Entropy, Probabilistic Harmonic Space, and the Harmony of Antonio Carlos Jobim

Entropia, o espaço harmônico probabilístico e a harmonia de Antonio Carlos Jobim

Carlos de Lemos Almada
Hugo Carvalho
Universidade Federal do Rio de Janeiro

Abstract: This paper introduces a theoretical framework derived from a deep and detailed harmonic analysis of songs composed by Antonio Carlos Jobim, focusing on two components, namely, "semantic" (related to the idea of *chord type*) and "syntactic" (involving *binary relations* between contiguous chords). The research is mainly focused on investigating the correlations between compositional style (here related to the harmonic construction) and the concepts of probability, expectance, and, especially *entropy*, being the latter defined as a measure of uncertainty or "surprise" of events along time. After a bibliographical review of these topics and their applications to music, a section exposes Markov Chains, a mathematical tool used to formalize the "semantic-syntactic" harmonic relations statistically inferred in the analyzed *corpus* of Jobim's works. Then it follows the formalization of a *probabilistic harmonic space* and the concept of *probabilistic index*, directly associated with the entropy of the observed binary relations. This approach opens a new analytical perspective, also allowing the generalization of the presented theoretical and methodological technology for the examination of other repertoires and posterior comparison, presenting then as a new mean of investigation on the nature of style.

Keywords: Jobim. "Semantic-syntactic" harmonic relations. Entropy and probability. Markov chains. Probabilistic Harmonic Space.

Resumo: Este artigo introduz um arcabouço teórico derivado de uma profunda e detalhada análise harmônica da integral de canções de Antonio Carlos Jobim, considerando aspectos "semânticos" (relacionado à ideia de tipos acordais) e "sintáticos" (envolvendo relações binárias entre acordes contíguos). O foco do estudo concentra-se na investigação das correlações existentes entre estilo composicional (no caso, relacionado à construção harmônica) e os conceitos de probabilidade, expectativa e, especialmente, entropia, definida como a medida da incerteza de eventos decorrentes no tempo. Após uma revisão

Revista da Associação Brasileira de Teoria e Análise Musical
Journal of the Brazilian Society for Music Theory and Analysis
@ TeMA 2022 – ISSN 2525-5541

bibliográfica sobre tais tópicos e de suas aplicações em música, uma seção dedicada a Cadeias de Markov prepara o exame central do objeto de estudo, a saber, as relações harmônicas "semântico-sintáticas" levantadas estatisticamente no repertório analisado das canções jobinianas. A formalização de um *espaço harmônico probabilístico* e do conceito de *índice probabilístico*, diretamente associados à entropia medida referente ao comportamento dos acordes das canções abrem uma perspectiva analítica original, permitindo ainda a generalização do aparato teórico-metodológico para o exame de outros repertórios e posterior comparação, apresentando-se, portanto, como um novo meio de investigação sobre a natureza do estilo.

Palavras-chave: Jobim. Relações harmônicas "semânticas-sintáticas". Entropia e probabilidade. Cadeias de Markov. Espaço Harmônico Probabilístico.

<p style="text-align:center">*   *   *</p>

1. Background on Information Theory

In 1948, American mathematician and electrical engineer Claude Elwood Shannon (1916–2001) published a seminal paper entitled "A Mathematical Theory of Communication", which settled the very basis for what would be later known as Information Theory.[1] Shannon was especially interested in what he considered a "fundamental problem of communication [namely,] that of reproducing at one point either exactly or approximately a message selected at another point" (Shannon 1948, p. 1). As he stresses in the introduction of his work, the notion of a meaning associated with a given message is not particularly relevant in his approach, and his goal should be an exclusively engineering problem. Despite the theory there developed being sufficiently general to allow for more general scenarios, a first approach proposed by Shannon was to consider messages produced by a discrete source, to model devices like the telegraph. Since messages of interest usually are not unstructured sequences of random symbols, a more sophisticated mathematical model for the informal term "message" was necessary. Therefore, Shannon's theory is strongly centered on a specific kind of stochastic process, the so-called *Markovian process*. In short, this

---

[1] In 1949 this paper was published as a book with the slightly different title "The Mathematical Theory of Communication". It seems a minor modification, but it indicates the promptly noticed generality and importance of the work. Indeed, in 2020 Google Scholar indicates that Shannon (1948) is the most cited paper in Mathematics and correlated areas. In Cover 2006, a modern and comprehensive introduction to Information Theory is presented.

mathematical object is specifically tailored to model and quantify the chance of observing a certain symbol after the observation of the preceding ones. More will be developed on Markov process further ahead (see Section 2), and for the moment let us continue following Shannon's argumentation.

Firstly, Shannon introduces the notion of *n-gram*, which could be defined as a sequence of informational unities with a generic number *n* of elements. He then proposes to the reader to imagine a hypothetic generator of typological signs with could produce sequences of different sizes over a fixed alphabet, that for simplicity, we will assume as the English alphabet composed of 26 letters, not making distinction between upper and lower cases: *monograms* (A, Z, E, O, ...), *digrams* (BC, AK, NW, …), *trigrams* (VIQ, ZMP, …), and so on. Note that the set of *n*-grams increases exponentially, each containing 26^n symbols.

For the sake of illustration, assume that the production of a given symbol by his "machine" (say "J") does not give any cue for the next one. Put another way, the probability of observing any symbol after "J", including its repetition, does not depend on the previous symbol being a "J". Without an underlying language, it is reasonable to assume that all 26 alternatives are equally probable, and equal to $\frac{1}{26} \approx 0.04$, for all letters. This can also be extended to larger-size grams, which would imply, of course, in dramatic diminution of the magnitude of the probabilities. Taken for example trigram "GJQ": the probability of it to be followed by, say, "TTX" would be $\frac{1}{26^3} \approx 0.00006$.

On the other hand, if these structures are contextualized, for example in a natural language as English, things change considerably, since symbols are not uniformly distributed, and, moreover, succession now is strongly conditioned by semantic and syntactic forces. Shannon proposes then the possibility of stochastic production of "words" of different sizes in his "machine", considering distinct *orders of approximation*: zero-order[2] (which means that the probability of occurrence of a given symbol depends only on its relative frequency within the context, being *not* dependent on the surrounding symbols), and *k*-order, for $k \geq$ 1 (meaning that the probability of observing a given symbol depends on the *k* previously observed symbols).

---

[2] Also denoted as *zeroth-order* expectation by modern authors, like David Huron (2006). The same applies to the remaining approximation types of Shannon (i.e., one-order/first-order, etc.).

Shannon then demonstrates how his "machine" functions, firstly based only on relative frequency of monograms in an English *corpus* of texts. This analysis confirms the two aforementioned aspects, namely: now the letters have distinct probabilities of occurrence (for example, usually "E" is many times more frequent than "X"), and the context informs about the occurrence of a particular symbol (observing a "Q" strongly conditions "U" as the next symbol). After this, Shannon shows that the increasing of order of approximation turns the production of hypothetical sentences gradually close to normal textual constructions. This amazing formulation (in an epoch where computers were almost an idealized conception) became one of the very foundations for algorithmic creation and machine-learning processes.

Maybe the most far-reaching aspect of Shannon's theory is the concept of *entropy of a random variable*. The term, borrowed from the field of Thermodynamics (in which is associated with the degree of disorder of a closed system), receives in his work a distinct, rather co-related meaning, adapted to the context of the Information Theory. For Shannon, the entropy (traditionally denoted by $\mathcal{H}$) of a given random variable represents the degree of *uncertainty* with respect to the context it is inserted to. In his words,

> Suppose we have a set of m possible events whose probabilities of occurrence are $p_1, p_2, \dots, p_m$. These probabilities are known, but that is all we know concerning which event will occur. Can we find a measure of how much "choice" is involved in the selection of the event, or of how uncertain we are of the outcome? If there is such a measure, say $\mathcal{H}(p_1, \dots, p_m)$, it is reasonable to require of it the following properties:
>
> 1. $\mathcal{H}$ should be continuous on each p_i.
>
> 2. If all the p_i are equal, $p_i = \frac{1}{m}$, then $\mathcal{H}$ should be a monotonic increasing function of m. With equally likely events there is more choice, or uncertainty, when there are more possible events.
>
> 3. If a choice be broken down into two successive choices, the original $\mathcal{H}$ should be the weighted sum of the individual values of $\mathcal{H}$ (Shannon, 1948, p. 10).

Shannon proves that there is only one function that satisfies these three axioms, up to a positive multiplicative constant:

$$\mathcal{H} = -\sum_{i=1}^{m} p_i . \log_2(p_i). \tag{1}$$

Regardless of the basis, the quantity $-\log(p_i)$ can be interpreted to measure the *surprise* of observing an event of probability p_i. Indeed, observing an event of probability one should give no surprise, and on the other hand, the occurrence of an impossible event brings an infinite amount of surprise. Moreover, this value decreases from infinity to zero as p_i increases from 0 to 1.[3] Note that this same interpretation also holds if one substitutes the term "surprise" by *information*. Therefore, the entropy of a random variable can be interpreted as the *average surprise* or *average information* that it carries.

The logarithmic base chosen (i.e., 2), among other possible (as Shannon discusses), has the important advantage to provide a quantification of entropy in terms of *binary digits* (or bits).[4] This choice becomes especially important if we interpret the entropy as the *average number of bits per symbol* necessary to encode a message written following the probability distribution $p_1, \ldots, p_m$. This interpretation is not so clear at the first sight, but it is an important theorem within Information Theory (Cover; Thomas 2006, p. 62).

Another interpretation of entropy is the average minimal number of binary questions that are necessary to identify a particular value observed from the probability distribution $p_1, \ldots, p_m$. Consider, for example, a game that consists of trying to guess a number between 1 and 6 before rolling an honest dice by means of binary questions, like "is the value contained in set S?", and so on. The probability of observing any value on the dice is equal to $\frac{1}{6}$. If we enter these probabilities in Eq. (1) we obtain an entropy of $\mathcal{H} = 2.585$. Now, suppose that the dice is someway modified, in such way that some faces become more likely than others, according to the following distinct probabilities: $p_1 = 0.05; p_2 = 0.10; p_3 = 0.35; p_4 = 0.02; p_5 = 0.20; p_6 = 0.28$. In this case we have $\mathcal{H} = 2.1699$, a lower value than that obtained when rolling the honest dice. This means that the second experiment is *less uncertain* than the first one, and therefore, requires, in average, less binary questions to guess the observed value.

---

[3] There are infinitely many continuous decreasing functions from the interval (0,1] to the set of real numbers, but in the same spirit of Shannon, it can be proven that the logarithm is the only reasonable function that captures the intuition behind the concept of *surprise*, up to a multiplicative constant (Ross 2010, p. 425).

[4] Shannon credits the creation of the nowadays popular abbreviation "bit" to J.W. Turkey.

2. Background on Markov Chains

The history of the Markov Chains dates to the beginning of the 20th century when Andrey Andreyevich Markov proposed this mathematical object as a counterexample to a statement made by his intellectual rival Pavel Nekrasov. Essentially, the same theory was independently rediscovered a decade later by Agner Krarup Erlang, a Danish mathematician and engineer, when studying queues of telephone calls on a hub. The history of the Markov Chains can be more explored in (Carvalho 2019), (Maia 2016), and references therein, and in this section, we provide the fundamentals necessary to apply it to modelling transition between musical objects. For a brief introduction to Markov Chains, with a more probabilistic approach, see Ross 2006, pp. 419–24 and de Groot; Schervish 2012, pp. 188–200.

