{1}$$

The following relationship is fulfilled,

$$H(\alpha) = n\,\frac{1}{n}\sum_i z(\mu(A_i)) \leq n\,z\left(\frac{1}{n}\sum_i \mu(A_i)\right) = n\,z(\tfrac{1}{n}) = \log_2 n$$

Let us remember that if the function of a random variable $f(x)$ is concave, the expected value satisfies $\mathrm{E}(f(x)) \leq f(\mathrm{E}(x))$ (Jensen's inequality).

Therefore, for a partition of n elements, the value $\log_2 n$ is the maximum value that the entropy can reach, and it is reached when the elements of the partition are of equal measure.

Thus, $\sum_i z(\mu(A_i)) = \sum_i \mu(A_i)\log_2 \frac{1}{\mu(A_i)}$ is the weighted average of the values $\log_2 \frac{1}{\mu(A_i)}$ (in probabilities, it would be the *information* associated with the measure of the random variable A_i, i.e., the expected value of the information content or self-information of the variable A_i). The smaller $\mu(A_i)$, the greater the above logarithm; so, this average gives an idea of how refined the partition is.

Shannon (1948) showed that the entropy defined in this way is the *only function* that, except for a multiplicative constant, satisfies the following postulates:

1. $H(\alpha)$ is a continuous function of $\mu(A_i)$.
2. If $\mu(A_1) = \cdots = \mu(A_n) = \frac{1}{n}$, then $H(\alpha)$ is an increasing function of n.

Let us consider partitions $\alpha = \{A_i\}_{i\in I}$ and $\beta = \{B_j\}_{j\in J}$ and a refinement of both, $\alpha \vee \beta = \{A_i \cap B_j\}_{i\in I,\,j\in J}$. The conditional entropy of α relative to β is defined as

$$H(\alpha/\beta) = \sum_j \mu(B_j)\sum_i z(\mu(A_i/B_j)) \equiv \sum_j \mu(B_j)\sum_i z\left(\frac{\mu(A_i\cap B_j)}{\mu(B_j)}\right) \quad (2)$$

where $\mu(A_i/B_j)$ is the conditional measure of A_i relative to B_j.

3. The following *sub-additivity* property holds,
$$H(\alpha \vee \beta) = H(\alpha) + H(\beta/\alpha) \leq H(\alpha) + H(\beta) \quad (3)$$
(In general, for three measurable partitions, α, β, and γ, then $H(\alpha \vee \beta/\gamma) = H(\alpha/\gamma) + H(\beta/\alpha \vee \gamma) \leq H(\alpha/\gamma) + H(\beta/\gamma)$.)
4. If φ is an automorfism of $(\mathcal{M}, \mu)$, then $\varphi(\alpha \vee \beta) = \varphi\alpha \vee \varphi\beta$ and $H(\alpha/\beta) = H(\varphi\alpha/\varphi\beta)$.

Therefore, the partition entropy seems like an appropriate parameter to measure how much a scale of n arbitrary tones differs from an n-TET scale. It is also the parameter that provides us with an estimate of how refined a scale is, because the entropy of a scale of n tones will always be greater than that of a subscale of $n-1$ tones, and less than or equal to that of an n-TET scale. This will be useful when dealing with cyclic scales. In addition, as the intervals of a partition are subdivided, the entropy is additive with regard to the intervals being refined, which saves calculations.

3. Cyclic Scales

The computation of entropy for cyclic scales requires a brief review of their properties, which hereinafter are summarized by following Cubarsi [16,17].

A ratio of 2 between frequencies corresponds to the range of one octave. For any frequency ratio $\nu \in \Omega \equiv (0, \infty)$, the values $2^k\nu$, $k \in \mathbb{Z}$, define one equivalence class. The set Ω is a commutative group for multiplication. The set of all the octaves of the fundamental frequency ratio ($\nu_0 = 1$) is a monogenous subgroup of Ω of an infinite cardinal, $\Omega^2 = \{2^k, k \in \mathbb{Z}\}$. The frequency classes are the elements of the quotient group $\Omega_0 = \Omega/\Omega^2$, also commutative for multiplication. For each equivalence class, we choose a representative in $[1, 2)$ (with identified extremes) as the reference octave, which we identify with Ω_0. A finite set of these representatives will be referred to as *scale tones*.

An n-tone cyclic scale E_n^h is a scale of one generator, a real positive value h. The scale tones satisfy the symmetry condition, consisting of displaying several degrees of rotational symmetry that are equivalent to the closure condition [18–20], although such an equivalence does not hold for scales with more than two generators [21,22]. The partition of the octave induced by the scale notes has exactly two sizes of scale steps, and each number of generic intervals occurs in two different sizes, which is known as Myhill's property [1,2]. For $h = 3$, we get the particular case of the 12-tone Pythagorean scale, generated by fifths, as well as those listed in Table 1. In general, for h other than a rational power of 2 (which would lead to degenerate cases of equal temperament scales) we would obtain generalized Pythagorean scales.

Table 1. Properties of cyclic scales: GRA = Good rational approximation (accurate scale); RID = Regular interval distributed; BRA = Best rational approximation (optimal scale); MB = Minimal bias; I | Xn = maximum distance $1/xn$ to a note of the n-TET scale.

n						N	m	M	δ
2						3	1	1	0
3	GRA	RID		I	2n		5	2	1	1
5	GRA	RID	BRA	I	3n	MB	8	2	3	1
7	GRA	RID				11	2	5	0
12	GRA	RID	BRA	I	4n	MB	19	7	5	0
17		RID				27	12	5	1
29	GRA					46	12	17	1
41	GRA	RID	BRA			65	12	29	1
53	GRA	RID	BRA	I	6n	MB	84	12	41	0
94						149	53	41	1
147						233	53	94	1
200	GRA					317	53	147	1
253	GRA					401	53	200	1
306	GRA	RID	BRA	I	2n		485	53	253	1
359	GRA	RID				569	53	306	0
665	GRA	RID	BRA	I	8n	MB	1054	359	306	0
971						1539	665	306	1
1636						2593	665	971	1
2301						3647	665	1636	1
2966						4701	665	2301	1
3631						5755	665	2966	1
4296						6809	665	3631	1
4961						7863	665	4296	1
5626						8917	665	4961	1
6291						9971	665	5626	1
6956						11,025	665	6291	1
7621						12,079	665	6956	1
8286	GRA					13,133	665	7621	1
8951	GRA					14,187	665	8286	1
9616	GRA					15,241	665	8951	1
10,281	GRA					16,295	665	9616	1
10,946	GRA					17,349	665	10,281	1
11,611	GRA					18,403	665	10,946	1
12,276	GRA					19,457	665	11,611	1
12,941	GRA					20,511	665	12,276	1
13,606	GRA					21,565	665	12,941	1
14,271	GRA					22,619	665	13,606	1
14,936	GRA					23,673	665	14,271	1
15,601	GRA	RID	BRA	I	2n		24,727	665	14,936	1
16,266		RID				25,781	665	15,601	0
31,867	GRA	RID	BRA	I	2n		50,508	16,266	15,601	0
47,468		RID				75,235	31,867	15,601	1
79,335	GRA	RID	BRA	I	2n		125,743	31,867	47,468	1
111,202	GRA	RID	BRA			176,251	31,867	79,335	0
190,537	GRA	RID	BRA	I	8n	MB	301,994	111,202	79,335	1
301,739						478,245	111,202	190,537	0
492,276						780,239	301,739	190,537	0
682,813						1,082,233	492,276	190,537	0
873,350						1,384,227	682,813	190,537	0
1,063,887						1,686,221	873,350	190,537	0

The scale tones are $\nu_k = \frac{h^k}{2^{[\![k]\!]}}$; $k = 0, \ldots, n-1$; with $[\![k]\!] = \lfloor k \log_2 h \rfloor$ (floor function). When the scale tones are ordered from lowest to highest pitch in $[1, 2)$ (say, in *cyclic order* or *by ordinal*), we find two extreme tones—the minimum tone ν_m and the maximum tone ν_M—which determine the two elementary factors $U = \nu_m = \frac{h^m}{2^{[\![m]\!]}}$ (up the fundamental) and $D = \frac{2}{\nu_M} = \frac{2^{[\![M]\!]+1}}{h^M}$ (down the fundamental) associated with the generic widths of the step interval such that $U^M D^m = 2$. The indices satisfy $n = m + M$, which are all coprime.

The tone $\nu_n = \frac{h^n}{2^{[\![n]\!]}}$, which does not belong to the scale E^h_n, provides the closure condition (either $\nu_n \to 1^+$ or $\nu_n \to 2^-$) determining the n-order comma $\kappa_n = \min(\nu_n, \frac{2}{\nu_n})$ (in the frequency space), i.e., the error in closing the scale near the fundamental *with no other scale tones between them*. The comma itself does not provide information about whether ν_n closes above or below the fundamental. In using the index $N = [\![m]\!] + [\![M]\!] + 1 = \lfloor n \log_2 h + \frac{1}{2} \rfloor$, two parameters provide this information: on the one hand, the *scale closure*, $\gamma_n = \frac{h^n}{2^N}$, which is a value close to 1 satisfying $\frac{U}{D} = \gamma_n$; and on the other hand, the *scale digit* $\delta = N - [\![n]\!]$, taking values 0 or 1. Then, $\delta = 0 \iff \gamma_n > 1$ $(\nu_n \to 1^+, \gamma_n = \kappa_n)$ or $\delta = 1 \iff \gamma_n < 1$ $(\nu_n \to 2^-, \gamma_n = \kappa_n^{-1})$. The value $|\log_2 \gamma_n| = \log_2 \kappa_n$ measures the distance from ν_n to 1. (The distance between two scale tones α, β is measured as $d(\alpha, \beta) = \min(|\log_2 \frac{\alpha}{\beta}|, 1 - |\log_2 \frac{\alpha}{\beta}|)$.)

Figure 1 (left and central panels) displays how the 12-tone cyclic scale for $h = 3$ (Pythagorean scale) is formed. The h iterates of indices $m = 7$ and $M = 5$ are the extreme tones, while $n = 12$ provides the closure, since there is no other tone with a frequency that is between ν_{12} and the fundamental. The ratio between consecutive iterates is either $\frac{3}{2}$ or $\frac{3}{4}$, except for the last iterate (the wolf fifth), which compensates for the comma. When the scale tones are arranged in pitch order, i.e., by ordinals, their ratios are either $U = \frac{3^7}{2^{11}}$ or $D = \frac{2^8}{3^5}$. Such a distinction is much clearer under the following alternative approach.