Intuitively, a Markov Chain is a process where the observed state at some time instant statistically depends only on the observed state in the immediately preceding time. Consider, for example, a quite boring sequence of pitch-classes where the n-th one can be either the same as the $(n-1)$-th, one semitone above or one semitone below it, all with equal probability 1/3 (consider B as the pitch class preceding C and C as the pich-class succeeding B). Assume also that the initial pitch class is randomly chosen with equal probability 1/12 from the set of the twelve pitch-classes. This is a very basic example of a Markov Chain, but it can be used to introduce some fundamental concepts of the more general theory: the set of all the possible observed states is called the *state space*, and in this example, it is given by $\Lambda = \{$C, C♯, D, D♯, E, F, F♯, G, G♯, A, A♯, B$\}$; vector $\lambda =$ [1/12 … 1/12], containing 12 entries, is the *initial probability distribution*; and finally, the *transition probabilities matrix*, denoted by $\boldsymbol{M}$, is constructed and given by:

$$
\boldsymbol{M} = \begin{bmatrix}
1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 \\
1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \\
1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 1/3
\end{bmatrix}.
$$

The value on the i-th row and j-th column is the probability of jumping to the pitch class j given that the current pitch class is i, for $i, j \in \Lambda$. Therefore, this Markov Chain can be abstractly described by the triple $(\Lambda, \lambda, \mathbf{M})$. Notice that if the set Λ is also abstracted from the musical intuition and being allowed to be any generic set with 12 elements, the same mathematical framework holds in several distinct scenarios. Endowed with this intuition, we now formulate the Markov Chains with more mathematical details.

Denote by $\Lambda = \{x_1, \ldots, x_N\}$ a finite set of objects of interest, which will be called the *state-space* and its elements will be called *states*. Let also $X_1, X_2, X_3, \ldots$ be a sequence of variables, assuming values on the set Λ but which specific value depends on random factors.[5] The sequence $X_1, X_2, X_3, \ldots$ is called a *Markov Chain* if the probability of the variable X_n assuming any particular value on the set Λ depends only on the value observed of the variable X_{n-1}, for all $n = 2, 3, 4, \ldots$. In mathematical terms, we can write that

$$P\big(X_n = x_j \big| X_{n-1} = x_i, \ldots, X_1 = x_k\big) = P\big(X_n = x_j \big| X_{n-1} = x_i\big), \tag{2}$$

where the symbol | denotes a *conditional probability*, meaning that the event on its left side is considered in the light of the occurrence of the event on the right side. At the beginning of the process, there is no "past" to condition on, so the probability distribution of the first-time instant X_1 is distinctively denoted by a vector of probabilities $\lambda = [p_1, \ldots, p_n]$, composed by positive real numbers that sum up to 1.

Of all these concepts, the most abstract when regarding musical applications are the states, because of its genericness. However, it is exactly this aspect that makes the Markov Chains theory widely applicable, not only to musical scenarios but also in other areas of Science and Engineering as well. The researcher can choose whatever it is desirable to use as a set of states, and particular instances will be discussed later in this section. For the moment, we focus instead on mathematical aspects of the theory.

The set of probabilities $P\big(X_n = x_j \big| X_{n-1} = x_i\big)$, for all $x_i, x_j \in \Lambda$, are called the *transition probabilities* and abbreviated by p_{ij}, for $i, j = 1, \ldots, N$. In this work it is assumed that these probabilities are constant along time, and this collection can be assembled in the *transition probabilities matrix*, denoted by $\mathbf{M}$, where the

---

[5] This object is called a *random variable* in the statistical literature.

MUSICA THEORICA   Revista da Associação Brasileira de Teoria e Análise Musical 2022,
v. 7, n. 1, p. 68–111 – Journal of the Brazilian Society for Music Theory
and Analysis @ TeMA 2022 – ISSN 2525-5541

75

element on the i-th row and j-th column precisely contains the value of p_{ij}. Notice that the sum of each row of this matrix is equal to 1, since it is given by

$$\sum_{j=1}^{N} p_{ij} = \sum_{j=1}^{N} P(X_n = x_j | X_{n-1} = x_i), \tag{3}$$

that is, a quantity that describes the situation: "conditioned on departing from the state x_i, the probability of arriving at any one of the n possible states".

The probabilities contained in matrix $\boldsymbol{M}$ also allow us to answer the following question: "if departing from the state x_i, what is the probability of arriving in state x_j in exactly k steps?". For $k = 2$, it is easily interpreted that this answer is given by:

$$\begin{aligned} p_{ij}^{(2)} &= P(X_n = x_j | X_{n-2} = x_i) \\ &= \sum_{k=1}^{N} P(X_n = x_j | X_{n-1} = x_k) P(X_{n-1} = x_k | X_{n-2} = x_i), \end{aligned} \tag{4}$$

since the right hand-side of this equation encodes exactly the sentence "the probability of transitioning from x_i to x_k in one step and then transitioning from x_k and x_j also in one step but considered among all the possible values of x_k, for $k = 1, \dots, N$. From the theory of Linear Algebra, the quantity $p_{ij}^{(2)}$ defined above is exactly the entry in the i-th row and j-th column of the matrix $\boldsymbol{M}^2$. It is also possible to prove that an analogous result also holds for larger steps, namely the *Chapman-Kolmogorov equations*: "the entry in the i-th row and j-th column of the matrix $\boldsymbol{M}^k$, denoted by $p_{ij}^{(k)}$, is exactly the probability of arriving in k steps in state x_j if departing from state x_i".

Finally, another important aspect of the Markov Chain theory, especially in musical applications, is the *order* of the chain. The example and formalization here presented are for the case of a *first-order* chain, that is, when the transition probability is given only by the previous and current step. In musical applications, it is sometimes reasonable to have a dependence on a more distant past to some chord passages, for example, and these phenomena are not captured by a first-order chain. A possible solution to this issue is to increase the order of the chain, by considering probabilities of transition of the following type:

$$p_{ijk} = P(X_n = x_k | X_{n-1} = x_j, X_{n-2} = x_i), \tag{5}$$

which gives rise to a *second-order* Markov Chain. Clearly this procedure can be generalized to any order of interest, but unfortunately it is not the most interesting solution, mainly because of the increase in the number of parameters to work with, since the size of the probability transition matrix will severely increase. A more parsimonious solution and the one adopted in this work is to maintain a first-order Markov Chain but with a special focus on modelling the state-space with more sophisticated objects than simply chords in such a way that the modelling becomes more realistic.

The first applications of Markov Chains to Music were not endowed with this sophistication on wisely modeling the state-space, but no less important, mainly because of the pioneering and the promising results obtained. In the beginning of the decade of 1950, Harry Olson and Hebert Belar analyzed a *corpus* of 11 melodies by Stephen Foster, properly transposed to the same tonality, and estimated transition probabilities of order 0,[6] 1, and 2 between pitches; the rhythmic patterns were analyzed only on its relative frequency of occurrence. With this information, they developed the first prototype of a "synthesizer", that generated random music according to these probabilities.

Previously to the publication of Olson's and Belar's work, which occurred in the beginning of the decade of 1960, Iannis Xenakis pioneered a theoretical study on the limits of generality when modelling the state-space. Indeed, in his book Formalized Music he proposes to consider Markov Chains on the set of *screens*, an object that described the frequency occurring at any particular time instant on a sonic manifestation. Nowadays the screens can be interpreted in the light of the Short-Time Fourier Transform, a powerful tool to perform time-frequency analysis of audio signals; Xenakis was aware of the existence of this concept, through the work of Gabor, but the computational power to make this approach practical was not available at the time. So Xenakis circumvented this issue by manually creating a set of screens and a probability transition matrix among them and used a slight modification of a realization of this chain to generate a sequence of screens, which then he transported to usual music notation. The result of this experiment are the pieces *Analogique A* and *Analogique B*. For more detail on the creation of these pieces see Carvalho (2019).

---

[6] A Markov Chain of order 0 is described only by the relative frequency of observation of its states.

MUSICA THEORICA   Revista da Associação Brasileira de Teoria e Análise Musical 2022,
v. 7, n. 1, p. 68–111 – Journal of the Brazilian Society for Music Theory
and Analysis @ TeMA 2022 – ISSN 2525-5541

77

The first experiment using Markov Chains to model harmony dates back from 1957, by Lejaren Hiller and Leonard Isaacson. Instead of directly estimating the transition probabilities from a *corpus* of interest they proposed one by giving more weight for consonant and smaller intervals. Using this material, they composed the *Illiac Suite*, later renamed to *String Quartet No. 4*, which is considered the first musical score generated by a computer.

The history of Markov models in music is very extensive and is still under development, not only in Music Theory but also in Signal Processing, since Hidden Markov Models are widely employed to perform automatic chord recognition, among other tasks (Müller 2015). For a more extensive bibliographical review on the beginning of this story, see Ames 1989.

3. Information Theory and Markov Chains in Music

Due to the attractive potentials of Shannon-Weaver's theory to be extended to other fields besides Electrical Engineering, it is not surprising that just a few years after its publication Leonard Meyer wrote an article entitled *Meaning in Music and Information Theory* (1957). In his study, Meyer intends, in fact, not only to evidence the, in his own words, "striking" parallelism and equivalences between the music experience and Information Theory, but mainly to investigate a closely related topic, *meaning*, an aspect, as aforementioned, not covered by Shannon. Meyer begins resuming one of his most famous theses,[7] which expresses that meaning in music results from "the arousal and subsequent inhibition of expectant tendencies in the shaping of musical experience," connecting it with a general definition of meaning, proposed by the logician Morris Cohen: "anything acquires meaning if it is connected with, or indicates, or refers to, something beyond itself, so that its full nature points to and is revealed in that connection" (Meyer 1957, p. 412). Since music can essentially be understood as information transmitted across time, probability has an enormous importance in the cognitive process of meaning and style (in fact, the very kernel of Meyer's interest), which orients the subsequent discussion in his work.

He firstly proposes a subdivision of musical meaning between two complementary but interacting categories: designative (or connotative, external

---

[7] Introduced in his seminal book *Emotion and Meaning in Music* (1956).

to musical events of a piece), and embodied (referring to events present in the musical context in question). Although arguing that designative meaning (associated with musical character) also influences the determination of style, Meyer prefers to address specifically the latter category in his paper. In this context, he defines style as "the universe of discourse within which musical meanings arise," adding that, apart from particularities of epoch, geographic, composers, musical systems and idioms, and so on,

> what remains constant from style to style [… is] in fact the psychology of human mental processes – the ways in which the mind, operating within the context of culturally established norms, selects and organizes the stimuli that are presented to it (Meyer 1957, p. 413).

By adding then that style "is a complex system of probabilities" that arises naturally "expectation" (ibid., p. 414), Meyer manages to properly connect his study with Information Theory.

On the other hand, expectation[8] can also be classified as *latent* or *active*. While the former concerns basically "the probability relations embodied in a particular musical style" (for example, the strong tendency of the leading tone toward the tonic in tonal music), expectation is activated when normality is in some way disturbed. As Meyer interestingly emphasizes, "meaning arises when an individual becomes aware, either affectively or intellectually, of the implications of a stimulus in a particular context" (ibid., p. 415). This leads the author to suggest that "musical meaning arises when an antecedent situation, requiring an estimate as to the probable modes of pattern continuation, produces uncertainty as to the temporal tonal nature of the expected consequent" (ibid., p. 416). Just after this, Meyer inserts the notion of entropy in his argumentation, relating properly meaning and information as dependent on uncertainty:

> Information is measured by the randomness of the choices possible in a given situation. If a situation is highly organized and the possible consequents in the pattern process have a high degree of probability, then information (or entropy) is low. If, however, the situation is characterized by a high degree of shuffledness so that the consequents are more or less equi-probable, then information (or entropy) is said to be high (ibid., p. 416).

---

[8] This notion of "musical expectation" can be related to the probabilistic concepts of conditional expectation and forecasting, but we will not draw this parallel here. For more details, see Temperley 2007.