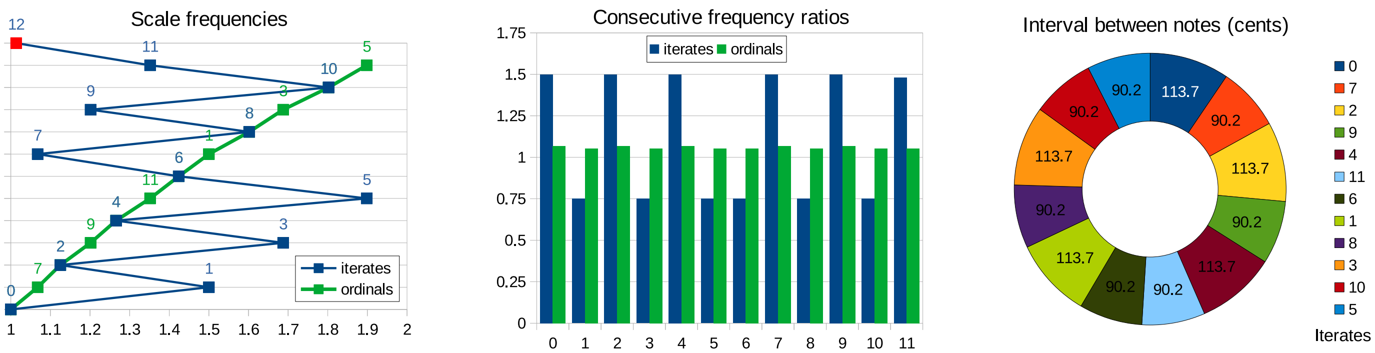

Figure 1. (**Left**) Tones (frequencies ν_k) of the cyclic scale for $h = 3$ and $n = 12$ in $[1, 2)$ in order of iterates (blue) and in pitch order (green). The red dot is assumed as closing the scale. (**Center**) Ratios between iterates (blue) and consecutive scale tones (green). (**Right**) Scale intervals in cents ($1200 \log_2 \nu_k$) along the circle of the octave clockwise direction (number in white for the first one) with two interval sizes.

Each tone ν_k in the frequency space is associated with a *note* or pitch class $\log_2 \nu_k$ in the octave $S_0 = \mathbb{R}/\mathbb{Z}$, so that the above quantities have the corresponding ones in S_0. Thus, the elementary intervals $u = \log_2 U$ and $d = \log_2 D$ generate the partition of the octave in n intervals, with $Mu + md = 1$, satisfying $u - d = \phi_n$, with interval closure $\phi_n = \log_2 \gamma_n$ and interval comma $|\phi_n| = \log_2 \kappa_n$. Sometimes, it is more useful to work in the multiplicative space of tones, and sometimes, in the additive space of notes. In the latter case, a frequency $x \in [1, 2)$ is usually expressed by musicians as $1200 \log_2 x \in [0, 1200)$ in *cents* (¢), so that each semitone of the 12-TET scale is divided into 100 parts. Figure 1 (right panel) displays how the intervals between the 12 notes of the Pythagorean scale are distributed along the circle of the octave (clockwise direction). All of the intervals, even the last one, are either $u = 113.7$¢ or $d = 90.2$¢.

The fraction $\frac{N}{n}$ is associated with the convergent and semi-convergent continued fraction expansions of $\log_2 h$ [23–25]. Among cyclic scales, two categories may be pointed out. The first category is formed by *optimal scales*, associated with the best closure $\gamma_n \approx 1$, corresponding to the *best rational approximations*, i.e., the convergents of its canonical continued fraction expansions from both sides. The second category, which we shall name *accurate scales*, is associated with the best estimations of the generator tone $\frac{2^{\frac{N}{n}}}{h} \approx 1$, corresponding to the *good rational approximations* $\frac{N}{n}$ of $\log_2 h$. Apart from *accurate scales*, which include *optimal scales*, there are cyclic scales not associated with good or best rational approximations, still corresponding to semiconvergents.

Notice that the terms "good" and "best" rational approximations [26] are equivalent to the best approximation "of the first kind" and "of the second kind", respectively [27]. The conditions mean the following.

A one-sided best approximation of $\log_2 h^+$ occurs when $\gamma_n < 1$ and $([\![k]\!] + 1) - k\log_2 h > N - n\log_2 h > 0,\ 0 < k \le n$.

A one-sided best approximation of $\log_2 h^-$ occurs when $\gamma_n > 1$ and $0 < n\log_2 h - N < k\log_2 h - [\![k]\!],\ 0 < k \le n$.

A best rational approximation satisfies $0 < |n\log_2 h - N| < |k\log_2 h - ([\![k]\!] + 1)|$ and $0 < |n\log_2 h - N| < |k\log_2 h - [\![k]\!]|,\ 0 < k \le n$.

A one-sided good approximation of $\log_2 h^+$ occurs when $\gamma_n < 1$ and $\frac{[\![k]\!]+1}{k} - \log_2 h > \frac{N}{n} - \log_2 h > 0,\ 0 < k \le n$.

A one-sided good approximation of $\log_2 h^-$ occurs when $\gamma_n > 1$ and $0 < \log_2 h - \frac{N}{n} < \log_2 h - \frac{[\![k]\!]}{k},\ 0 < k \le n$.

A good rational approximation satisfies $0 < |\log_2 h - \frac{N}{n}| < |\log_2 h - \frac{[\![k]\!]+1}{k}|$ and $0 < |\log_2 h - \frac{N}{n}| < |\log_2 h - \frac{[\![k]\!]}{k}|,\ 0 < k \le n$.

For all these scales, the values U, D, and γ_n have bounds according to Appendix A. In particular, $|\log_2 \gamma_n|$ quantifies the error of the rational approximation of $\log_2 h$. These bounds determine the interval between a note of the cyclic scale E_n^h and the one with the same ordinal in the n-TET scale, which is analyzed in Appendix B.

Finally, the family of cyclic scales follow a chain, $E_n^h \subset E_{n^+}^h$, so that in starting from the indices (m, M) of E_n^h, the same values for the next scale $E_{n^+}^h$ are (i) $m^+{=}m + M,\ M^+{=}M \iff \delta{=}0$ and (ii) $m^+{=}m,\ M^+{=}m + M \iff \delta{=}1$ (see Table 1).

4. Entropy of a Cyclic Scale

For a cyclic scale E_n^h, with the measure defined for tone intervals as $\mu([x_0, x_1]) = \log_2 \frac{x_1}{x_0}$ for $0{<}x_0{\le}x_1$, consider the partition of the octave $\alpha{\equiv}(u^M, d^m)$ composed of M intervals of width u and m intervals of width d, regardless of the order in which the intervals follow each other. The partition entropy is

$$H(u^M, d^m) = Mz(u) + mz(d) \tag{4}$$

We will explicitly write it in terms of the indices of the minimum and maximum tones m and M. In order to do so, we define the values

$$|x\rangle = \{\log_2 h^x\}, \quad \langle x| = 1 - \{\log_2 h^x\}; \qquad x \in R$$

where $\{\log_2 x\}$ is the mantissa of $\log_2 x$, that is, $\log_2 h^x - \lfloor \log_2 h^x \rfloor$, such that the elemental intervals u and d can be expressed in terms of the indices of the minimum and maximum tones as

$$u = |m\rangle, \quad d = \langle M| \tag{5}$$

In this way, Equation (4) becomes

$$H(|m\rangle^M, \langle M|^m) = Mz(|m\rangle) + mz(\langle M|) \tag{6}$$

Let us see how to express the entropy of the cyclic scale $E^h_{n^+}$, the one following E^h_n in the chain of cyclic scales. The scale refinement process is as follows. The partition (u^M, d^m) breaks, so that the major interval splits into two, one of the same size as the minor one plus a remainder. This residual is only smaller than the size of the smaller interval when the scale is optimal. This process is iterated.

As explained in Section 3, we must distinguish two cases, depending on whether γ_n is greater or less than 1, or equivalently, if δ equals 0 or 1:

(i) $\delta = 0$, $\gamma_n > 1$, $U > D$, $u > d$. In this case, $N = [\![n]\!]$, and $m^+ = n, \quad M^+ = M$.

The refinement is performed in the M intervals of size u, and the m intervals of size d are maintained:

$$\begin{array}{l} u' = u - d = \phi_n \iff u' \equiv |m^+\rangle = |m\rangle - \langle M| \\ d' = d \iff d' \equiv \langle M^+| = \langle M| \end{array}$$

Since $b[z(\frac{a}{b}) + z(1 - \frac{a}{b})] = z(a) - z(b) + z(b-a)$ and $u - d = \kappa_n$, it holds that $u\ H(\frac{d}{u}, 1 - \frac{d}{u}) = z(d) - z(u) + z(\kappa_n)$. Thus, we express the entropy according to Equation (3), with the new partition $\alpha \vee \beta = ({u'}^{M^+}, {d'}^{m^+})$:

$$\begin{aligned} H({u'}^{M^+}, {d'}^{m^+}) &= H(u^M, d^m) + M\ u\ H(\tfrac{d}{u}, 1 - \tfrac{d}{u}) = \\ &= H(u^M, d^m) + M\ [z(d) - z(u) + z(\kappa_n)] \end{aligned}$$

In terms of the indices of the extreme tones, in noting $H_n = H(E^h_n)$, it can be written as

$$H_{n^+} = H_n + \Delta_M; \quad \Delta_M = M\ [z(\langle M|) - z(|m\rangle) + z(\kappa_n)] \tag{7}$$

(ii) $\delta = 1$, $\gamma_n < 1$, $U < D$, $u < d$. In this case, $N = [\![n]\!] + 1, \quad m^+ = m$, and $\quad M^+ = n$.

The refinement is performed in the m intervals of size d, and the M intervals of size u are maintained:

$$\begin{array}{l} d' = d - u = -\phi_n \iff d' \equiv \langle M^+| = \langle M| - |m\rangle \\ u' = u \iff u' \equiv |m^+\rangle = |m\rangle \end{array}$$

The entropy of the new partition $\alpha \vee \beta = ({u'}^{M^+}, {d'}^{m^+})$ is

$$\begin{aligned} H({u'}^{M^+}, {d'}^{m^+}) &= H(u^M, d^m) + m\ d\ H(\tfrac{u}{d}, 1 - \tfrac{u}{d}) = \\ &= H(u^M, d^m) + m\ [z(u) - z(d) + z(\kappa_n)]) \end{aligned}$$

In terms of the indices of the extreme tones, they are equivalent to

$$H_{n^+} = H_n + \Delta_m; \quad \Delta_m = m\ [z(|m\rangle) - z(\langle M|) + z(\kappa_n)] \tag{8}$$

Figure 2 displays how entropy increases in terms of the number of tones of a Pythagorean scale. The left panel refers to the scales with a lower number of tones, while the right panel shows the larger trend (in logarithmic scale), by making explicit the values of δ.

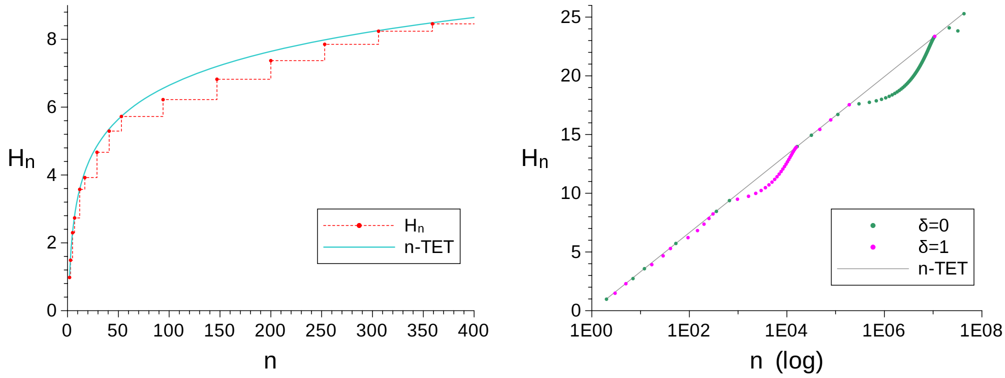

Figure 2. Small and large scale trends of the entropy for cyclic scales with $h = 3$.