MUSICA THEORICA   Revista da Associação Brasileira de Teoria e Análise Musical 2022, v. 7, n. 1, p. 68–111 – Journal of the Brazilian Society for Music Theory and Analysis @ TeMA 2022 – ISSN 2525-5541

79

Another central concept of Information Theory that is perfectly applicable to the understanding of musical processing is *redundancy*. Like expectation, redundancy is governed by statistical conditions and, in a sense, can be seen as much more intense in musical situations.[9] As affirmed by Meyer,

> Just as letters can be left out of a written statement or words omitted from a message without affecting our ability to understand and reconstruct the word or message, so tones can be omitted from a musical passage without affecting our ability to grasp its meaning (ibid., p. 418).

This means that, due to redundancy, one is capable of reconstructing musical omitted information given its context (properly associated with stylistic norms). Fig. 1 provides a simple illustration of this aspect, depicting an archetypical cadential formula[10] of a hypothetical piece. Despite lacking the necessary elements for an unequivocal analytical labelling of the harmonic functions involved (more specifically, pitch classes C in the first chord, B in the second, and E and G in the third), anyone minimally familiarized with the common-practice conventions would have no difficulties to hear functionally the passage as it is expressed in the Roman numerals written below the score.

Figure 1: Example of a perfect authentic cadence with omitted notes

Redundancy is a common resource of the compositional palette (normally associated with manipulation of textures and densities), and it is in general easily decoded by listeners (like exemplified above). However, this shared property is frequently explored by composers to promote ambiguity, a very useful spice in musical construction. Consider, for example, an alternative version of the

---

[9] See for example David Huron's commentaries about the larger amount of redundancy in music in comparison with that associated with language, which is almost negligible (Huron 2006, p. 244).

[10] In more technical terms, a perfect authentic cadence, or PAC.

cadence of Fig. 1, in which the high expectancy of tonic resolution provoked by the dominant preparation is abruptly frustrated, as shown in Fig. 2. The arrows attempt to reconstruct the cognition of the detour: initially, the very surprising arrival of the second-inversion B-major triad arouses some confusion. In the immediate continuation, the listener recognizes the resumption of the same cadential formula, but transposed a semitone downwards; finally, the so-presumed dominant rooted in G is retroactively reinterpreted as a German-sixth chord of the new key of B major, which also implies a functional reinterpretation of its constituents: the root G becomes the flatted fifth related to an omitted root C♯; the previously fifth D is functionally associated with a flatted ninth. The chord is completed by two "physically" absent (but inferred) pitch classes, B and F,[11] that have their functions inverted, as seventh and third (in this case, as the enharmonically equivalent E♯). Using Meyer's conceptualization, the striking break of expectations in this situation arises a new possible meaning for the harmonic formula to the surprised hypothetical listener, so far habituated only to the default resolution.

Figure 2: Example of ambiguous harmonic situation. Arrows are intended to reconstruct the listener's interpretation.

---

[11] Which were, at the first impression, "mistakenly" labeled as third and seventh.

Consider now that the same listener, amazed by the new experience, becomes interested in knowing other pieces of that composer or style. We can conceive that now the listener's expectancy related to this specific cadence has somehow changed. The *uncertainty* of the formula for him/her was definitively increased. Mathematically, we can think that the cadences employed by the hypothetical composer are the output of some unknown random variable. As the listener hears only the excerpt in Fig. 1, he infers that this is the only possible cadence, implying that this random variable has zero entropy.[12] But, after being aware of the cadence in Fig. 2, he may associate a positive probability to it. However, this value should be not as high as that of the first cadence: since in this second case the musical content brings more surprise to the listener, its probability of occurrence should be quite low. Finally, since now there is some uncertainty associated with the observed cadence, the respective random variable has more entropy than in the first case.

This simple example allows us to see musical styles as very complex and dynamic systems, formatted by sets of specific norms, but also by the breaking of these rules, in distinct degrees of probabilities. As pointed out by Meyer, statistical analysis becomes a powerful tool for depicting particularities of a style. However, as he says "the mere collection and counting of phenomena do not lead to significant concepts. Behind any statistical investigation must be hypotheses that determine which facts shall be collected and counted" (Meyer 1957, p. 421). In other words, the understanding of the *conditions* that characterize the processes to be studied is a primordial factor for initiating the analysis.

Several empirical studies followed closely the pioneering explorations of Meyer in the confluence between music and Information Theory. In 1958, Joseph Youngblood, in an article entitled "Style as Information", proposed "to explore the usefulness of Information Theory as a method of identifying musical style" (p. 24). The main goal of his study is to "attempt to determine the extent of the restrictions under which certain composers worked; [however] it will not attempt to explain why certain combinations were favored and others eschewed" (p. 25). Youngblood examines a *corpus* of 20 pieces (whose melodic lines were previously analyzed) composed by Schubert (eight songs), Mendelssohn (six arias), and

---

[12] Recall that sure events do not bring any information and no surprise. Mathematically, $\mathcal{H}(1) = -1.\log_2(1) = 0$.

Schumann (six songs),[13] mainly with respect to first-order probabilities (i.e., involving bi-grams, that is, blocks of two contiguous notes), through the construction of probability transition matrices. With the zeroth- and first-order probabilities associated with each composer, the author calculated then their entropies. Additionally, he also calculated the relative entropy ($\mathcal{H}_r$) for any case, which, as already defined, corresponds in this specific case to the ratio between the nominal entropy and the entropy which would result if all chromatic notes were equally distributed (i.e., $p(C) = p(C\sharp) = p(D) = \ldots = p(B) = 1/12$), which equals 3.58 bits. Finally, he proposes to quantify redundancy (R) as the difference between the maximum probability 1.00 and $\mathcal{H}_r$. Table 1 summarizes the data obtained by Youngblood.

	zeroth-order			first-order (average)		
	$\mathcal{H}$	$\mathcal{H}_r$	R	$\mathcal{H}$	$\mathcal{H}_r$	R
Schubert	3.13	0.87	0.13	5.37	0.75	0.25
Mendelssohn	3.04	0.85	0.15	5.54	0.74	0.26
Schumann	3.05	0.85	0.15	5.52	0.77	0.23

Table 1: Entropy (H), relative entropy (Hr), and redundancy (R) in a *corpus* of works by Schubert, Mendelssohn, and Chopin (adapted from Youngblood 1958, p. 32)

In the article "Information Theory Analyses of Four Sonata Expositions", published in 1966, Lejaren Hiller and Calvert Bean, motivated by the argument that "most music not only has an average information level somewhere between chaos and total order, but also usually has increases and decreases of information during its time duration" (p. 102), propose a statistical analysis of the expository sections of the initial movements of four piano sonatas (Mozart's K. 545, Beethoven's Op. 90, Berg's Op. 1,[14] and Hindemith's second sonata), in order to investigate if pitch distribution can be in some way associated with structural expectations, as preconized by Meyer. Differently from Youngblood, the authors take for their universe, rather than 12, 21 pitch classes (that is, disregarding

---

[13] He also applies the method in a statistical analysis of selected pieces of Gregorian chant, aiming at a comparison with the first *corpus*.

[14] In this case, the unique movement.

MUSICA THEORICA   Revista da Associação Brasileira de Teoria e Análise Musical 2022, v. 7, n. 1, p. 68–111 – Journal of the Brazilian Society for Music Theory and Analysis @ TeMA 2022 – ISSN 2525-5541

83

enharmonic equivalence), in order to capture with more efficiency certain tendencies of notes depending on contextual conditions.[15]

In the next year (1967), another article written by Hiller, this time co-authored by Ramon Fuller, extends the informational analytical approach to serial music, by examining Webern's Symphony Op.21. The alphabet considered (that is, the set of symbols under investigation) is formed now by melodic intervals, rather than pitch classes. At the end of the study, the authors detect two practical limitations of the analytical method, concerning the sizes of the sample and of the alphabet. As recognized by them, "if the analyst works with large samples and small alphabets, he obtains reliable but generalized results; if he works with smaller samples and/or larger alphabets, he obtains more specific but less reliable results" (pp. 110–1). Despite this dilemma, they conclude that analysis from an Information-Theory viewpoint is a potential valuable tool in the investigation of style, given that it can "provide a valuable summary of the effect of the number of symbols used, their relative frequency, and their combinatorial arrangements upon the structural complexity of a musical composition" (p. 101).

The problem of the sample length is also a major concern in a paper written in 1983 by Leon Knopoff and William Hutchinson. Taking as reference Youngblood's study, the authors argue about the relative size of samples in statistical analysis addressing musical styles. The authors pose several questions in order to orient their approach, which expands considerably the Youngblood's repertoire. Limiting the study to zeroth-order probabilities of isolated notes, they obtain some discrepant results for entropy considering a same composer (Mozart, for example) and different *corpora*, which allows them for questioning the efficacy of the calculation of entropy as a tool for determining style.

In the work "On the Entropy of Music: An Experiment with Bach Chorale Melodies" (1992), Leonard Manzara proposes an experiment to test the capacity of listeners for predicting the continuation of a melodic event given an already familiarized context. Manzara assumes that "the information content, or 'entropy', of a piece of music cannot be determined in the abstract, but depends on the listener's familiarity with, and knowledge of, the genre to which it belongs" (p. 81). For the experiment, Manzara selects Bach's collection of 371

---

[15] That is, the authors differ in this approach flats from sharps, according to the tonal context. For example, normally pitch class B♭ in F major has a diametrically distinct behavior from its enharmonic equivalent A♯ in B major.

chorales as *corpus* of study and elaborates a computational game, named "Chorale Casino",[16] in which a user is asked to guess the proper continuation for a given note of the soprano melody of a random chosen chorale. As in a slot machine of a real casino, bets (associated with imaginary money in this case) are made according to the individual plausibility that a user has with respect to a continuation of the probed event. More specifically, this ingenious experiment explores entropy of the distribution of pitches[17] (or the information conveyed by these) considering the familiarization of listeners with a musical idiom and their latent expectation (in Meyer's terms). Manzara organized then a tournament in which fifteen contestants (grouped in three levels of musical experience) competed using the program. Two pieces were selected for the tournament, the chorales 61 (in E♭ major) and 151 (in G major). Afterwards, Manzara calculated the pitch entropy for both melodies: 1.529 (chorale 61) and 1.974 (chorale 151), relatively low values that indicate a considerable (average) predictability in both cases. He also elaborated *instantaneous entropy profiles* for the two pieces, aiming at a "detailed note-by-note of the melody for each chorale" (p. 86). These profiles, plotted graphically, reveal that the lowest values for entropy coincide with the fermatas of the scores, which represent cadential reposes matching the end of textual verses. More specifically, authentic cadences have, in average, still lower entropy than half cadences.

Among a multitude of relations between music and mathematics studied in the first volume of *Musimathics*, written in 2006 by Gareth Loy, there is a short section dedicated to Information Theory, focusing especially on the concepts of entropy and redundancy (pp. 343–9). After a review of Shannon's theory, Loy illustrates the relation between entropy and music expectancy with an interesting example, addressing motivic construction in a hypothetical piece. Fig. 4 aims to reconstruct Loy's basic idea by including a graphical representation of the involved relations and elements. Consider that a listener is exposed to a simple eight-note motive, contextualized in the key of D major. For simplicity, this motive is assumed as occurring at the very beginning of the piece (that is, no

---

[16] Based on a similar experiment organized by T. Cover and R. King, addressing entropy in English printed texts.