5. Partition Modulation

According to Equation (A2), we write the elementary intervals of a cyclic scale as

$$u = \frac{1+m\phi_n}{n}, \quad d = \frac{1-M\phi_n}{n} \tag{9}$$

We have referred to one degenerate case of a cyclic scale, the limiting case when the two intervals are one, $u \to d$. In this case, $u = d = \frac{1}{n}$ and $E_n^h = E_n^\top$, the n-TET scale. This is equivalent to $\phi_n \to 0$ in the expressions of Equation (9). However, we should consider two more degenerate cases. The larger is one interval, the smaller is the other, always filling a full octave, $Mu + md = 1$. Therefore, if $d \to 0$, then the cyclic scale becomes an equal temperament scale of M tones, and if $u \to 0$, the scale becomes an equal temperament scale of m tones. In other words, one of the following cases applies:

(i) According to Appendix A, for $\phi_n > 0$, by Equation (A5), $\phi_n < \frac{1}{M}$. If $\phi_n \to \frac{1}{M}$, then $d \to 0$. In this case, the intervals of the cyclic scale satisfy $0 < d < \phi_n < u$. Since $u - d = \phi_n$, then when $u \to \phi_n$, the scale is not non-degenerate anymore and becomes an M-TET scale, with entropy

$$H_n = H(u^M, d^m) \xrightarrow[d \to 0]{} H(u^M) = H_M = \log_2(M)$$

Obviously, an optimal cyclic scale is far from this situation, because it satisfies $0 < \phi_n < d < u$.

(ii) Also, according to Appendix A, for $\phi_n < 0$, owing to Equation (A8), $\phi_n > -\frac{1}{m}$. If $\phi_n \to -\frac{1}{m}$, then $u \to 0$. In this case, the intervals of the cyclic scale satisfy $0 < u < |\phi_n| < d$. Then, $d \to |\phi_n|$, so that the scale becomes a degenerate m-TET scale with entropy

$$H_n = H(u^M, d^m) \xrightarrow[u \to 0]{} H(d^m) = H_m = \log_2(m)$$

An optimal cyclic scale case is also far from this situation, since it satisfies $0 < |\phi_n| < u < d$.

Lemma 1. *The scales formed from the two elementary intervals $u' = \frac{1+m\xi}{n}$ and $d' = \frac{1-M\xi}{n}$ with values $-\frac{1}{m} < \xi < \frac{1}{M}$ generate an infinite and continuous family of n-tone cyclic scales $C_n^h(\xi)$, which are neighbors of E_n^h, such that $C_n^h(\phi_n) = E_n^h$ and $C_n^h(0) = E_n^\top$. In addition, if $-\frac{1}{n+m} < \xi < \frac{1}{n+M}$, the scales $C_n^h(\xi)$ are optimal.*

These results are immediate consequence of the bounds obtained in Appendix A.

5.1. Modulating Temperament Scales

The most usual case of cyclic scale E_n^h is the one associated with a generator corresponding to a harmonic $h \in \mathbb{Z}^+$ of the fundamental tone. Then, the tonal class of the generator is $g = \frac{h}{\lfloor \log_2 h \rfloor} \in (1,2)$, which can be written as $g = 2^{\frac{\mu+\phi n}{n}}$, where $\mu = N - [\![1]\!] n$ is the ordinal of the scale tone that better approximates g [17], known as the chromatic length of the pitch class of the generator. The value $\mu = \lfloor n \log_2 g + \frac{1}{2} \rfloor$ also determines the indices m and M of the extreme tones, corresponding to the ordinals 1 and $n-1$ of the scale notes. Since as $\mu\, m - n\, [\![m]\!] = 1$, m is the positive integer, $0 < m < n$ so that $\mu\, m = 1 \bmod n$.

For fixed n (and also m and M), the family of cyclic scales $C_n^h(\xi)$ has generators that, in general, are not the tonal class of any harmonic; that is, they are real scales. In this case, we write them as $g' = 2^{\frac{\mu+\xi}{n}}$ so that, for $\xi \in (-\frac{1}{m}, \frac{1}{M})$, we obtain a partition modulation, which is a continuum of irregular temperaments [28,29] close to E_n^h and $E_n^\top$, also corresponding to cyclic scales. Hence, the n-tone cyclic scales $C_n^h(\xi)$ are a family of *modulating temperament scales* around the generator $2^{\frac{\mu}{n}}$, namely, $\mathcal{T}_n^\mu = \{C_n^h(\xi),\ \xi \in (-\frac{1}{m}, \frac{1}{M})\}$.

For example, for $n = 12$ and $g' = \sqrt[4]{5}$, the value $\xi = \frac{12}{4}\log_2 5 - \lfloor \frac{12}{4}\log_2 5 + \frac{1}{2} \rfloor = -0.03422$ gives rise to the *quarter-comma meantone temperament*, which fits the class of the fifth harmonic well, in exchange for decreasing by 5¢ in the accuracy of the third one. Since the value of $|\xi|$ is lower than $\frac{1}{n+m} = \frac{1}{19} = 0.05263$, it results in an optimal scale.

5.2. Entropy in Terms of the Closure

Let us consider an n-tone modulating the temperament scale $C_n^h(\xi) \in \mathcal{T}_n^\mu$. According to Equations (4) and (9), we write the entropy of $C_n^h(\xi)$ as

$$H_n = \mathcal{H}(m, M, \xi) = Mz(\tfrac{1+m\xi}{n}) + mz(\tfrac{1-M\xi}{n})$$

In using the relationship $z(\frac{a}{b}) = \frac{1}{b}z(a) + az(\frac{1}{b})$, it is immediate to see that

$$\mathcal{H}(m, M, \xi) = \log_2 n + \tfrac{M}{n} z(1+m\xi) + \tfrac{m}{n} z(1-M\xi)$$

that is,

$$\mathcal{H}(m, M, \xi) = \log_2 n - \tfrac{M}{n}(1+m\xi)\log_2(1+m\xi) - \tfrac{m}{n}(1-M\xi)\log_2(1-M\xi) \qquad (10)$$

By deriving Equation (10) with respect to ξ, we obtain

$$\frac{\partial \mathcal{H}}{\partial \xi} = -\frac{Mm}{n}\left[\log_2(1+m\xi) - \log_2(1-M\xi)\right] \qquad (11)$$

If $0 < \xi < \frac{1}{M}$, the two terms of the previous equation are positive and the derivative is negative; therefore, in this interval, the entropy is a decreasing function. If $-\frac{1}{m} < \xi < 0$, both terms in the above equation are negative and the derivative is positive, so in this interval, the entropy is an increasing function. At $\xi = 0$, there is a local maximum, since $\log_2(1+m\xi) - \log_2(1-M\xi) = 0$ if and only if $\xi = 0$. Thus, $\mathcal{H}$ in terms of ξ is a concave function, since $\frac{\partial^2 \mathcal{H}}{\partial \xi^2} = -\frac{Mm}{n \ln 2}(\frac{m}{1+m\xi} + \frac{M}{1-M\xi}) < 0$. Therefore, the following apply:

(a) If $0 < \xi < \frac{1}{M}$, then $\mathcal{H}(m, M, 0) = \log_2 n > \mathcal{H}(m, M, \xi) > \log_2 M = \mathcal{H}(m, M, \frac{1}{M})$.
(b) If $-\frac{1}{m} < \xi < 0$, then $\mathcal{H}(m, M, -\frac{1}{m}) = \log_2 m < \mathcal{H}(m, M, \xi) < \log_2 n = \mathcal{H}(m, M, 0)$.

In addition, it is straightforward to see the following properties:

(c) If $m < M$, $\mathcal{H}(m, M, \xi) < \mathcal{H}(m, M, -\xi)$ for $\xi > 0$ and $\mathcal{H}(m, M, \xi) > \mathcal{H}(m, M, -\xi)$ for $\xi < 0$.
(d) If $m > M$, $\mathcal{H}(m, M, \xi) > \mathcal{H}(m, M, -\xi)$ for $\xi > 0$ and $\mathcal{H}(m, M, \xi) < \mathcal{H}(m, M, -\xi)$ for $\xi < 0$.
(e) $\mathcal{H}(m, M, \xi) > \mathcal{H}(M, m, -\xi)$ for $-\frac{1}{m} < \xi < \frac{1}{M}$.

Properties (c) and (d) refer to the slight asymmetry of the entropy far from $\xi = 0$, which, as we shall see in the last section, is negligible around $\xi = 0$. Property (e) means that a cyclic scale generated by h and its inverse scale, generated counter-clockwise with swapped indices of the extreme tones, although they have different notes that differ in one comma, have the same entropy.

Therefore, the entropy of any n-tone modulating temperament scale in the family $\mathcal{T}_n^{\mu}$ only depends on the closure.

6. Normalized Entropy

Since the entropy increases as the partition is refined, to be able to compare qualities of different scales, whether they are cyclic or not, we will consider the *normalized entropy*, i.e., relative to the maximum value that it can reach, as that of the equally tempered scale. For an n-tone scale E_n, it is defined as

$$\eta(E_n) = \tfrac{1}{\log_2(n)}\, H(E_n); \quad 0 \leq \eta \leq 1$$

such that η tends to 1 as E_n approaches an n-TET scale. This ratio has also been called efficiency and relative entropy [30]. For a cyclic scale E_n^h, we will write $\eta_n = \eta(E_n^h)$.

Let us see in which case the normalized entropy increases when a scale is being refined. Let us assume two scales, not necessarily cyclic: $E_n \subset E_{n^+}$ with $n < n^+$ (i.e., E_{n^+} is a refinement of E_n).

Lemma 2. *The normalized entropy increases if and only if the relative increment of entropy in refining the partition is greater than the relative increment of the entropy of the corresponding equal temperament scales.*

Proof. By the sub-additivity property, $H(E_{n^+}) = H(E_n) + \Delta$. It will be $\eta(E_n) < \eta(E_{n^+})$ if and only if $\frac{H(E_n)}{\log_2 n} < \frac{H(E_{n^+})}{\log_2 n^+}$, so $\frac{H(E_{n^+})}{\log_2 n} < \frac{H(E_{n^+})+\Delta}{\log_2 n^+}$. Hence,

$$\frac{\log_2 n^+ - \log_2 n}{\log_2 n} < \frac{\Delta}{H(E_n)} \tag{12}$$

□

The deviation of an n-tone scale E_n, not necessarily cyclic, relative to the regular n-TET scale will be measured from the complementary of the normalized entropy, which we will call *bias*:

$$\theta(E_n) = 1 - \eta(E_n); \quad 0 \leq \theta \leq 1 \tag{13}$$

7. Bias of a Cyclic Scale

For a non-degenerate n-tone cyclic scale $C_n^h(\xi) \in \mathcal{T}_n^{\mu}$, the bias $\theta(C_n^h(\xi)) = \theta_n(\xi)$ depends on the divergence of its elementary intervals u and d with regard to the elementary interval of the n-TET scale. Thus, in bearing in mind Equation (10),

$$\theta_n(\xi) = \tfrac{1}{n \ln n}\, [M(1+m\xi)\ln(1+m\xi) + m(1-M\xi)\ln(1-M\xi)] \tag{14}$$

The graph of the function θ_n is the mirror image up-to-down of H_n, scaled by the factor $\frac{1}{\log_2 n}$; hence, it is convex (Figure 3). The value $\theta_n(0) = 0$ is its minimum in the interval $\xi \in (-\frac{1}{m}, \frac{1}{M})$, and at the extremes (corresponding to degenerate scales), it takes the values $\theta_n(-\frac{1}{m}) = 1 - \frac{\ln m}{\ln n}$ and $\theta_n(\frac{1}{M}) = 1 - \frac{\ln M}{\ln n}$. Therefore, in the interval $(-\frac{1}{m}, 0)$, $\theta_n(\xi)$ is a decreasing function, and in $(0, \frac{1}{M})$, it is an increasing function of ξ.