[17] Since Bach's chorales are essentially homorhythmic (that is, quarter note is largely the most recurrent durational figure in the pieces), Mancara manages to isolate pitch from other dimensions, decreasing the number of variables of his empirical study.

intraopus information was yet stored in the listener's memory). Consequently, the expectancy of the next "move" (represented in the figure by the level above that of the real music events) is virtually null, which means that any possible musical situation can, theoretically, follow the motive. In terms of information, entropy at this initial moment is relatively high.[18] Consider now that a diatonic sequence transposed a second higher is played. This new event triggers a series of cognitive processes in the listener's mind: firstly, he/she recognizes retroactively, with the help of the short-term memory, that this new melodic fragment is the second unity of a model-sequence scheme, and immediately projects the logic continuation of the pattern, which will start with F♯, the third scalar degree. This leads entropy to drop radically. The subsequent matching with listener's expectancy turns entropy even lower since the projection of a new sequence (beginning with G) becomes almost a certainty in his/her mind. However, the melodic chain is suddenly broken, replaced by an apparently cadential closure. The amount of information conveyed by the new fragment is raised to the initial level. Memory is not anymore capable of providing support for expectancy of the immediate continuation.
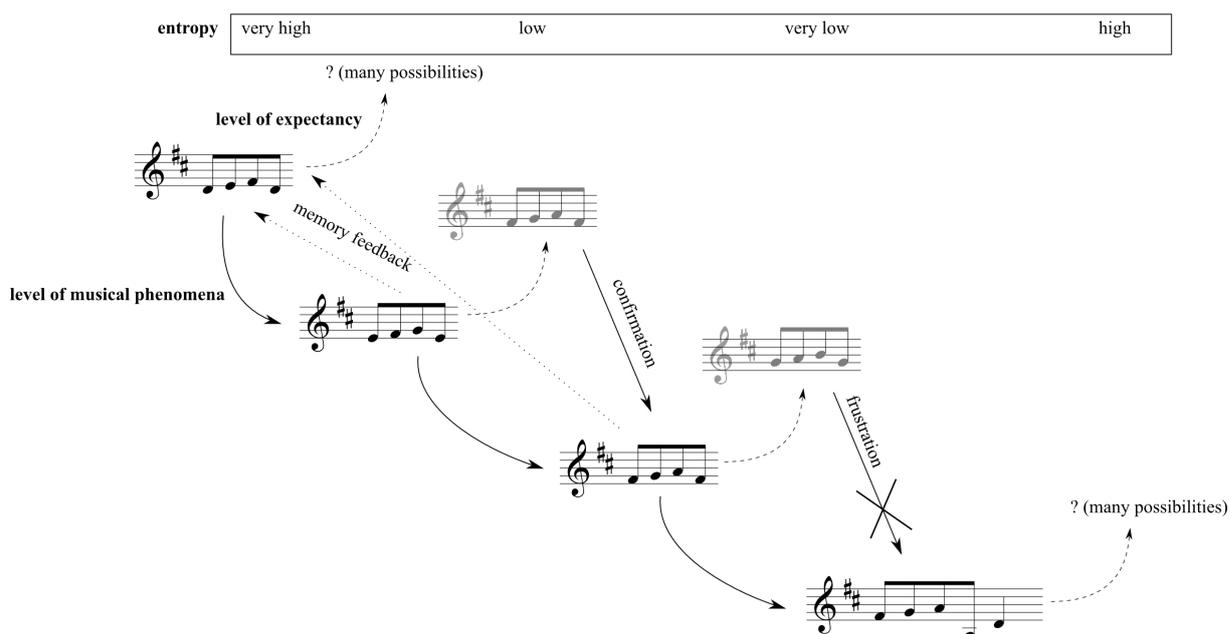

Figure 3: Example of entropy in the manipulation of a melodic sequence (adapted from Loy (2006, pp. 348–9))

---

[18] For the purposes of this illustration, it is enough the qualification of the entropy, rather than its precise quantification.

In the introduction of *Sweet Anticipation* (2006), David Huron affirms his intention of revisiting Meyer's pioneer studies, and to "recast the discussion in light of contemporary findings. [Huron's principal purpose] is to fill in the details and to describe a comprehensive theory of expectation" (Huron 2016, p. 3). The central aspect of Huron's comprehensive and deep study lays the notion of statistical learning, as "the origin of auditory expectations" (viii), a topic supported by many discussions involving original ideas, concepts, premises, and experiments, which ultimately associate musical experience with realms like evolutionary survival strategies, psychological and cognitive mechanisms, as well as cultural conventions.

The concept of entropy is firstly addressed in the section entitled "Subjective probability and uncertainty" (p. 53). After summarizing the basic aspects of Shannon's ideas, Huron presents an interesting experiment made with two groups, formed by Balinese and American musicians. An unfamiliar Balinese melody played by a *peng ugal* (a typical instrument with a range of only ten possible pitches) is presented, note to note, to both groups. Any subject is asked to guess which would be the best continuation at each point. Since there are only ten alternatives for each note, maximum entropy (calculated by using Eq. 1 with $p_i = \frac{1}{10}$, for $i = 1, \dots, 10$) equals 3.22 bits, corresponding to total uncertainty at a given point with respect to the next event. This was almost the value estimated considering American listeners (3.2 bits) after hearing the first note of the melody. In contrast, the estimate of the entropy related to the expectations of Balinese listeners after the first note was considerably lower: 2.8 bits, which indicates that learning from data is a process with a strong cultural component. As calculated by Huron, 2.8 bits corresponds to approximately seven equiprobable states for the second note. Continuing the experiment, the estimated entropy decreases as the melody is played for Balinese subjects, reaching 2.35 bits at the fifth note (roughly, five possible continuations uniformly chosen), while for American musicians it is kept in higher values. After this, uncertainty starts to increase for Balinese musicians, and at about the tenth note the estimated entropy in both groups became almost equal (rounding 3 bits), but near the end of the melody they differ again, with a relatively lower entropy experimented by the Balinese.

In Chapter 13, entitled "Creating predictability", Huron returns to entropy:

MUSICA THEORICA   Revista da Associação Brasileira de Teoria e Análise Musical 2022,
v. 7, n. 1, p. 68–111 – Journal of the Brazilian Society for Music Theory
and Analysis @ TeMA 2022 – ISSN 2525-5541

87

> Using information theory, we might say that what distinguishes one work
> from another are those elements that have a higher entropy than other works,
> but a low entropy in the context of the work itself. Said another way, we
> would look for passages or features that are (1) not commonly found in other
> works, but (2) occur frequently in the work under consideration. Formally,
> we can define a "distinctive feature" to be those passages or figures that
> exhibit a high ratio of external-to-internal entropy. Such measures of
> "distinctiveness" have long been used in quantitative stylistics, such as
> research used to determine the authorship of some text (p. 262).

The author then comments that after the enormous enthusiasm arisen from Meyer's pioneer study about the potential connections between music and the Information Theory, the interest gradually declined over the subsequent decades, due to technological limitations, regarding not only computational capacity but also the lack of sufficiently large and comprehensive databases involving distinct musical repertoires. Arguing that both obstacles were plainly straightforward to overcome in more recent times, Huron advocates the idea of statistical learning as one of the most efficient ways for studying and explaining musical style and expectations with the necessary deepness.

Entropy also plays an essential role in David Temperley's book *Music and Probability* (2007). As stated by the author in the preface, the study was motivated by an interest about the cognitive mechanisms involved in the parsing of syntactic structures, a well-studied linguistic problem that has been more recently also directed to music. For addressing the question, Temperley adopts a probabilistic approach, especially associated with Bayesian theory[19], which provides the necessary elements for the elaboration of several computer-based models, both with analytical and compositional purposes, considering metrical, rhythmic, harmonic, tonal, and pitch organizations. A general aspect considered by Temperley along the book is how some structure of a given musical situation (e.g., meter or key) can be perceived (or deduced) by a listener given superficial elements, and vice-versa, a task which involves the notion of conditional probability, the first also relying on the Bayesian interpretation of probability. In the last chapter of the book, the author defines his concept of *communicative pressure* as a decisive way for understanding musical style from the interaction of

---

[19] Bayesian Statistics is a branch of the main discipline on which Bayes' Theorem plays a central role, by allowing the probabilistic incorporation of prior knowledge of the researcher and its further update as the stream of data arrives and is processed. For more details, see Jaynes 2003.

listeners' and composers' perspectives in function of probability, bringing some refinement to Meyer's original proposal. With the help of this "powerful tool for explaining differences between styles" (p. 184), Temperley addresses perceptual discrimination of syncopated and rubato rhythms, especially when both categories are combined in a musical context.

In a further work, Temperley (2009) investigates harmony in common-practice period, interested on the principles that govern the transition between contiguous chords. Motivated by a scarcity of empirical studies on this specific subject (in contrast with the more common statistical analysis on frequency of chords in different repertoires), the author adopts as database a *corpus* formed by harmonic progressions extracted from 46 excerpts of several common practice works (by varied composers, as Bach, Haydn, Beethoven, Grieg, etc.).[20]After encoding and performing the harmonic analysis, Temperley produced a number of aggregate statistics, including a counting matrix, whose structure is reproduced in Fig. 4.[21]

Roman numerals refer to chord roots regardless the associated chordal qualities, such that, as exemplified by Temperley himself, ii and V/V are treated as the same entity. On the other hand, this strategy allows for normalizing transitions occurring in different keys and modes. The rows are occupied by the antecedent chords of a transition, while the consequent chords are disposed along the columns. A cell represents the number of occurrences of a transition formed by the corresponding pair of row and column. For example, I (first row) is followed by IV (sixth column) in 45 instances in the *corpus*. The last column (named Σ) contains the sum of the values on each line and represent the total number of "departures" from that chord root. After dividing each line by the sum of its values we obtain an *estimated probability transition matrix*, containing the estimated transition probabilities between pairs of chord roots within the K-P *corpus*. This matrix is displayed in Fig. 5.

---

[20] These progressions come from the workbook that accompanies Stefan Kostka and Dorothy Payne's book *Tonal Harmony* by McGraw-Hill (1995). Temperley names it the Kotska-Payne *corpus* (K-P *corpus* for short).

[21] For a detailed description of the processes and conventions adopted, as well as the of several explorations of the statistical data, see Temperley (2009).

MUSICA THEORICA   Revista da Associação Brasileira de Teoria e Análise Musical 2022,
v. 7, n. 1, p. 68–111 – Journal of the Brazilian Society for Music Theory
and Analysis @ TeMA 2022 – ISSN 2525-5541

89

	I	♭II	II	III	♭III	IV	♯IV	V	♭VI	VI	♭VII	VII	Σ
I	0	7	31	1	4	45	2	116	11	17	3	19	256
♭II	3	0	8	0	0	0	1	2	0	0	0	1	15
II	22	3	0	1	4	1	7	45	2	8	0	6	99
III	1	1	0	0	0	0	0	4	4	0	0	0	10
♭III	1	0	2	0	0	7	0	1	0	7	0	1	19
IV	32	2	10	0	4	0	3	11	0	1	1	4	68
♯IV	7	0	0	0	0	0	0	9	0	0	0	0	16
V	167	0	8	1	2	4	0	0	7	6	0	2	197
♭VI	5	2	8	0	1	3	0	2	0	3	2	0	26
VI	4	2	28	0	1	4	2	1	0	0	0	1	43
♭VII	0	0	0	5	0	0	0	1	0	0	0	0	6
VII	27	0	0	0	3	0	1	1	1	0	0	0	33
													788

Figure 4: Counting matrix related to K-P *corpus*, adapted from the data produced by
Temperley (2015)

	I	♭II	II	III	♭III	IV	♯IV	V	♭VI	VI	♭VII	VII
I	0	0,027	0,121	0,004	0,016	0,176	0,008	0,453	0,043	0,066	0,012	0,074
♭II	0,200	0	0,533	0	0	0	0,067	0,133	0	0	0	0,067
II	0,222	0,030	0	0,010	0,040	0,010	0,071	0,454	0,020	0,081	0	0,061
III	0,100	0,100	0	0	0	0	0	0,400	0,400	0	0	0
♭III	0,053	0	0,105	0	0	0,369	0	0,053	0	0,368	0	0,053
IV	0,471	0,029	0,147	0	0,059	0	0,044	0,162	0	0,015	0,015	0,059
♯IV	0,437	0	0	0	0	0	0	0,563	0	0	0	0
V	0,848	0	0,041	0,005	0,010	0,020	0	0	0,030	0,030	0	0,010
♭VI	0,192	0,077	0,308	0	0,038	0,115	0	0,077	0	0,115	0,077	0
VI	0,093	0,046	0,651	0	0,023	0,093	0,046	0,023	0	0	0	0,023
♭VII	0	0	0	0,833	0	0	0	0,167	0	0	0	0
VII	0,818	0	0	0	0,091	0	0,030	0,030	0,030	0	0	0

Figure 5: Transition matrix related to K-P *corpus*, produced from the matrix of Fig. 4

Interesting possible usages for the transition matrix, not explored by
Temperley in this work, is to provide a calculation of entropy associated with the

chords in the repertoire. Consider, for example, the eighth row of the matrix, which corresponds to the possible continuations for chord V, as shown in Fig. 6. This means that in the analyzed *corpus*, out of 197 transitions initiated by V, this chord is followed by I in 167 instances, in no case by ♭II, in eight by II, and so on.