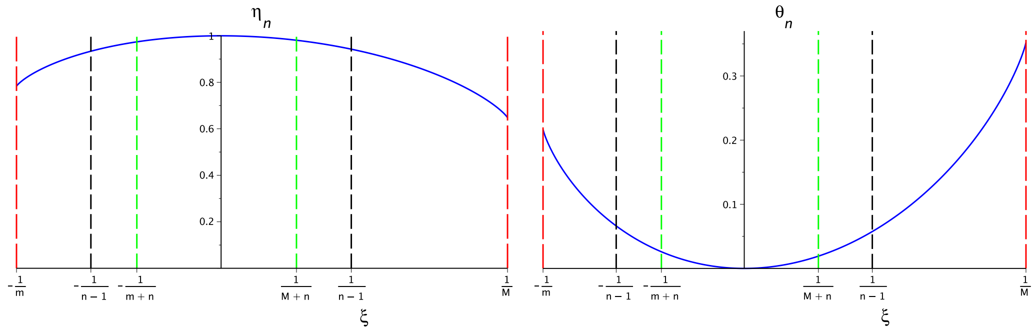

Figure 3. Graphs of normalized entropy and bias.

Note that this behavior of θ_n in terms of ξ holds when n is fixed. In other words, $\theta_n = \theta(\xi; m, M)$. Then, although for another n'-tone cyclic scale, $n' \neq n$, the bias $\theta_{n'}$ has a similar behavior with respect to ξ', we cannot assure that if $\xi < \xi'$, then $\theta(\xi; m, M) < \theta(\xi'; m', M')$, since this also depends on the values of m' and M', such that $m' + M' = n'$.

Figure 4 shows the trends of the bias θ_n and the interval comma $|\phi_n|$ for Pythagorean scales (generated by $h = 3$) in terms of n. In general, optimal scales have low bias, but the bias not always decreases as a scale is refined. For example, the first optimal scales are those of $n = 5; 12; 41; 53; 306; 665 \ldots$, and the bias decreases for $n = 5; 12; 53; 665$, but the scale of $n = 41$ has greater bias than that of $n = 5; 12$, and the one of $n = 306$ has greater bias than those of $n = 12; 53$. Also notice that there are intervals where $|\phi_n|$ decreases but θ_n increases.

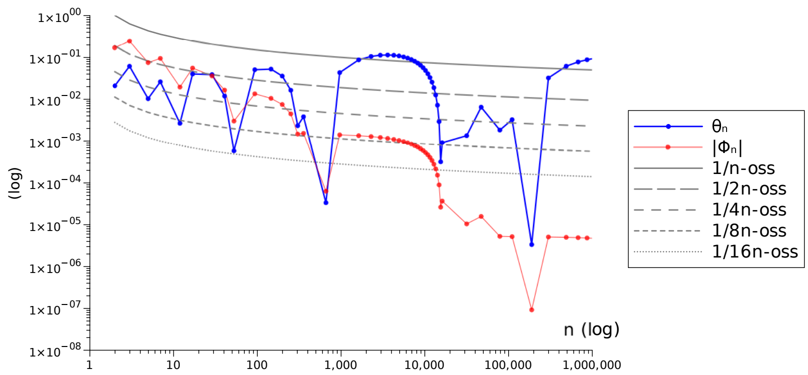

Figure 4. Behavior of $|\phi_n| = \log_2 \kappa_n$ (red) and $\theta_n = 1 - \eta_n$ (blue) in terms of the number of tones n (bilogarithmic scales). In gray is maximum bias for several similarity levels to n-TET scales.

We will say that a cyclic scale E_n^h is of *minimal bias* (MB) if for any cyclic scale $E_{n'}^h$ with $n' < n$, $\theta_n < \theta_{n'}$. Obviously, it is tantamount to say that $\eta_n > \eta_{n'}$; hence, the scale has greater normalized entropy than the previous ones. Not all optimal cyclic scales are MB.

Before a deeper analysis, in an approximate way, we may estimate how close an MB scale and an equal temperament scale of the same number of tones are. We may use the criterion for which a cyclic scale (not necessarily optimal) unambiguously approximates an n-TET scale if every note of the former is at a distance equal or less than half an elementary interval from the latter. According to Equation (A12) with $\lambda = 2$, this leads to the condition $\text{¢}(\kappa_n) \leq \frac{600}{n-1}$ (condition I|2n).

This condition can be satisfied by optimal and non-optimal cyclic scales. For example, for $n = 3$, condition I|2n is holds although the scale is not optimal. On the contrary, $n = 41; 306; 111\,202$ are optimal scales but the condition I|2n is not met.

Let us see what happens with the MB condition. In this case, for up to $n = 2 \cdot 10^5$, all scales that satisfy MB also satisfy I|2n, but there are scales that are I|2n and not MB scales, for example, for $n = 3; 306; 15\,601; 31\,867; 79\,335$. Hence, the condition that a note of the cyclic scale is closer than half an interval of a note of a tempered scale is weaker than the

MB condition. On the contrary, if we further restrict the above condition, for example, the respective notes are at most at a third interval, i.e., $¢(\kappa_n) \leq \frac{400}{n-1}$ (condition I | 3n), then for up to $n = 2 \cdot 10^5$, all I | 3n scales are MB scales. In Table 1, these and other properties are displayed.

8. Scales with Limited Bias

8.1. Scale Distributed within Regular Intervals

We study several families of scales for which it is possible to estimate a priori the lower limit of the entropy.

We say that two n-tone scales E_n and E'_n sharing the fundamental tone *alternate* if their notes in cyclic order—$s_j = \log_2 \sigma_j \in E_n$ and $s'_j = \log_2 \sigma'_j \in E'_n$, where $0 < j < n$—fulfill one of the following conditions:

$$\begin{array}{lll} \text{(a)} & s_{j-1} < s'_j \leq s_j; \quad 0 < j < n & (E_n \text{ alternates by the right of } E'_n\,) \\ \text{(b)} & s'_{j-1} < s_j \leq s'_j; \quad 0 < j < n & (E_n \text{ alternates by the left of } E'_n\,) \end{array} \tag{15}$$

If E'_n is the equal temperament scale $E_n^{\top}$, then we write $\sigma'_j = \vartheta'_j$ and say that E_n is *regular interval distributed* (RID). In this case, the alternation can also be defined from the following relations involving intervals between tones,

$$\text{(a)} \quad 0 \leq I(\sigma_j, \vartheta'_j) < \frac{1}{n} \qquad \text{(b)} \quad 0 \leq I(\vartheta'_j, \sigma_j) < \frac{1}{n}; \qquad 0 < j < n \tag{16}$$

(The interval between two tones $\nu_1 < \nu_2$ in $(0, \infty)$ is $I(\nu_2, \nu_1) = \log_2 \nu_2 - \log_2 \nu_1$.)

Note that these conditions are generally more restrictive than the condition

$$d(\sigma_j, \vartheta'_j) < \frac{1}{n} \iff -\frac{1}{n} < I(\sigma_j, \vartheta'_j) < \frac{1}{n} \tag{17}$$

Obviously, Equation (16) implies the condition of Equation (17), although, as seen in Appendix B, for cyclic scales, they are equivalent.

For example, suppose case (a). If $0 \leq I(\sigma_j, \vartheta'_j) < \frac{1}{n}$, then the interval between two consecutive notes of E_n satisfies $I(\sigma_j, \sigma_{j+1}) = I(\sigma_j, \vartheta'_{j+1}) + I(\vartheta'_{j+1}, \sigma_{j+1}) = \frac{1}{n} - I(\sigma_j, \vartheta'_j) + I(\vartheta'_{j+1}, \sigma_{j+1}) < \frac{2}{n}$, so that $d(\sigma_j, \sigma_{j+1}) < \frac{2}{n}$. Instead, if $d(\sigma_j, \vartheta'_j) < \frac{1}{n}$, then $d(\sigma_j, \sigma_{j+1}) \leq d(\sigma_j, \vartheta'_j) + d(\vartheta'_j, \vartheta'_{j+1}) + d(\sigma_{j+1}, \vartheta'_{j+1}) < \frac{3}{n}$. In a similar way, we would reason case (b).

Lemma 3. *The interval between two consecutive notes of an n-tone RID scale is lower than $\frac{2}{n}$.*

8.2. Scale r-Similar to n-TET

A criterion for measuring the proximity between scales is *similarity* [31]. Two n-tone scales T and T' are similar at level $r > 0$ (r-similar) if for each tone $\tau_i \in T$, $\exists \tau'_j \in S$ such that $d(\tau_i, \tau'_j) \leq r$ and $d(\tau_i, \tau'_k) > r, \forall k \neq j$.

For example, a cyclic scale E_n^h and the n-TET scale have a level of similarity $\frac{1}{2n}$ if and only if any pair of notes with ordinal j, i.e., $\vartheta_j \in E_n^h$ and $\vartheta'_j \in E_n^{\top}$, satisfy $d(\vartheta_j, \vartheta'_j) \leq \frac{1}{2n}$. This condition is equivalent to the condition of Equation (A12) with $\lambda = 2$, $¢(\kappa_n) \leq \frac{600}{n-1}$.

However, for the current purpose of evaluating and comparing entropy partitions, we will slightly modify such a concept by assuming that the fundamental is shared by both scales. Then, we say a scale E_n is *similar* to the n-TET scale at level r, with $0 < r \leq \frac{1}{2n}$, if their tones satisfy

$$d(\sigma_j, \vartheta'_j) \leq r\,; \quad 0 < j < n \tag{18}$$

8.3. Scale One-Side r-Similar to n-TET

We introduce a concept that will mix the one of similarity with the one of distribution within regular intervals, which will be appropriate for studying the bounds of the entropy for cyclic scales.

Let ϑ'_j be the tones of the n-TET scale. We say that a scale E_n is *one-side r-similar* at level r (r-OSS) to the n-TET scale, with $0 < r \leq \frac{1}{n}$, if

$$\text{(a)} \quad 0 \leq I(\sigma_j, \vartheta'_j) < r \qquad \text{(b)} \quad 0 \leq I(\vartheta'_j, \sigma_j) < r; \qquad 0 < j < n \tag{19}$$

It is OSS by the right (a) or by the left (b), respectively. Therefore, when $r = \frac{1}{n}$ it matches the definition of an RID scale.

9. Entropy of a Scale $\frac{1}{\lambda n}$-OSS to n-TET

Consider an n-tone scale E_n, not necessarily cyclic, which is $\frac{1}{\lambda n}$-OSS to n-TET for $\lambda \geq 1$. We already know that the maximum entropy is $\log_2 n$. Let us see the minimum entropy it can reach. We assume case (a) of Equation (19) and consider the n notes ($n \geq 2$) of E_n in cyclic order in $[0, 1) \cup \{1\}$, since the entropy does not change if we add a null set. We also extend the range of the possible variation in the intervals between notes at the extremes to include the limiting degenerate cases. The points determining the division of the octave are

$$s_0 = 0; \quad s_i = \frac{i}{n} + t_i, \quad 0 \leq t_i \leq \frac{1}{\lambda n}, \quad 1 \leq i \leq n-1; \quad s_n = 1$$

By writing $T = \frac{1}{n}$ and assuming $t_0 = t_n = 0$, the respective intervals are

$$T_i = s_i - s_{i-1} = T + t_i - t_{i-1}, \; 1 \leq i \leq n$$

Then, the scale E_n generates the partition $\alpha = \{T_i\}$, $i \in \{1, \ldots, n\}$, with entropy

$$H(\alpha) = \sum_{i=1}^{n} z(T_i) = -\sum_{i=1}^{n} (T + t_i - t_{i-1}) \log_2(T + t_i - t_{i-1})$$

The entropy $H(\alpha)$ is a concave and differentiable function $f(t_1, \ldots, t_{n-1})$, $f{:}V \to \mathrm{R}$, with $V = [0, \frac{1}{\lambda n}]^{n-1}$, a hypercube, which is a convex and compact space. Then, f has a local and global maximum at $(t_1, \ldots, t_{n-1}) = (0, \ldots, 0)$, corresponding to an n-TET scale with $T_i = T$ and $H = \log_2 n$, and the global minimum is reached at one or some of the vertices of V. These vertices are determined using the possible values $t_j \in \{0, \frac{1}{\lambda n}\}$, $1 \leq j \leq n-1$, and the resulting intervals T_i can only take the following values: interval T_1, values $\frac{1}{n}$ and $\frac{\lambda+1}{\lambda n}$; interval T_n, values $\frac{1}{n}$ and $\frac{\lambda-1}{\lambda n}$; and the intermediate intervals, values $\frac{\lambda-1}{\lambda n}$, $\frac{1}{n}$ and $\frac{\lambda+1}{\lambda n}$.

The octave is covered by a number of different intervals satisfying

$$A + B + C = n \tag{20}$$

of which A in number have width $\frac{\lambda-1}{\lambda n}$, B width $\frac{1}{n}$, and C width $\frac{\lambda+1}{\lambda n}$, so that $A\,\frac{\lambda-1}{\lambda n} + B\,\frac{\lambda}{\lambda n} + C\,\frac{\lambda+1}{\lambda n} = 1$. Therefore,

$$A\,(\lambda - 1) + B\,\lambda + C\,(\lambda + 1) = \lambda n \tag{21}$$

(Some of these intervals may have width zero, giving rise to a degenerate scale with less than n non-null intervals.)

The entropy at one of these vertices can be written in terms of the respective number of intervals as

$$\begin{aligned} f &= A\,z(\tfrac{\lambda-1}{\lambda n}) + B\,z(\tfrac{\lambda}{\lambda n}) + C\,z(\tfrac{\lambda+1}{\lambda n}) = \\ &= \log_2 n + \tfrac{1}{n}\,g(A, C); \quad g(A, C) = A\,z(\tfrac{\lambda-1}{\lambda}) + C\,z(\tfrac{\lambda+1}{\lambda}) \end{aligned} \tag{22}$$

where z is the concave function defined in Equation (1), also extended for values $t > 1$, where $z(t) < 0$. Among possible configurations, we look for the minimum value that f can take, which will correspond to the minimum value of g. Notice that this quantity is added to the entropy of an n-TET scale, as it is consistent with the fact that the B intervals of width $\frac{1}{n}$ are not involved in the expression. However, the intervals in number C, of greater width than $\frac{1}{n}$, contribute to decreasing the entropy, while those in number A, of width lower than $\frac{1}{n}$, contribute to increasing it, since the corresponding function z evaluated in values less than 1 is positive.

In the current case, by combining Equations (20) and (21), we obtain $A = C$ and $B + 2C = n$. Therefore,

$$g(C) = C\,Z(\lambda, 1); \qquad Z(\lambda, a) = z(\tfrac{\lambda - a}{\lambda}) + z(\tfrac{\lambda + a}{\lambda}), \quad \lambda \geq a \tag{23}$$

If n is even, the maximum value for C, by assuming $B = 0$, is $C = \frac{n}{2}$. Hence, the number of intervals of null-size is also $A = \frac{n}{2}$.

The function $Z(\lambda, a) = z(\frac{\lambda-a}{\lambda}) + z(\frac{\lambda+a}{\lambda})$ with $a > 0$ is defined for $\lambda \geq a$, where it satisfies $Z(\lambda, a) < 0$, always increasing from $Z(a, a) = -2$ until $Z(\lambda, a) \to 0$ when $\lambda \to \infty$. Then, $g(\frac{n}{2}) = \frac{n}{2} Z(\lambda, 1)$, which correspond to a degenerate scale of $\frac{n}{2}$ non-null intervals. Therefore, if $n = \dot{2}$, the entropy satisfies $\log_2 n + \frac{1}{2} Z(\lambda, 1) \leq H(\alpha)$.

If n is odd, the maximum value for C and A is $\frac{n-1}{2}$, with $B = 1$. Then, $g = \frac{n-1}{2} Z(\lambda, 1)$, which correspond to a degenerate scale of $\frac{n+1}{2}$ non-null intervals. Therefore, if $n \neq \dot{2}$, the entropy satisfies $\log_2 n + \frac{1}{2} Z(\lambda, 1)(1 - \frac{1}{n}) \leq H(\alpha)$.

Case (b) is similar. It would be reasoned by considering that $f(t_1, \ldots, t_{n-1}) = f(-t_{n-1}, \ldots, -t_1)$, i.e., it is equivalent to case (a), but following the intervals from right to left starting at 1.

Theorem 1. *The entropy and bias of a scale* $\frac{1}{\lambda n}$*-OSS to n-TET satisfy*

$$\begin{aligned} n = \dot{2}, &\quad \log_2 n + \tfrac{1}{2} Z(\lambda, 1) \leq H(\alpha) \leq \log_2 n; &&\quad 0 \leq \theta(\alpha) \leq -\tfrac{1}{2\log_2 n} Z(\lambda, 1) \\ n \neq \dot{2}, &\quad \log_2 n + \tfrac{1}{2} Z(\lambda, 1)(1 - \tfrac{1}{n}) \leq H(\alpha) \leq \log_2 n; &&\quad 0 \leq \theta(\alpha) \leq -\tfrac{1}{2\log_2 n} Z(\lambda, 1)(1 - \tfrac{1}{n}) \end{aligned} \tag{24}$$

We explicitly write two cases. For $\lambda = 1$, since $Z(1, 1) = -2$, the entropy of an RID scale satisfies, if $n = \dot{2}$, $\log_2 \frac{n}{2} \leq H(\alpha) \leq \log_2 n$, and if $n \neq \dot{2}$, it satisfies $\log_2 \frac{n}{2} + \frac{1}{n} \leq H(\alpha) \leq \log_2 n$.

For $\lambda = 2$, since $Z(2, 1) = -0.189$, the entropy of a scale $\frac{1}{2n}$-OSS to n-TET satisfies, if $n = \dot{2}$, $\log_2 n - 0.189 \leq H(\alpha) \leq \log_2 n$ and, if $n \neq \dot{2}$, $\log_2 n - 0.189(1 - \frac{1}{n}) \leq H(\alpha) \leq \log_2 n$.

Levels for $\lambda = 1, 2, 4, 8, 16$ are displayed in Figure 4.

10. Entropy of a Scale $\frac{1}{\lambda n}$-Similar to *n*-TET

Let us calculate the minimum entropy that can reach an n-tone scale E_n, not necessarily cyclic, with a similarity level $\frac{1}{\lambda n}$ with the n-TET scale for $\lambda \geq 2$. According to the previous notation and considerations, the points that determine the division of the octave are

$$s_0 = 0; \quad s_i = \frac{i}{n} + t_i, \quad -\frac{1}{\lambda n} \leq t_i \leq \frac{1}{\lambda n}, \quad 1 \leq i \leq n - 1; \quad s_n = 1$$

The respective intervals are $T_i = s_i - s_{i-1} = \frac{1}{n} + t_i - t_{i-1}$, $1 \leq i \leq n$, by assuming $t_0 = t_n = 0$. As before, we write the entropy of the partition $\alpha = \{T_i\}$, $i \in \{1, \ldots, n\}$, as $H(\alpha) = \sum_{i=1}^{n} z(T_i) = f(t_1, \ldots, t_{n-1})$, where $f{:}V \to \mathrm{R}$, with $V = [-\frac{1}{\lambda n}, \frac{1}{\lambda n}]^{n-1}$, a convex and compact space. The local and global maximum of f takes place at $(t_1, \ldots, t_{n-1}) = (0, \ldots, 0)$, corresponding to the scale of n-TET, with $T_i = T$ and $H = \log_2 n$. The global minimum is reached at one or some of the vertices of V. These vertices are determined using the possible values $t_j \in \{-\frac{1}{\lambda n}, \frac{1}{\lambda n}\}$, $1 \leq j \leq n - 1$, and the resulting intervals T_i can only take the following values: the values $\frac{\lambda-1}{\lambda n}$ and $\frac{\lambda+1}{\lambda n}$ of the extreme intervals T_1 and T_n, while the intermediate intervals can have values $\frac{\lambda-2}{\lambda n}$, $\frac{1}{n}$, and $\frac{\lambda+2}{\lambda n}$. The octave is covered by a number of different intervals satisfying

$$A + B + C + D + E = n \tag{25}$$

of which A in number have width $\frac{\lambda-2}{\lambda n}$, B width $\frac{\lambda-1}{\lambda n}$, C width $\frac{1}{n}$, D width $\frac{\lambda+1}{\lambda n}$, and E width $\frac{\lambda+2}{\lambda n}$, so that $A\,\frac{\lambda-2}{\lambda n} + B\,\frac{\lambda-1}{\lambda n} + C\,\frac{\lambda}{\lambda n} + D\,\frac{\lambda+1}{\lambda n} + E\,\frac{\lambda+2}{\lambda n} = 1$. Therefore,

$$A\,(\lambda-2) + B\,(\lambda-1) + C\,\lambda + D\,(\lambda+1) + E\,(\lambda+2) = \lambda n \tag{26}$$

The entropy at one of these vertices can be written in terms of the respective number of intervals:

$$\begin{aligned} f &= A\,z(\tfrac{\lambda-2}{\lambda n}) + B\,z(\tfrac{\lambda-1}{\lambda n}) + C\,z(\tfrac{\lambda}{\lambda n}) + D\,z(\tfrac{\lambda+1}{\lambda n}) + E\,z(\tfrac{\lambda+2}{\lambda n}) = \\ &= \log_2 n + \tfrac{1}{n}\,g(A,B,D,E), \quad g(A,B,D,E) = A\,z(\tfrac{\lambda-2}{\lambda}) + B\,z(\tfrac{\lambda-1}{\lambda}) + D\,z(\tfrac{\lambda+1}{\lambda}) + E\,z(\tfrac{\lambda+2}{\lambda}) \end{aligned} \tag{27}$$

Once again, the intervals of width $\frac{1}{n}$ are not involved in this expression. The intervals in numbers D and E, of greater width than $\frac{1}{n}$, contribute to decreasing the entropy, while those who are there in numbers A and B, of width lower than $\frac{1}{n}$, contribute to increasing it.

The minimum value of g is obtained using the highest possible values of E and D (in that order) and minimum of A and B.

In the current case, by combining Equations (25) and (26), we obtain $A = E + \frac{1}{2}D - \frac{1}{2}B$ and

$$B + 2C + 3D + 4E = 2n \tag{28}$$

If n is even, at most there can be $E = \frac{n}{2}$ intermediate intervals of width $\frac{\lambda+2}{\lambda n}$. Through the substitution of this value into Equation (28), we obtain $B + 2C + 3D = 0$, so that B=C=D=0 and $A = \frac{n}{2}$. Hence, the minimum g is $g(\frac{n}{2}, 0, 0, \frac{n}{2}) = \frac{n}{2}Z(\lambda, 2)$, corresponding to a degenerate scale of $\frac{n}{2}$ non-null intervals. Therefore, if $n \doteq 2$, the entropy satisfies $\log_2 n + \frac{1}{2}Z(\lambda, 2) \le H(\alpha)$.