	I	♭II	II	III	♭III	IV	♯IV	V	♭VI	VI	♭VII	VII	Σ
V	167	0	8	1	2	4	0	0	7	6	0	2	197

Figure 6: Eighth row of the matrix of Fig. 4

The calculating the entropy of V is preceded by the following instructions:

(1) Form a vector disregarding the zeroed continuations:

$$v = \langle 167, 8, 1, 2, 4, 7, 6, 2 \rangle;$$

(2) By dividing each element of vector v by the total of continuations, 197 in this case, transform its values into probabilities. Thus,

$$v = \langle 0.8477, 0.0406, 0.0051, 0.0102, 0.0203, 0.0355, 0.0305, 0.0102 \rangle.$$

(3) Input v in Eq. (1), which returns entropy $\mathcal{H} = 1.0015$ bits.

The value obtained matches the quite low uncertainty associated with this particular chord, that in almost 85% of the cases is followed by the first degree in the *corpus*. On the other hand, if we apply the same algorithm to another chord with a more even distribution among continuations, say the ♭VI, the entropy is considerably higher, in the case $\mathcal{H} = 2.7343$ bits.

For Temperley, the harmony rules support his findings, since the most recurrent root motions in his analysis are V–I, I–V, ii–V, and I–IV, which match to the general knowledge about these progressions in tonal music. Indeed, this is a quite consistent finding of the statistical analysis, related to transitions between primary triads and functions, however, the lack of a clear-cut distinction of the chords with respect to their qualities (as well as by the presence or absence of extensions, like sixths, sevenths, ninths, and so on) represents a considerable limitation of his study, preventing it from exploring more high levels of chordal relations. Apart from this, Temperley's analytical method is an important reference for the present work, as it will be evidenced in Section 5.

Recently, a study conducted by Mathieu Barthet and collaborators (2014) use modern mining techniques for exploring big database, focusing on harmonic progressions of six musical styles (classical, jazz, blues, rock, folk, and reggae). In contrast with previous statistical studies on musical aspects, this one works with

MUSICA THEORICA   Revista da Associação Brasileira de Teoria e Análise Musical 2022,
v. 7, n. 1, p. 68–111 – Journal of the Brazilian Society for Music Theory
and Analysis @ TeMA 2022 – ISSN 2525-5541

91

large-scale *corpora* (involving about 200,000 progressions) and relies on Music Information Retrieval rather than symbolic representation. Although the authors are not especially interested in expectancy and entropy issues, some of their general goals are quite close of the present study, as to "uncover similarities and discrepancies between sets of musical pieces [… and] exemplar or idiomatic chord progressions directly from empirical data analyses", as we as to extract necessary "information to find the commonalities and specificities of musical styles and composers".

Unlike in Temperley's analysis, Barthet *et al* discriminated the computed chords according to sixteen qualities (comprising also some inversions), which resulted into a dictionary of 192 entries (16 × 12 chromatic roots),[22] and harmonic functions. The analysis identified progressions of different lengths, from two to sixteen chords (2- and 16-gram sequences), in each one of the six musical genres selected. Considering, for example, the case of the 4-gram unities, jazz music has the largest number of distinct sequences (19,820).

4. Probabilistic Description of Harmonic Aspects of Antonio Carlos Jobim's Music

The Brazilian composer Antonio Carlos Jobim (1927–1994) was one of the greatest exponents of Brazilian music, being specially known for his musical contributions to the Bossa Nova style. The strong appeal of Jobim's music among both the general and specialized public (i.e., listeners, performers, and music theoreticians) is certainly related to its harmony, which can be briefly described as a large set of very complex and denser sonorities connected by mainly unexpected, idiosyncratic relations. Surprisingly, this rich harmonic universe lacks a proper in-depth prospection in strictly structural terms, which motivated the pursuing of this research. In a first step, a recently concluded analysis addressed the harmonic preferences of Jobim's music, through an analysis

---

[22] Despite of this expansion, 16 qualities seem a very modest number, if we consider the size of the *corpora* in question (especially if inversions are also counted among them). As it will be presented, Jobim's *corpus* (composed by 146 songs) encompasses 88 distinct chordal qualities.

covering a *corpus* of 145 of his songs,[23] from which the present work is a continuation.

The analysis of Jobim's harmony focused on two basic aspects: (1) the structure of the individual chords, which can be seen as a "semantic" component,[24] and (2) the possible types of relations involving two contiguous chords, or *binary relations*, which corresponds to a "syntactic" component of the investigation (Almada 2022), being this latter aspect the focus of the present work. Essentially, a binary relation refers not to specific chords, but rather on their possible *interactions*, which involves both their chordal qualities and the intervals that separate their roots.

A secondary, although very meaningful, aspect associated with this syntactic analysis addresses the measurement of expectancy for the binary relation's second element, given the first one. This aspect is intimately related to how stylistic preferences are formed (recalling Meyer's pioneering study), which, in turn, is strongly dependent on statistical concepts, especially the entropy and Markov chains. We now discuss this point in more detail.

4.1 Binary harmonic relations and the augmented transition matrix

Formally, a binary harmonic relation (BHR) is expressed as the triple

$$\langle a(ib) \rangle, \tag{6}$$

here variables a and b denote the *antecedent* and *consequent* chord types,[25] respectively, and i represents the melodic interval between the roots of a and b, measured in semitones, that is, $0 \leq i \leq 11$. Therefore, the relation $\langle a(ib) \rangle$ codify the following information: go form chord type a to b, being their roots an interval of i semitones apart.

---

[23] The research team is formed by Carlos Almada (coordinator), Ana Miccolis, Claudia Usai, Eduardo Cabral, Igor Chagas, João Penchel, Max Kühn, and Vinicius Braga. For a description of goals, theoretical framework, and methodological strategies involved, see Almada *et al* (2019).

[24] This approach involves the concept of chordal quality *chord type* (CT). A CT shall be seen as an abstraction of a specific chord, considering not the real component notes and root, but only its internal intervallic structure.

[25] In this research there exist two possible symbolic notations for expressing concrete manifestations of antecedent or consequent chord types: (1) by conventional, alphanumeric chord labeling used in popular harmony ("C7", "Em7.9", etc.); and (2) by *CT genealogic notation* (to be introduced ahead). For simplicity, this article will use preferentially the former notation.

MUSICA THEORICA   Revista da Associação Brasileira de Teoria e Análise Musical 2022,
v. 7, n. 1, p. 68–111 – Journal of the Brazilian Society for Music Theory
and Analysis @ TeMA 2022 – ISSN 2525-5541

93

The analysis of a musical *corpus* allows us to estimate an *augmented transition matrix*[26] regarding the binary harmonic relations, which we will denote by M. More specifically, the rows of M refer to the antecedent element of the binary relation, while the columns are related to the consequent element. Therefore, the element in row j and column k of M is relative frequency of occurrence of the binary relation $\langle a_j(ib)_k \rangle$ within the considered *corpus*, being this value denoted by p_{jk}. This structure is summarized in Fig. 7.

Note that since the elements related to the rows and columns of M are distinct, we do not expect this matrix to be square. Indeed, the matrix estimated with the data from Jobim-*corpus*' analysis has 88 rows and 547 columns, which gives a total of 48,163 cells or possible binary relations. Therefore, it cannot be formally associated with a Markov chain.
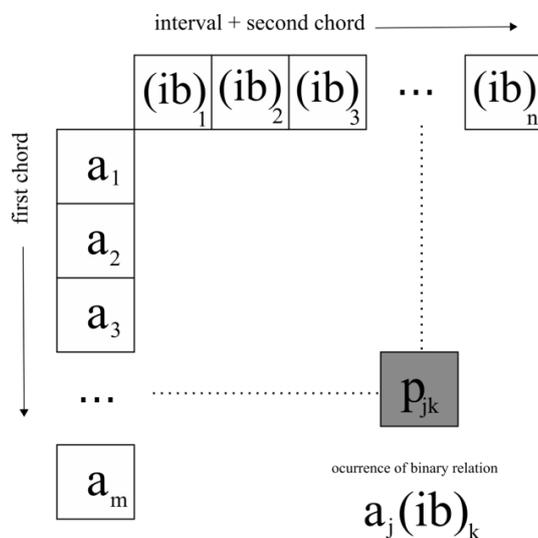
Figure 7: Generic structure of matrix M

4.2 Genera of chord types

Recall that our main object of interest is the harmonic relation between successive pairs of chords, so that the key of the music is of lesser importance. For example, the cadence $\lhd\flat VII \rightarrow V^7 \rightarrow I\rhd$ has the same feeling independently on

---

[26] The matrix is integrated only by situations that really occur in the *corpus*. The adoption of this format prevents the matrix from being extremely large and, especially, reduces computational cost. A detailed description of the data in question and the structure of the transition matrix is addressed in Miccolis *et al* (2021).

which key it is being played. Therefore, a process called *normalization* was employed, which also possess the advantage of reducing the size of matrix ***M***. It is described as follows, when regarding a window of analysis consisting of two consecutive chords (see Fig. 8):

1) Transpose chord *a* such that its root becomes C (pitch class 0), which corresponds to the chord's *prime form*, according to the adopted terminology;

2) Transpose chord *b* by same interval used in step (1);

3) Compute interval *i* between *a* and *b.*

4) Compose the formula *a*(*ib*), which represents the occurrence of that particular binary relation, converting it into the previously defined intervallic ambit ($0 \leq i \leq 11$), if necessary;

5) Re-write the formula using codes for the specific chord types of *a* and *b*, returning the formal notation of the binary relation.[27]

Regarding the latter step, the need for concision and compactness of data for the computational process led to the elaboration of an encoding system for the chord types. The importance of this strategy can be easily perceived when one becomes aware of the total of qualities employed by Jobim in the *corpus*, not more than 88 distinct types. This number becomes still more striking if compared with the universe of possibilities in usage in popular music, namely 184 chord types.[28]

The encoding system consists of the conversion of an ordinary chord label (considering its prime-form version, that is, with root "C") into a string that combines a letter and an order number. According with the Chord-Type Genera Theory, there are ten basic CTs (named *proto-chords*, anyone heading a *genus*), from which derived types are obtained through recursive transformations by application of three sorts of operations (substitution, addition, and alteration). For avoiding confusions with the alphabetic convention for naming roots, the *genera* are identified with the alphabet backwards, adopting uppercase for

---

[27] The encoding process is described ahead.

[28] A detailed discussion about this and other related aspects of this theoretical model, entitled Chord-Type Genera Theory, is presented in a book about Jobim's harmony, written by the Carlos Almada, currently in editorial preparation.