If n is odd, we should examine two alternatives. Firstly, if we assume that there are $E = \frac{n-1}{2}$ intermediate intervals of width $\frac{\lambda+2}{\lambda n}$, through substitution into Equation (28), we obtain $B + 2C + 3D - 2 = 0$. Possible interval values are $(B, C, D) = (0, 1, 0); (2, 0, 0)$. In this case, the minimum value of g is provided by the first triad, i.e., B=D=0, $C = 1$, $A = \frac{n-1}{2}$, which yields $g(\frac{n-1}{2}, 0, 0, \frac{n-1}{2}) = \frac{n-1}{2}Z(\lambda, 2)$, corresponding to a degenerate scale of $\frac{n+1}{2}$ non-null intervals. In this case,

$$\log_2 n + \tfrac{1}{2}Z(\lambda, 2)(1 - \tfrac{1}{n}) \le H(\alpha) \tag{29}$$

As the second alternative, if there are $E = \frac{n-3}{2}$ intermediate intervals of width $\frac{\lambda+2}{\lambda n}$, $B + 2C + 3D - 6 = 0$ holds. Possible interval values are $(B, C, D) = (0, 0, 2); (1, 1, 1); (3, 0, 1); (0, 3, 0); (4, 1, 0)$. The minimum of g is provided by the first triad, so that B=C=0, $D = 2$, $A = \frac{n-1}{2}$, and $g(\frac{n-1}{2}, 0, 2, \frac{n-3}{2}) = \frac{n-1}{2}z(\frac{\lambda-2}{\lambda}) + 2z(\frac{\lambda+1}{\lambda}) + \frac{n-3}{2}z(\frac{\lambda+2}{\lambda})$, corresponding to a degenerate scale of $\frac{n+1}{2}$ non-null intervals. Hence,

$$\log_2 n + \tfrac{1}{2}Z(\lambda, 2)\left[1 - \tfrac{1}{n}\,\frac{z(\frac{\lambda-2}{\lambda}) - 4z(\frac{\lambda+1}{\lambda}) + 3z(\frac{\lambda+2}{\lambda})}{Z(\lambda, 2)}\right] \le H(\alpha) \tag{30}$$

For $\lambda \ge 2$, the factor multiplying $\frac{1}{n}$ is a positive value between $3(2 - \log_2 3) = 1.245$ and $\frac{3}{2}$. Since the value $Z(\lambda, 2)$ is negative, the above factor increases it. Hence, the entropy of Equation (29) is lower than the entropy of Equation (30).

Theorem 2. *The entropy and bias of a scale* $\frac{1}{\lambda n}$*-similar to n-TET satisfy*

$$n = \dot{2}, \quad \log_2 n + \tfrac{1}{2} Z(\lambda, 2) \le H(\alpha) \le \log_2 n; \qquad 0 \le \theta(\alpha) \le -\tfrac{1}{2\log_2 n} Z(\lambda, 2)$$
$$n \ne \dot{2}, \quad \log_2 n + \tfrac{1}{2} Z(\lambda, 2)(1 - \tfrac{1}{n}) \le H(\alpha) \le \log_2 n; \quad 0 \le \theta(\alpha) \le -\tfrac{1}{2\log_2 n} Z(\lambda, 2)(1 - \tfrac{1}{n}) \tag{31}$$

We explicitly evaluate the case for $\lambda = 2$. Since $Z(2,2) = -2$, if $n = \dot{2}$, then $\log_2 \frac{n}{2} \le H(\alpha) \le \log_2 n$, and if $n \ne \dot{2}$, then $\log_2 \frac{n}{2} + \frac{1}{n} \le H(\alpha) \le \log_2 n$, which are the same bounds as for an RID scale.

11. Cyclic RID Scales

We may reformulate Theorem A1 according to the above definitions.

Theorem 3. *Optimal cyclic scales are RID scales.*

Nevertheless, there are many non-optimal cyclic scales that are also RID scales. Cyclic RID scales correspond to partial convergents of continued fractions where the comma is not as low as for optimal scales, i.e., their closure is not a best approximation, although they belong to a family of relatively good partial convergents. However, although in most cases, a convergent that gives a good approximation (i.e., an accurate scale) generates a RID scale, there are exceptions, such as for $n = 200$. Between the notes $j = 83$ of the respective scales E^3_{200} and $E^{\top}_{200}$, there is a distance of nearly 1.5 elementary intervals of the equal temperament scale, and between two consecutive notes of the cyclic scale, there may be a distance equivalent to 2.1 regular intervals, which means that within some regular intervals, there are two notes of the cyclic scale.

Therefore, there exist accurate scales that are not RID, and RID scales that are not accurate.

11.1. Comma and Elementary Intervals of an RID Scale

It is immediate to identify an RID scale from its interval comma $|\phi_n| = |\log_2 \gamma_n| = |n \log_2 h - N|$. As seen in Appendix B, the distance between the tones $\vartheta'_j \in E^{\top}_n$ and $\vartheta_j = \nu_k \in E^h_n$ for $0 \le j < n$ is $\frac{k}{n}|\phi_n|$. So, $\frac{k}{n}|\phi_n| < \frac{1}{n}$ for all $0 \le k < n$ if and only if $\frac{n-1}{n}|\phi_n| < \frac{1}{n}$, that is, the interval comma is limited as $|\phi_n| < \frac{1}{n-1}$.

Lemma 4. *A cyclic scale is RID if and only if $|\phi_n| < \frac{1}{n-1}$.*

This condition, which is satisfied by all optimal scales, assures us that the value of $|\phi_n|$ is small enough to be far from the previously seen degenerate cases of less than n notes, since for every RID scale, $|\phi_n| < \frac{1}{n-1} \le \min(\frac{1}{m}, \frac{1}{M})$ is fulfilled.

From another point of view, while considering RID scales, we are excluding "bad approximations" of $\log_2 h$ that satisfy any of the following: if $\phi_n > 0$, then $\frac{1}{n-1} \le n \log_2 h - N < \frac{1}{M}$, or, if $\phi_n < 0$, then $-\frac{1}{m} < N - n \log_2 h \le -\frac{1}{n-1}$. This fact also has implications on the bounds of the two elementary intervals of cyclic RID scales, We distinguish the following cases:

(i) If $0 < \phi_n < \frac{1}{n-1}$, according to Equation (9), $u < \frac{1}{n}(1 + \frac{m}{n-1}) < \frac{1}{n}(1 + \frac{n-1}{n-1}) = \frac{2}{n}$. Hence,
$$\tfrac{1}{n} < u < \min(\tfrac{1}{M}, \tfrac{2}{n}) \tag{32}$$
(We should distinguish two cases: (a) $M > m$: $\frac{1}{n} < \frac{1}{M} < \frac{2}{n}$ and $\frac{2}{n} < \frac{1}{m}$; since $Mu + md = 1$, we have $\frac{1}{n} < u < \frac{1}{M}$ and $0 < d < \frac{1}{n}$, which does not add any new limitation. (b) $M < m$: $\frac{1}{n} < u < \frac{2}{n}$ and $\frac{1}{n}(1 - \frac{M}{m}) < d < \frac{1}{n}$. Therefore, the extremes still correspond to n-tone scales, by avoiding degenerate cases.)

(ii) If $-\frac{1}{n-1} < \phi_n < 0$, according to Equation (9), $d < \frac{1}{n}(1 + \frac{M}{n-1}) < \frac{1}{n}(1 + \frac{n-1}{n-1}) = \frac{2}{n}$. Hence,
$$\tfrac{1}{n} < d < \min(\tfrac{1}{m}, \tfrac{2}{n}) \tag{33}$$

(We should distinguish two cases: (a) $M > m$: $\frac{1}{n} < \frac{1}{M} < \frac{2}{n}$ and $\frac{2}{n} < \frac{1}{m}$; hence, $\frac{1}{n} < d < \frac{2}{n}$ and $\frac{1}{n}(1 - \frac{m}{M}) < u < \frac{1}{n}$. Once again, the extremes correspond to n-tone scales not degenerating toward scales of fewer tones. (b) $M < m$: $\frac{2}{n} < \frac{1}{M}$ and $\frac{1}{n} < \frac{1}{m} < \frac{2}{n}$; hence, $\frac{1}{n} < d < \frac{1}{m}$ and $0 < u < \frac{1}{n}$, which does not add any new limitation.)

In both cases, the size of the greatest elementary interval of a cyclic RID scale is lower than $\frac{2}{n}$, as stated in Lemma 3.

11.2. Bias of RID Scales

For large values of $|\phi_n|$, that is, $|\phi_n| > \frac{1}{m}$ or $|\phi_n| > \frac{1}{M}$, Equation (14) can have a quite arbitrary and non-symmetrical behavior, but for values $|\phi_n| < \frac{1}{n-1}$, and in particular, for optimal scales with $|\phi_n| < \frac{1}{n+M}$ or $|\phi_n| < \frac{1}{n+m}$, depending on the value δ, the bias θ_n behaves as proportional to ϕ_n^2.

With $|m\phi_n| < 1$ and $|M\phi_n| < 1$, we can approximate the logarithms in the following expressions as

$$(1 + m\phi_n)\ln(1 + m\phi_n) = m\phi_n + \tfrac{1}{2}m^2\phi_n^2 - \tfrac{1}{6}m^3\phi_n^3 + O_4(m\phi_n)$$

$$(1 - M\phi_n)\ln(1 - M\phi_n) = -M\phi_n + \tfrac{1}{2}M^2\phi_n^2 + \tfrac{1}{6}M^3\phi_n^3 + O_4(M\phi_n)$$

and through substitution into Equation (14), we obtain

$$\theta_n = \frac{1}{\ln n}\left[\frac{Mm^2+mM^2}{2n}\phi_n^2 - \frac{Mm^3-mM^3}{6n}\phi_n^3 + \frac{Mm^4+mM^4}{12n}\,O_4(\phi_n)\right] =$$

$$= \frac{mM}{2\ln n}\left[\phi_n^2 - \frac{m-M}{3}\phi_n^3 + \frac{m^2-mM+M^2}{6}\,O_4(\phi_n)\right]$$

Notice that $m|\phi_n|^3 = \frac{1}{m^2}|m\phi_n|^3 = O_3(|\phi_n|) < O_3(|m\phi_n|)$ with $|m\phi_n| < 1$ and $M|\phi_n|^3 = \frac{1}{M^2}|M\phi_n|^3 = O_3(|\phi_n|) < O_3(|M\phi_n|)$ with $|M\phi_n| < 1$. The same happens with the higher-order terms. Therefore, for enough small $|\phi_n|$, we can use the approximation derived from the following result.