MUSICA THEORICA   Revista da Associação Brasileira de Teoria e Análise Musical 2022,
v. 7, n. 1, p. 68–111 – Journal of the Brazilian Society for Music Theory
and Analysis @ TeMA 2022 – ISSN 2525-5541

95

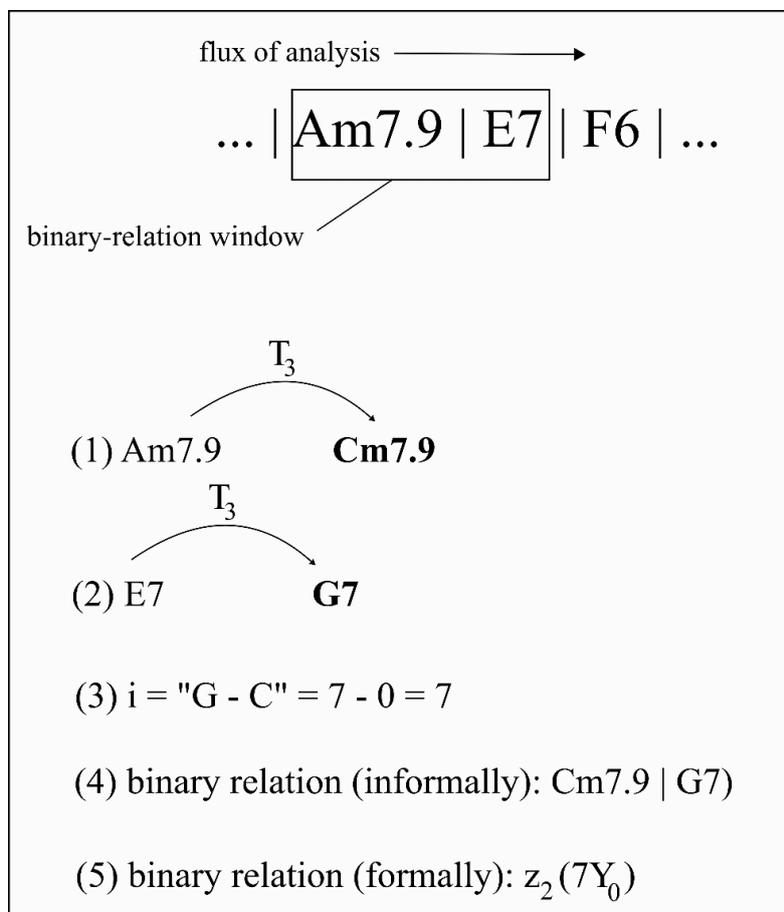naming major-third qualities (Z, Y, X, W, and V), and lowercase for the minor ones (z, y, x, w, and v).

Figure 8: Algorithm for encoding binary relations

The basic information related to the encoding system for the chord types is summarized in Table 2. Observe that "zero" subscripts added to the letters denote them as proto-chords, which shall be seen as both representative of the respective CT genera, and potential *generators* of chord-type variants. Put another way, CM7, encoded as Z_0, is the *proto-chord* that represents genus Z, containing all chord types of "major-with-major-seventh" quality, and can be taken as basis for the production of variants, such as C6 (Z_1), CM7.9 (Z_2), CM7.9(♯11) ($Z_{2.2}$), and so on, by the action of transformational operators.

From the results obtained in the analysis, several derived approaches were initiated or just idealized, basically subdivided into analytical and compositional

applications.[29] Among these, a very special point of interest in the research concerns a stylistic investigation about Jobim's preferences concerning harmonic construction. The next section describes how the notions of entropy and chordal expectations can be related to the findings of the analysis.

Chord label (prime form)[30]	Pitch-class content	Code
CM7	{0,**4**,7,11}	Z_0
C7	{0,**4**,7,10}	Y_0
C(♭5)7	{0,**4**,6,10}	X_0
C(♯5)7	{0,**4**,8,10}	W_0
C	{0,**4**,7}	V_0
Cm7	{0,**3**,7,10}	z_0
Cø	{0,**3**,6,10}	y_0
C°7	{0,**3**,6,9}	x_0
Cm(M7)	{0,**3**,7,11}	w_0
Cm	{0,**3**,7}	v_0

Table 2: List of the ten proto-chords, considering prime-form label (root C), pitch-class content (**4** refers to major third, **3** to minor), and their respective genealogical codes adopted in the system

5. Entropy and Expectation in Jobim's Harmony

As depicted in Fig. 7, any cell of the matrix **M** informs the estimated probability of occurrence of the respective binary relation in the *corpus*. Consider, for example, the seventh row of M (referred to chord type "C6" or Z_1),[31] as shown in Fig. 9. Note how some cells depict zero occurrences (actually, zeroed cells represent most of the cases, regardless of the chosen row), meaning that such alternatives are entirely absent in the repertoire.[32] It shall be recalled that we are

[29] For a study on composition of original harmonic sequences based on statistical modeling of Jobim's practice, see Miccolis *et al* (2021).

[30] The conventions for labelling chords adopted in this system are described in the book mentioned in footnote 28.

[31] As it can be observed in Fig. 9, the rows and columns of **M** are *not* ordered.

[32] It is a quite relevant fact for the understanding of Jobim's harmonic preferences that only 2,695 cells of **M** have non-zero values (4.8% of the total of possible relations considering the 88 chord types).

MUSICA THEORICA Revista da Associação Brasileira de Teoria e Análise Musical 2022, v. 7, n. 1, p. 68–111 – Journal of the Brazilian Society for Music Theory and Analysis @ TeMA 2022 – ISSN 2525-5541

97

dealing here with prime forms, rather than concrete chord labels. Take, for example, the binary relation #76, "C6→A7.13" (or, in the formal notation of the binary relations, $Z_1(9Y_4)$. This means that not only this special progression, but also all twelve transpositions of it[33] (i.e., "C#6→A#7.13", "D6→B7.13", …, "B6→G#7.13") are absent in the *corpus*.
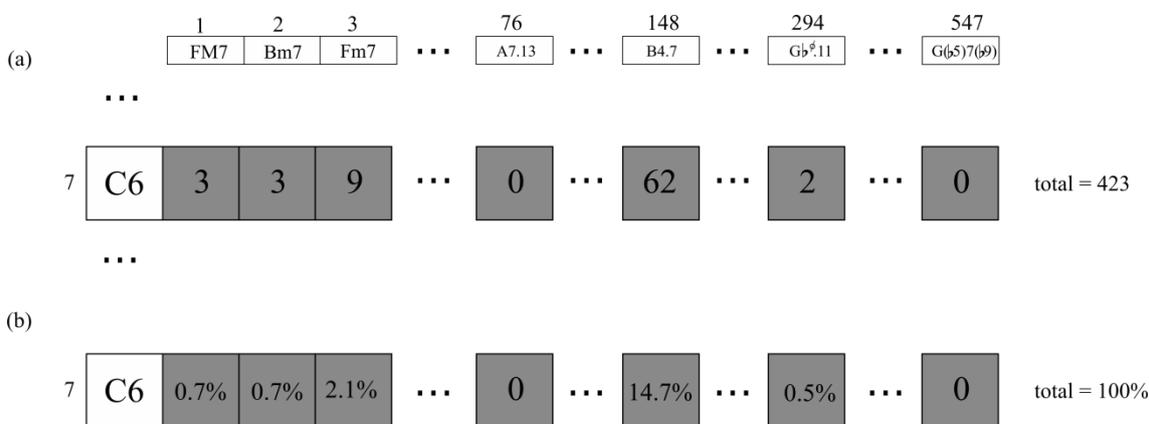
Figure 9: Occurrences expressed in *M* of binary relations initiated by chord type Z_1 ("C6") considering the Jobim's *corpus*: (a) counting of transitions; (b) relative frequency

Similarly to what was experimented with the transition matrix of the K-P *corpus* in Temperley's study, *M* can be seen as a source of data for the calculation of the entropy associated with each one of the 88 chord types detected in the analysis (and listed in *M*'s rows). As discussed in the first section of this article, the entropy (or, equivalent to say, the measure of uncertainty) of a given event depends on the number of possible immediate continuations for it, as well as on the distribution and magnitude of the probabilities of these continuations.

Consider once again "C6" as an illustration of this process. By plugging the values on the row of *M* corresponding to "C6" into Eq. (1), we obtain an estimated entropy of $\mathcal{H} = 5.913$. By considering that only 117 entries of this line are non-null, the maximum value attained by the entropy would be 6.870, if the distribution were uniform over these values.[34] Therefore, the estimated entropy indicates a quite high uncertainty associated with chord type "C6" in Jobim's *corpus*.

---

[33] Considering enharmonic equivalence for roots.

[34] Since the other 430 transitions were not observed on the analyzed *corpus*, it is not reasonable to include them even in hypothetical scenarios to compare the estimated entropy.

The estimated transition probabilities also provide a visual identification of the possible continuations for a selected chord type. This kind of representation is exemplified in Fig. 10, by selecting chord type "C6" again for exam. Aiming at a faster visual understanding, the edges connecting the central chord type to its possible continuations are proportionally adjusted to the respective estimated transition probabilities, being disposed in a spiral-like trajectory (for clarity, only the twelve more frequent transitions are depicted). The procedure can evidently be replicated for any chord type of the *corpus*.
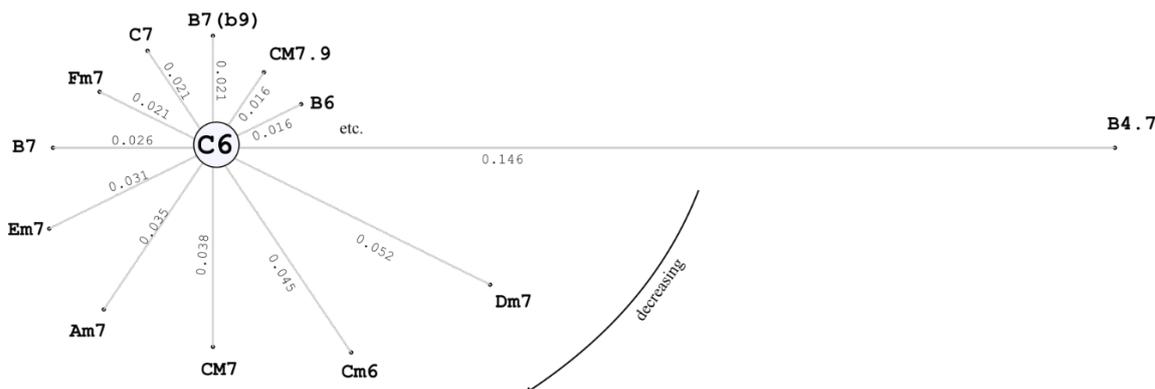
Figure 10: The twelve most probable continuations for chord type "C6" considering the Jobim's *corpus*

The scheme reveals, for example, that in this particular case (considering, of course, the repertoire analyzed), the most recurrent continuation ("B4.7") is almost three times more probable than the second alternative ("Dm7").

Fig. 11 compares the entropy and the number of possible continuations of the most populous genus (thirty members) Y, the class of dominant-seventh chord types. The central axis (dashed line) separates the number of possible continuations (left) and the entropy (measured in bits, right) of the respective member of genus Y (along the lines). Observe that, although there seems to be a positive correlation between these two quantities (that is, the greater the number of continuations, the greater tends to be the entropy), this is not always true. Indeed, entropy tends to increase when the diversity of continuations also increases, but the distribution of the probabilities also plays an important role. This becomes very clear if we compare, for example, types "C4.7" and "C7(♭9. ♯11)", framed in Fig. 11. Despite the discrepant numbers of possible continuations (respectively, 133 and 11), their entropies are considerably similar, 3.311 and 3.279 bits, respectively.
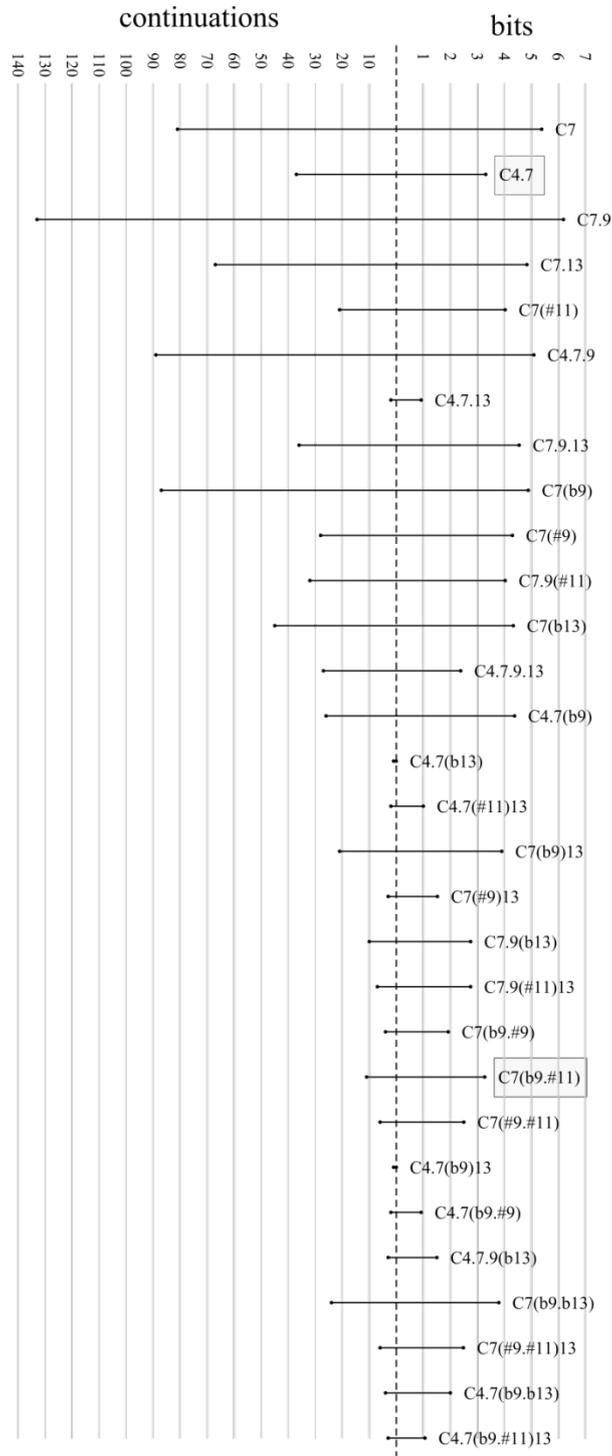
MUSICA THEORICA Revista da Associação Brasileira de Teoria e Análise Musical 2022, v. 7, n. 1, p. 68–111 – Journal of the Brazilian Society for Music Theory and Analysis @ TeMA 2022 – ISSN 2525-5541

99

Figure 11: Entropy (measured in bits) and number of possible continuations of chord types member of quality class Y ("dominant seventh"), considering the Jobim's *corpus*

This apparently controversial situation is easily explained through the exam of the probability vectors of the two chord types, as shown in Table 3. For a better visualization, the entries of the two vectors were ordered from the largest to the smallest value. If we focus on both probability distributions, it is possible to perceive a much more intense asymmetry in the case of "C4.7", since the first continuation (C7) concentrates almost half of the total probability mass. On the other hand, the probabilities of the eleven continuations of "C7(♭9.♯11)" are more evenly distributed, which contributes to elevate the uncertainty of this chord type, in spite of the relatively low number of alternatives for connection it has, in the comparison with the previous case. Indeed, the maximum entropy on this second case is 3.459, where the probability distribution is uniform over the eleven transition possibilities, close to the estimated value of 3.279.[35]

continuations	C4.7		C7(♭9.♯11)	
	chord-type	p	chord-type	p
1st	C7	0.4340	Dm7	0.1906
2nd	F	0.1319	F4.7.9	0.1429
3rd	C7(♭9)	0.0833	C4.7	0.1429
4th	F4.7.9.13	0.0660	C7.9	0.0952
5th	FM7	0.0347	C	0.0952
6th	FM7.9	0.0208	Em(M7)	0.0952
7th	C(♯5)7	0.0208	FM7	0.0476
8th	F6	0.0174	Gm7	0.0476
9th	B♭M7	0.0174	C7	0.0476
10th	E♭M7.9	0.0174	Cm7.9	0.0476
11th	*(plus 123 alternatives)*		F♯6	0.0476
		1.0000		1.000

Table 3: Comparison between chord types "C4.7" and "C7(♭9.♯11)" concerning the probabilities of their possible continuations, considering the Jobim's *corpus*

---

[35] Notice that the approach here presented does not consider functional relations between consecutive chords. However, functionality is slightly embedded in the assembling of the Jobim *corpus* since some instances of chords needed to be reinterpreted regarding their functional context.

6. A Probabilistic Harmonic Space

Motivated by Meyer's reflections about the formation of the style of a composer, as resulting from choices facing a set of possible alternatives within a determined cultural context, we propose to use the notions of entropy and probability in an analytical model destined to evidence how idiosyncrasies (in different degrees of intensity) of harmonic paths can in some way be linked to stylistic characterization.

Firstly, we define a structure called the *probabilistic harmonic space* as a conceptual construct that congregates the repertoire of all harmonic choices made by a given composer in respect to a representative *corpus* of his/her works, considering both material (chord types) and relations (the manner which the material elements are connected to each other), forming a particular subset of the universe of harmonic possibilities. Mathematically, a probabilistic harmonic space can be described by the pair $(\mathcal{S}, \mathcal{P})$, where $\mathcal{S}$ is a set of elements used to describe the harmonic objects employed by the composer, and $\mathcal{P}$ is a stochastic process on the set $\mathcal{S}$. This is a very general definition, and in this particular context, the data conveyed by matrix **M** together with the set of all observed binary relations is a possible description to Jobim's probabilistic harmonic space (evidently, concerning the *corpus* analyzed). This section aims to present this space in a more formal approach, adopting some of graph-theory tools.

Fig. 12 introduces the most basic element that integrates this space, corresponding to a generic chord type, in a neuron-like representation (it can also be alternatively referred as *unit* or *node*). A unit is visually identified by three complementary information: label (at the center),[36] entropy (in bits, at the top), and number of possible continuations (bottom). The diamond geometry of a node is especially designed to facilitate the understanding of the relations between the chord types. The leftmost vertex (marked by a black circle) is reserved as input slot, used for connection with the previous unit. The remaining vertices are used to connect the node with three of its continuations,[37] selected according to the

---

[36] Although only traditional chord labels are in the examples of this article, alternatively genealogical symbols (Z_2, $w_{3.1}$, etc.) can also be employed.

[37] The number "three" was chosen as the optimal alternative since, normally, the three first continuations of a given chord type concentrate more than half of the total probability. The increase of the number of connections would provoke exponential growth of the network (see ahead), which would turn unfeasible any analysis.

magnitude of the probabilities. For convention, the order of preference for the linkages is indicated by the type of line of the respective arrows: filled (for the highest probability), dashed (for the second place), and dotted (for the third).[38]
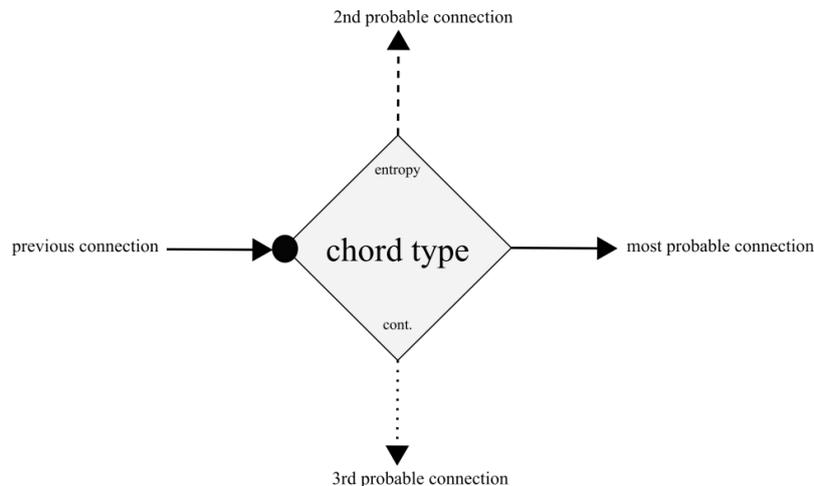

Figure 12: Generic model of a node inhabitant of a probabilistic space

We define a *unit-seed* as the chord type that launches a harmonic sequence (or path) mapped in the space, graphically denoted by a black circle in the input slot with a "X" inside (symbolizing a closed entry). Evidently, a harmonic path can be constructed with different lengths, from two to an undefined number of units. In the terms of the Information Theory (see the introduction), this means that sequences can be formed as digrams, trigrams, and so on. Nodes will be then connected to each other forming *probabilistic harmonic networks* (or PHN, for short), visually representing a very tiny portion of the whole space.

Let us start with the simplest instance of PHN, a digram, selecting, for example, chord type "C6" as unit-seed (Fig. 13). We can read this graph as a kind of map that offers three possible routes departing from a starting point ("C6"), in order of preference, established by the magnitude of probabilities (or weights),[39] leading to: (1) a suspended-fourth-dominant chord type one minor second lower; (2) a minor-with-seventh chord type a major second higher; and (3) a minor-with-sixth chord type with same root.

---

[38] Eventually, the dotted-line arrow can also refer to options beyond the third.

[39] Unlike customary paths in graph theory, here the highest weight corresponds to the simplest (and, therefore the best) path.
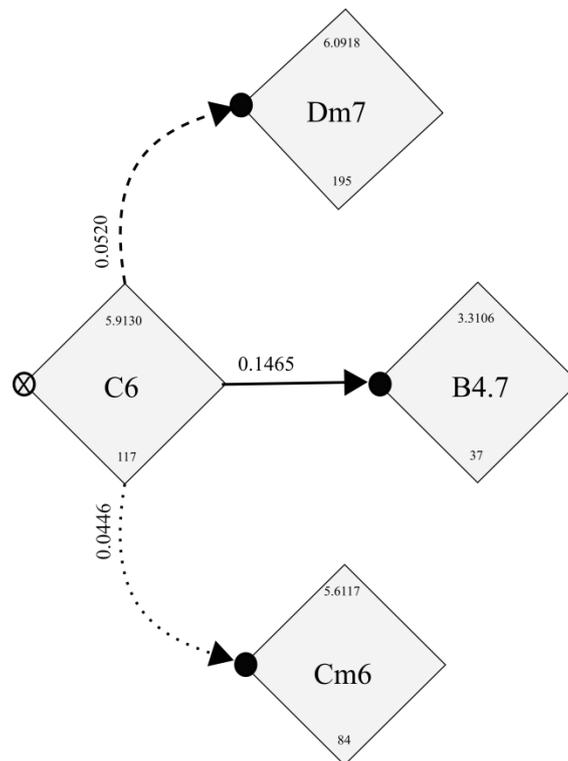
MUSICA THEORICA   Revista da Associação Brasileira de Teoria e Análise Musical 2022, v. 7, n. 1, p. 68–111 – Journal of the Brazilian Society for Music Theory and Analysis @ TeMA 2022 – ISSN 2525-5541

103

Figure 13: Three most probable paths of chord type "C6" considering the Jobim's *corpus*

Complexity highly increases with the inclusion of just one layer to a digram PHN (and, therefore, forming a trigram), since the three probable continuations of anyone of the three first branches are also added to the network. Fig. 14 illustrates this situation by electing another chord type, "Cm7", for unit-seed. As it can be observed, there exist nine (3 × 3) possible harmonic paths in this very simple PHN.[40] For the sake of clarity, the path formed by the first-default choices in each layer is indicated by the "activation" of the corresponding nodes (which is represented by the thicker frames).