Lemma 5. *The bias of a cyclic RID scale satisfies*

$$\theta_n = \frac{mM}{2\ln(m+M)}\,\phi_n^2 + O_3(|\phi_n|) \tag{34}$$

11.3. Cyclic Scales of Minimal Bias

As explained in Section 4, while refining a cyclic scale, in each iteration, one of the indices of the extreme tones remains fixed, while the other increases by a value equal to the one that remains fixed. Let us see that in each refinement, the following function appearing in Equation (34) always increases:

$$\psi(m, M) = \frac{mM}{\ln(m+M)} \tag{35}$$

Lemma 6. *For $n = m + M \geq 2$, $\psi(m^+, M) > \psi(m, M)$ with $m^+ = m + M$, and $\psi(m, M^+) > \psi(m, M)$ with $M^+ = m + M$.*

Proof. Indeed, it suffices to check that the function $\psi(m, M) = \frac{mM}{\ln(m+M)}$ satisfies $\frac{\partial\psi}{\partial m} > 0$ and $\frac{\partial\psi}{\partial M} > 0$.

$$\frac{\partial}{\partial m}\frac{mM}{\ln(m+M)} = \frac{M\ln(m+M) - \frac{mM}{m+M}}{\ln^2(m+M)} = \frac{M^2\ln(m+M) + mM(\ln(m+M)-1)}{\ln^2(m+M)} > 0 \iff m + M \geq 2$$

$$\frac{\partial}{\partial M}\frac{mM}{\ln(m+M)} = \frac{m\ln(m+M) - \frac{mM}{m+M}}{\ln^2(m+M)} = \frac{m^2\ln(m+M) + mM(\ln(m+M)-1)}{\ln^2(m+M)} > 0 \iff m+M \geq 2$$

Notice that for $m + M > 2$, the above inequalities hold, and if $m + M = 2$, then $M = m = 1$, so that the above results are also valid. □

Corollary 1. *Let us write $\psi_n = \psi(m, M)$. If $n > n'$, then $\psi_n > \psi_{n'}$.*

With this notation, the bias of a cyclic RID scale can be estimated from

$$\theta_n = \frac{1}{2}\psi_n\phi_n^2 \tag{36}$$

Theorem 4. *Every cyclic RID scale of minimal bias is optimal.*

Proof. We prove this by denying the consequent. Assume two cyclic RID scales E_n^h and $E_{n'}^h$ such that $n > n'$ with interval commas satisfying $|\phi_n| > |\phi_{n'}|$, i.e., E_n^h is not optimal. Then, applying the previous corollary to Equation (36), we have

$$\psi_n\phi_n^2 > \psi_{n'}\phi_{n'}^2 \Rightarrow \theta_n > \theta_{n'}$$

Therefore, if a cyclic RID scale is not optimal, it cannot be MB. □

The first cyclic MB scales for $h = 3$, i.e., Pythagorean scales, are for $n = 5; 12; 53; 665\ldots$, as shown in Figure 4, as well as in Table 1.

12. Conclusions

In the current paper, it was proposed to measure the regularity of the intervals of a music scale from their partition entropy. Among other properties, the fact of being a continuous increasing function of n for an n-TET scale, which is always the maximum value that the entropy of any n-tone scale can reach, together with the sub-additivity property, which guarantees that while refining the partition, the entropy always increases, make this parameter very suitable for our purpose. In order to compare scales with different numbers of tones, the entropy relative to the corresponding regular scale is used, which is the normalized entropy, so that their complementary to 1 quantifies the bias relative to the n-TET scale.

The main application of these concepts has been to cyclic scales, and their properties were reviewed and further investigated in the Appendices. Since non-degenerate cyclic scales are maximal even sets [3,4] and their intervals come in two possible sizes [1,2], the remaining properties allowing us to distinguish between scale distributions are the number of intervals of each size and the ratio between them.

Two situations have been analyzed. First, cyclic scales with a fixed number of tones, which is a family of modulating temperament scales around one generator. In this case, the bias only depends on the closure, i.e., the relative size of both elementary intervals. Second, as cyclic scales are refined, the bias also depends on how many intervals of each size there are.

In order to study such a dependency, it was necessary to restrict the scales in two ways. We centered our attention to scales with a lower limit of the entropy, determined using the condition that their notes are distributed along each of the intervals of a regular scale (RID scales). Such a study was conducted in a general way, by calculating the maximum bias of several scales, not necessarily cyclic, with different levels of similarity with an n-TET scale, either from one side or from both sides. Figure 4 displays the similarity levels for cyclic scales. In addition, we considered scales with the comma not exceeding that of an RID scale, since in this case, the dependency between bias and closure is well defined.

We proved that any cyclic scale of minimal bias (MB scale), i.e., with a bias that is lower than that of the cyclic scales of fewer tones, is necessarily optimal, i.e., corresponds to a best rational approximation of $\log_2 h$.

Notice that the bias, θ_n, according to Equations (35) and (36), is proportional to the product of $\psi(m, M)$, depending on each number of elementary intervals, and ϕ_n^2, depending on their relative size, so that θ_n accounts for the degrees of freedom allowed for cyclic scales.

Among the optimal scales, it was possible to select the ones for which intervals are distributed along the octave as regularly as possible relative to an equal temperament scale of the same number of tones. Therefore, in relation to the closure, scales can be ordered in nested families, from worst to best, as cyclic, accurate, and optimal scales, whilst in relation to their regularity, they can be ordered in nested families as cyclic, RID, optimal, and MB scales.

Although the current entropy-based measure has been particularly used to deepen the study of cyclic scales, the present work clearly suggests future applications to more general cases, also through using alternative metrics, either based on the distribution of the scale notes or the intervals, such as the recent Boltzmann–Shannon Interaction Entropy [32], allowing us to estimate a normalized entropy from a finite sample of points on a bounded interval.

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BRA	Best rational approximation (optimal scale);
GRA	Good rational approximation (accurate scale);
I|Xn	Maximum distance $1/x_n$ to n-TET;
MB	Minimal bias;
n-TET	n-tones of equal temperament;
OSS	One-side similar;
RID	Regular interval distributed.

Appendix A. Bounds for the Elementary Intervals and Closure of Cyclic Scales

The notes of a cyclic scale E_n^h divide the circle of the octave into n intervals of width $\log_2 U$ or $\log_2 D$, beginning and ending in the pitch class of the fundamental so that

$$M \log_2 U + m \log_2 D = 1 \tag{A1}$$

According to the definitions of U, D, and γ_n, these intervals can be referred to the elementary interval of the equal temperament scale as follows:

$$\log_2 U = \tfrac{1}{n} + \tfrac{m}{n} \log_2 \gamma_n; \quad \log_2 D = \tfrac{1}{n} - \tfrac{M}{n} \log_2 \gamma_n \tag{A2}$$

We may determine their bounds by distinguishing between the cases of γ_n being greater or less than 1:

(i) If $\gamma_n > 1$, then $U > D$. Depending on the scale, we have the following:

(a) The closure of an *optimal cyclic scale* satisfies $1 < \gamma_n < \frac{2}{\nu_M} < \nu_m$; hence, $1 < \gamma_n < D < U$. By taking logarithms and taking into account Equation (A2), we obtain

$$0 < \log_2 \gamma_n < \log_2 D = \tfrac{1}{n} - \tfrac{M}{n} \log_2 \gamma_n \quad \Longrightarrow \quad 0 < \log_2 \gamma_n < \tfrac{1}{n+M} \tag{A3}$$

The elementary intervals satisfy

$$\tfrac{1}{n+M} < \log_2 D < \tfrac{1}{n}; \quad \tfrac{1}{n} < \log_2 U < \tfrac{2}{n+M}$$

(b) The closure of a *non-optimal cyclic scale* satisfies $1 < \frac{2}{\nu_M} < \gamma_n < \nu_m$; hence, $1 < D < \gamma_n < U$. Then, in taking logarithms,

$$\begin{array}{rcl} \log_2 D = \frac{1}{n} - \frac{M}{n}\log_2 \gamma_n < \log_2 \gamma_n & \Longrightarrow & \frac{1}{n+M} < \log_2 \gamma_n \\ \log_2 \gamma_n < \log_2 U = \frac{1}{n} + \frac{m}{n}\log_2 \gamma_n & \Longrightarrow & \log_2 \gamma_n < \frac{1}{M} \end{array} \tag{A4}$$

Therefore,

$$0 < \log_2 D < \frac{1}{n+M}; \quad \frac{2}{n+M} < \log_2 U < \frac{1}{M}$$

(c) For an *accurate cyclic scale* for the above condition, $0 < \log_2 D < \log_2 \gamma_n < \log_2 U$, we must add that of a *good rational approximation*, which can be written as

$$0 < \frac{1}{n}\log_2 \gamma_n < \frac{1}{M}\log_2 D < \frac{1}{m}\log_2 U$$

Then, $\frac{1}{n}\log_2 \gamma_n < \frac{1-M\log_2 \gamma_n}{Mn}$; hence, $\log_2 \gamma_n < \frac{1}{2M}$, and, since Equation (A4) is still valid, we obtain the following bounds:

$$\frac{1}{n+M} < \log_2 \gamma_n < \frac{1}{2M}; \quad \frac{1}{2n} < \log_2 D < \frac{1}{n+M}; \quad \frac{2}{n+M} < \log_2 U < \frac{1}{2n} + \frac{1}{2M}$$

Thus, in case (i), either for optimal or non-optimal scales, the following is satisfied:

$$\gamma_n > 1 \Longrightarrow \quad 0 < \log_2 \gamma_n < \frac{1}{M}; \quad 0 < \log_2 D < \frac{1}{n}; \quad \frac{1}{n} < \log_2 U < \frac{1}{M} \tag{A5}$$

(ii) If $\gamma_n < 1$, then $D > U$. Depending on the scale, we have the following:

(a) The closure of an *optimal cyclic scale* satisfies $\frac{\nu_M}{2} < \frac{1}{\nu_m} < \gamma_n < 1$; hence, $\frac{1}{D} < \frac{1}{U} < \gamma_n < 1$. Then, in taking logarithms,

$$-\log_2 U = -\frac{1}{n} - \frac{m}{n}\log_2 \gamma_n < \log_2 \gamma_n < 0 \quad \Longrightarrow \quad -\frac{1}{n+m} < \log_2 \gamma_n < 0 \tag{A6}$$

The elementary intervals satisfy

$$\frac{1}{n+m} < \log_2 U < \frac{1}{n}; \quad \frac{1}{n} < \log_2 D < \frac{2}{n+m}$$

(b) The closure of a *non-optimal cyclic scale* satisfies $\frac{\nu_M}{2} < \gamma_n < \frac{1}{\nu_m} < 1$; hence, $\frac{1}{D} < \gamma_n < \frac{1}{U} < 1$. Then, in taking logarithms,

$$\begin{array}{rcl} -\log_2 D = -\frac{1}{n} + \frac{M}{n}\log_2 \gamma_n < \log_2 \gamma_n & \Longrightarrow & -\frac{1}{m} < \log_2 \gamma_n < 0 \\ \log_2 \gamma_n < -\log_2 U = -\frac{1}{n} - \frac{m}{n}\log_2 \gamma_n & \Longrightarrow & \log_2 \gamma_n < -\frac{1}{n+m} \end{array} \tag{A7}$$

Then, the elementary intervals satisfy

$$0 < \log_2 U < \frac{1}{n+m}; \quad \frac{2}{n+m} < \log_2 D < \frac{1}{m}$$

(c) For an *accurate cyclic scale* for the above condition, $-\log_2 D < \log_2 \gamma_n < -\log_2 U < 0$, we must add that of a *good rational approximation*, which can be written as

$$-\frac{1}{M}\log_2 D < -\frac{1}{m}\log_2 U < \frac{1}{n}\log_2 \gamma_n < 0$$

Then, $\frac{1}{n}\log_2 \gamma_n > -\frac{1+m\log_2 \gamma_n}{mn}$; hence, $-\frac{1}{2m} < \log_2 \gamma_n$, and, since Equation (A7) is still valid, we obtain the following bounds:

$$-\frac{1}{2m} < \log_2 \gamma_n < -\frac{1}{n+m}; \quad \frac{2}{n+m} < \log_2 D < \frac{1}{2n} + \frac{1}{2m}; \quad \frac{1}{2n} < \log_2 U < \frac{1}{n+m}$$

Thus, in case (ii), either for optimal or non-optimal scales, the following is satisfied:

$$\gamma_n < 1 \Longrightarrow \quad -\frac{1}{m} < \log_2 \gamma_n < 0; \quad 0 < \log_2 U < \frac{1}{n}; \quad \frac{1}{n} < \log_2 D < \frac{1}{m} \tag{A8}$$

Table A1 summarizes the above results over the octave S_0. Notice that, for optimal scales, Equations (A3) and (A6) give the following bounds, similarly to the theory of continued fractions:

$$\text{(i)} \quad \left|\log_2 h - \frac{N}{n}\right| < \frac{1}{n(n+M)} \;; \quad \text{(ii)} \quad \left|\log_2 h - \frac{N}{n}\right| < \frac{1}{n(n+m)}$$

Table A1. Bounds for $\phi_n = \log_2 \gamma_n$, $u = \log_2 U$, and $d = \log_2 D$, depending on the type of cyclic scale.

Cyclic Scale	$\phi_n > 0$	$\phi_n < 0$
optimal	$0 < \phi_n < \frac{1}{n+M}$ $\frac{1}{n} < u < \frac{2}{n+M}$ $\frac{1}{n+M} < d < \frac{1}{n}$	$-\frac{1}{n+m} < \phi_n < 0$ $\frac{1}{n+m} < u < \frac{1}{n}$ $\frac{1}{n} < d < \frac{2}{n+m}$
non-optimal	$\frac{1}{n+M} < \phi_n < \frac{1}{M}$ $\frac{2}{n+M} < u < \frac{1}{M}$ $0 < d < \frac{1}{n+M}$	$-\frac{1}{m} < \phi_n < -\frac{1}{n+m}$ $0 < u < \frac{1}{n+m}$ $\frac{2}{n+m} < d < \frac{1}{m}$
accurate	$\frac{1}{n+M} < \phi_n < \frac{1}{2M}$ $\frac{2}{n+M} < u < \frac{1}{2n} + \frac{1}{2M}$ $\frac{1}{2n} < d < \frac{1}{n+M}$	$-\frac{1}{2m} < \phi_n < -\frac{1}{n+m}$ $\frac{1}{2n} < u < \frac{1}{n+m}$ $\frac{2}{n+m} < d < \frac{1}{2n} + \frac{1}{2m}$
in general	$0 < \phi_n < \frac{1}{M}$ $\frac{1}{n} < u < \frac{1}{M}$ $0 < d < \frac{1}{n}$	$-\frac{1}{m} < \phi_n < 0$ $1 < u < \frac{1}{n}$ $\frac{1}{n} < d < \frac{1}{m}$

Appendix B. Deviation of a Cyclic Scale from *n*-TET

We calculate the interval between the notes with ordinal j of the respective scales E_n^h and $E_n^\top$, corresponding to the tones $\vartheta_j \in E_n^h$ and $\vartheta'_j = 2^{\frac{j}{n}} \in E_n^\top$. According to Equations (A2) and (A1), the deviations of each note of E_n^h with regard to $E_n^\top$ compensate each other, and in the end, they close the octave exactly. If the tone ϑ_j corresponds to the iteration ν_k, then $j = Nk - n[\![k]\!]$ (Equation (28) [17]). Therefore, $[\![k]\!] = \frac{Nk-j}{n}$. In bearing in mind that $\log_2 \gamma_n = n \log_2 h - N$, the scale note $\log_2 \vartheta_j = \log_2 \nu_k$ is

$$\log_2 \vartheta_j = k \log_2 h - [\![k]\!] = \frac{k}{n} \log_2 \gamma_n + \frac{j}{n} \tag{A9}$$

Lemma A1. *The interval that separates the two notes is*

$$I(\vartheta_j, \vartheta'_j) = \log_2 \vartheta_j - \log_2 \vartheta'_j = \frac{k}{n} \log_2 \gamma_n; \quad j = 0, \ldots, n-1 \tag{A10}$$

Therefore, all of these intervals have the same sign, which is positive if $\gamma_n > 1$ and negative if $\gamma_n < 1$. In taking absolute value, the interval between these close notes becomes a distance:

$$d(\vartheta_j, \vartheta'_j) = \frac{k}{n} \left|\log_2 \gamma_n\right| = \frac{k}{n} \kappa_n; \quad j = 0, \ldots, n-1 \tag{A11}$$

As expected, this distances increases with the iterations, so that the maximum is reached by the last iterate, $k = n-1$, i.e., $d_{\max} = \max_j d(\vartheta_j, \vartheta'_j) = \frac{n-1}{n} \kappa_n$. Then, a condition such as $d_{\max} \leq \frac{1}{\lambda n}$, for $\lambda > 0$, becomes $\kappa_n \leq \frac{1}{\lambda (n-1)}$, which in cents is

$$¢(\kappa_n) \leq \frac{1200}{\lambda (n-1)} \tag{A12}$$

For an optimal cyclic scale, according to Equations (A3) and (A6), $-\frac{1}{n+m} < \log_2 \gamma_n < \frac{1}{n+M}$, so that $|\log_2 \gamma_n| < \frac{1}{n}$ and $k|\log_2 \gamma_n| < 1$ in Equation (A11). Therefore, we conclude that $d(\vartheta_j, \vartheta'_j) < \frac{1}{n}$.

Theorem A1. *The notes ϑ_i, $i \neq 0$, of an* optimal *cyclic scale E_n^h and the ones of the corresponding n-TET scale $E_n^\top$ alternate, i.e., for $0 < i < n$, either for* (i) $\vartheta'_i < \vartheta_i < \vartheta'_{i+1}$ *or* (ii) $\vartheta'_{i-1} < \vartheta_i < \vartheta'_i$.

Such a situation is also possible for non-optimal scales, although there are exceptions where the tones do not alternate. For instance, in a non-optimal scale with $\gamma_n < 1$, Equation (A7) implies that $m|\log_2 \gamma_n| < 1$. But in a cyclic scale where $M > m$, the condition $k|\log_2 \gamma_n| < 1$, $\forall k$, stated in case (ii), is not fulfilled. An example of such a situation is the 29-tone scale, with $m = 12$ and $M = 17$, where $|\log_2 \gamma_n|^{-1} = 27.7$. The 28-th tone $\nu_{28} = \vartheta_{12}$ (454.74 cents) is below ϑ'_{12} (496.55 cents) more than a factor of $2^{\frac{1}{29}}$ (41.38 cents), and between ϑ_{12} and $\vartheta_{13} = \nu_{11}$, there are two consecutive tones, ϑ'_{11} and ϑ'_{12}, of the 29-TET scale.

References

1. Clough, J.; Myerson, G. Variety and multiplicity in diatonic systems. *J. Music. Theory* **1985**, *29*, 249–270.
2. Clough, J.; Myerson, G. Musical scales and the generalized circle of fifths. *Am. Math. Mon.* **1986**, *93*, 695–701.
3. Clough, J.; Douthett, J. Maximally even sets. *J. Music. Theory* **1991**, *35*, 93–173.
4. Douthett, J.; Krantz, R. Maximally even sets and configurations: Common threads in mathematics, physics, and music. *J. Comb. Optim.* **2007**, *14*, 385–410.
5. Pinkerton, R. Information theory and melody. *Sci. Am.* **1956**, *194*, 77–87.
6. Youngblood, J. Style as information. *J. Music. Theory* **1958**, *2*, 24–35.
7. Lewin, D. Some applications of communication theory to the study of twelve-tone music. *J. Music. Theory* **1968**, *12*, 50–84.
8. Cohen, J. Information Theory and Music. *Behav. Sci.* **1962**, *7*, 137–163.
9. Knopoff, L.; Hutchinson, W. Information theory for musical continua. *J. Music. Theory* **1981**, *25*, 17–44.
10. Knopoff, L.; Hutchinson, W. Entropy as a measure of style: The influence of sample length. *J. Music. Theory* **1983**, *27*, 75–79.
11. Snyder, J. Entropy as a Measure of Musical Style: The Influence of a Priori Assumptions. *Music. Theory Spectr.* **1990**, *12*, 121–160.
12. Temperley, D. *Music and Probability*; The MIT Press: Cambridge, MA, USA, 2006. https://doi.org/10.7551/mitpress/4807.001.0001.
13. Sakellariou, J.; Tria, F.; Loreto, V.; Pachet, F. Maximum entropy models capture melodic styles. *Sci. Rep.* **2017**, *7*, 9172.
14. Gündüz, G. Entropy, energy, and instability in music. *Phys. A Stat. Mech. Its Appl.* **2023**, *609*, 128365.
15. Arnold, V.; Avez, A. *Ergodic Problems of Classical Mechanics*; Mathematical Physics Monograph Series; Benjamin: New York, NY, USA, 1968.
16. Cubarsi, R. An alternative approach to generalized Pythagorean scales. Generation and properties derived in the frequency domain. *J. Math. Music.* **2020**, *14*, 266–291. https://doi.org/10.1080/17459737.2020.1726690.
17. Cubarsi, R. On the divisions of the octave in generalized Pythagorean scales and their bidimensional representation. *J. Math. Music.* **2024**, *18*, 1–19. https://doi.org/10.1080/17459737.2023.2194301.
18. Carey, N.; Clampitt, D. Aspects of Well-Formed Scales. *Music. Theory Spectr.* **1989**, *11*, 187–206.
19. Carey, N.; Clampitt, D. Two Theorems Concerning Rational Approximations. *J. Math. Music.* **2012**, *6*, 61–66.
20. Carey, N.; Clampitt, D. Addendum to "Two Theorems Concerning Rational Approximations". *J. Math. Music.* **2017**, *11*, 61–63.
21. Žabka, M. Generalized Tonnetz and well-formed GTS: A scale theory inspired by the Neo-Riemannians. In *Mathematics and Computation in Music*; Springer: Berlin, Germany, 2009; pp. 286–298.
22. Žabka, M. Well-Formedness in Two Dimensions: A Generalization of Carey and Clampitt's Theorem. *J. Math. Music.* **2010**, *4*, 1–30.
23. Hellegouarch, Y. Gammes naturelles I. *Gaz. Des Mat.* **1999**, *81*, 25–39.
24. Hellegouarch, Y. Gammes naturelles II. *Gaz. Des Mat.* **1999**, *82*, 13–26.
25. Kassel, A.; Kassel, C. On Hellegouarch's definition of musical scales. *J. Math. Music.* **2010**, *4*, 31–43.
26. Loya, P. *Amazing and Aesthetic Aspects of Analysis*; Springer: New York, NY, USA, 2017.
27. Khinchin, A. *Continued Fractions*; University of Chicago: Chicago, IL, USA, 1964.
28. Milne, A.; Sethares, W.; Plamondon, J. Tuning continua and keyboard layouts. *J. Math. Music.* **2008**, *2*, 1–19.
29. Sethares, W.; Milne, A.; Tiedje, S.; Prechtl, A.; Plamondon, J. Spectral Tools for Dynamic Tonality and Audio Morphing. *Comput. Music. J.* **2009**, *33*, 71–84.
30. Wilcox, A. *Indices of Qualitative Variation*; Oak Ridge National Lab.: Oak Ridge, TN, USA, 1967.

31. Liern, V. On the construction, comparison, and exchangeability of tuning systems. *J. Math. Music.* **2015**, *9*, 197–213.
32. Diggans, C.T.; AlMomani, A. Boltzmann–Shannon interaction entropy: A normalized measure for continuous variables with an application as a subsample quality metric. *Chaos Interdiscip. J. Nonlinear Sci.* **2023**, *33*, 123131.