---

[40] Generically, the number of harmonic paths (h) in a n-gram PHN is obtained by the formula $h = 3^{n-1}$. Given this, a sequence of eight elements (8-gram), for example, would be represented by a probabilistic network with not less than 2,907,152 possible paths.
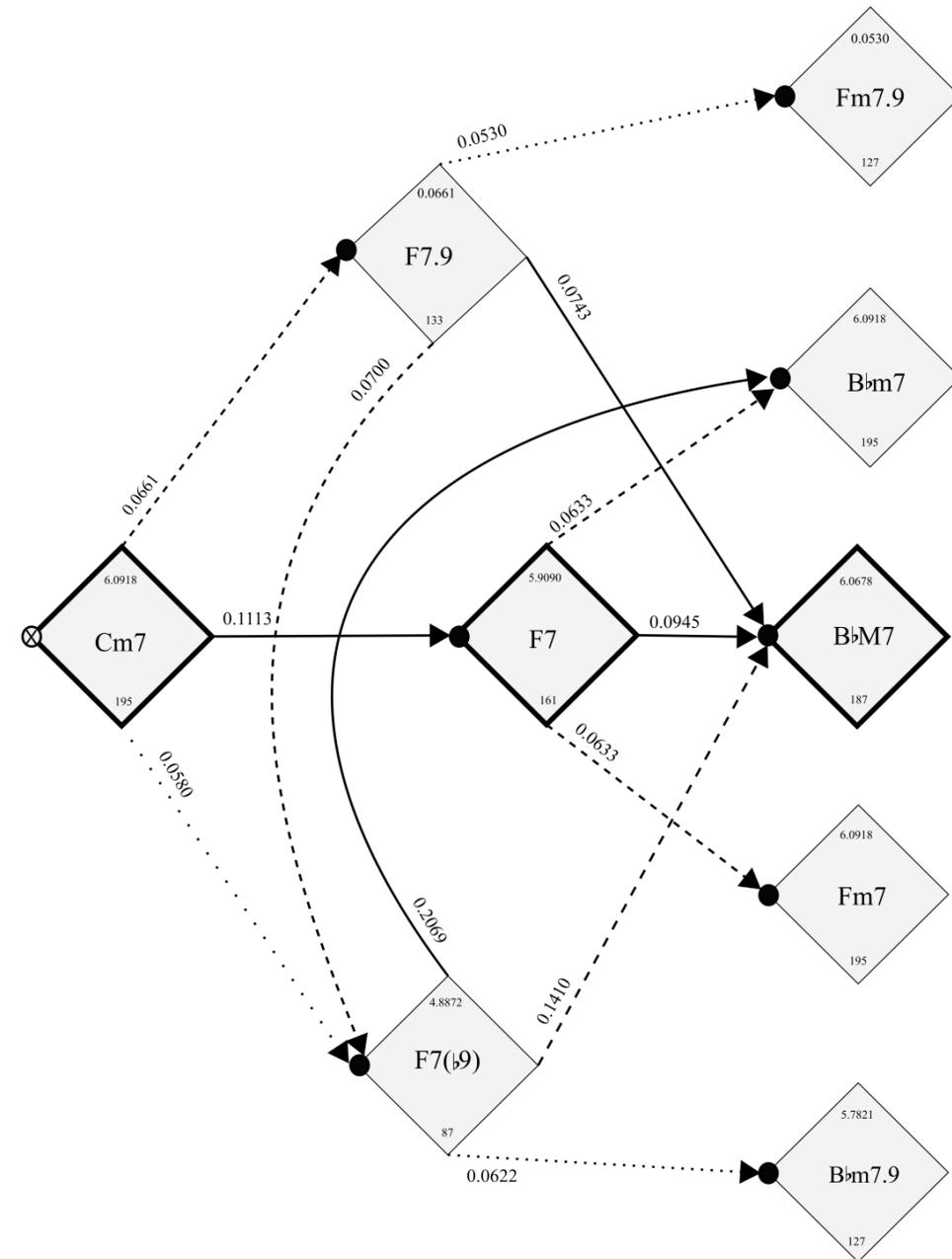
Figure 14: Possible 3-gram paths departing from seed "Cm7" (considering the Jobim's *corpus*)

With the purpose for elaborating a strategy for quantification of the paths within probabilistic terms, we propose the creation of a *probabilistic index*, which is calculated by the following formula

$$q = -log_2 \prod_{k=1}^{K} p_k, \tag{7}$$

MUSICA THEORICA Revista da Associação Brasileira de Teoria e Análise Musical 2022, v. 7, n. 1, p. 68–111 – Journal of the Brazilian Society for Music Theory and Analysis @ TeMA 2022 – ISSN 2525-5541

105

$$q = \frac{1}{K} log_2 \prod_{K=1}^{K} p_k \qquad (7)$$

where q is the probabilistic index of a given path, K is the path's length (i.e., the number of relations in question, or of edges travelled along the path), and p_k is the probability of the binary relations considered. Likewise in the calculation of the entropy, the adoption of the logarithmic function is a necessity, considering the extremely low results obtained in multiplications of probabilities. Notice that it is only reasonable to compare probabilistic indexes for paths of the same length. A possible way to circumvent this issue is to define q as an average, by dividing Eq. (7) by K, but since in this work only paths of the same length are being compared, this definition was not adopted, for simplicity.

Another measure aiming systematization of the analytical process concerns a precise and simple method for identification of the paths. They will be expressed as vectors formed by the sequence of the numbers that represent the order of choice at each layer ("0" indicates the unit-seed).[41] Table 4 lists the information conveyed by the PHD of Fig. 14, including the new elements, path notation and index q. It reveals that the simplest path (i.e., which has the highest value for q) does not correspond to the first-default path 0.1.1 ("Cm7-F7-B♭M7"), as it would be expected, but rather to path 0.3.1 ("Cm7-F7(♭9)-B♭m7"), which is due to the highest probability in the network (0.2069) that occurs with the last linkage.

seed	p_1	path	Layer 1	p_2	path	Layer 2	q
Cm7	**0.1113**	0.1	F7	0.0945	0.1.1	B♭M7	6.571
				0.0630	0.1.2	B♭m7	7.156
				0.0630	0.1.3	Fm7	7.156
	0.0661	0.2	F7.9	0.0743	0.2.1	B♭M7	7.783
				0.0700	0.2.2	F7(♭9)	7.869
				0.0530	0.2.3	Fm7.9	8.271
	0.0580	0.3	F7(♭9)	**0.2069**	0.3.1	B♭m7	**6.381**
				0.1410	0.3.2	B♭M7	6.934
				0.062	0.3.3	B♭m7.9	8.119

Table 4: Calculation of index q for paths depicted in Fig. 14

Let us take it a step further and consider a lengthier sequence, under a different standpoint. This time, instead of exploring possibilities of connections from a selected seed in the probabilistic harmonic space, we will use the structure

---

[41] For avoiding eventual confusions (when paths involve numbers greater than 9), the entries of the vector are always separated by dots.

of the model for comparing functionally equivalent harmonic progressions by their respective probabilistic properties. Fig. 15 depicts two 5-grams that, despite describing quite different trajectories, have the same general syntactic goal, namely, to prepare the arrival of B♭M7, supposedly a stable harmonic point in a hypothetical piece. For clarity, the networks were summarized only by the chosen alternatives. While the probabilistic path in (a) is constructed by the first- and second-default choices at each station, the sequence (b) adopts in all but one case very low-rated options. Consequently, the probabilistic indices of the two cases diverge by a factor of two: $q_a = 10.807$ and $q_b = 22.677$.
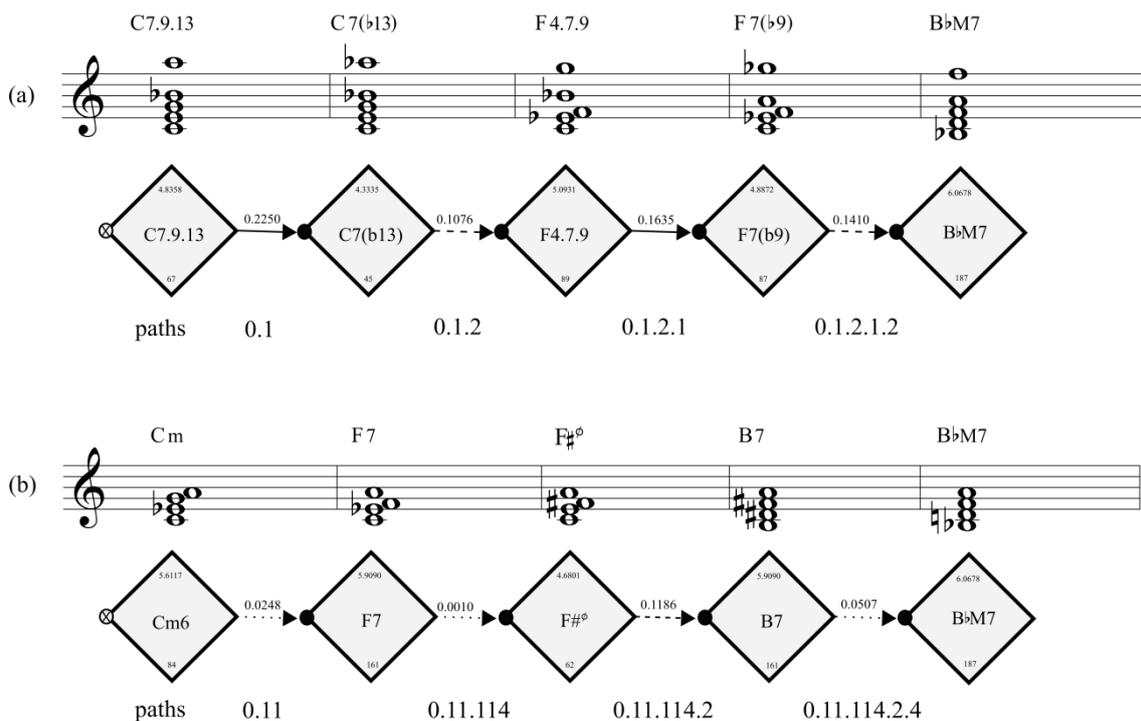
Figure 15: Two alternative 5-gram paths arriving at chord "B♭M7" (considering the Jobim's *corpus*)

Another attractive possibility is to evaluate graphically how the *corpus* "behaves" with respect to selected harmonic formulas. This model plots the sequential elements (on the horizontal axis) according to accumulated value of q (vertical axis). Therefore, the greater the slope of the edge that departs from a given chord (evidenced by the formed angle), the lower will be the probability of the following chord (or, alternatively, the more surprising it will be in face of other possible continuations).

MUSICA THEORICA   Revista da Associação Brasileira de Teoria e Análise Musical 2022, v. 7, n. 1, p. 68–111 – Journal of the Brazilian Society for Music Theory and Analysis @ TeMA 2022 – ISSN 2525-5541

107

This technology can therefore be applied for the exam of specific progressions in the repertoire, leading to a more profound investigation on stylistic issues. Fig. 16 compares *hexagrams* (i.e., harmonic progressions formed by six chords), referred to four well-known of Jobim's songs: *Desafinado*, *Ela é carioca*, *Insensatez*, and *Triste*.[42] All progressions correspond to the first six chords of each song. Functioning as a control for the comparison, the graph also includes a six-chord enriched cadence, a relatively common formula shared by composers of same context. The sequences are depicted in Table 5.
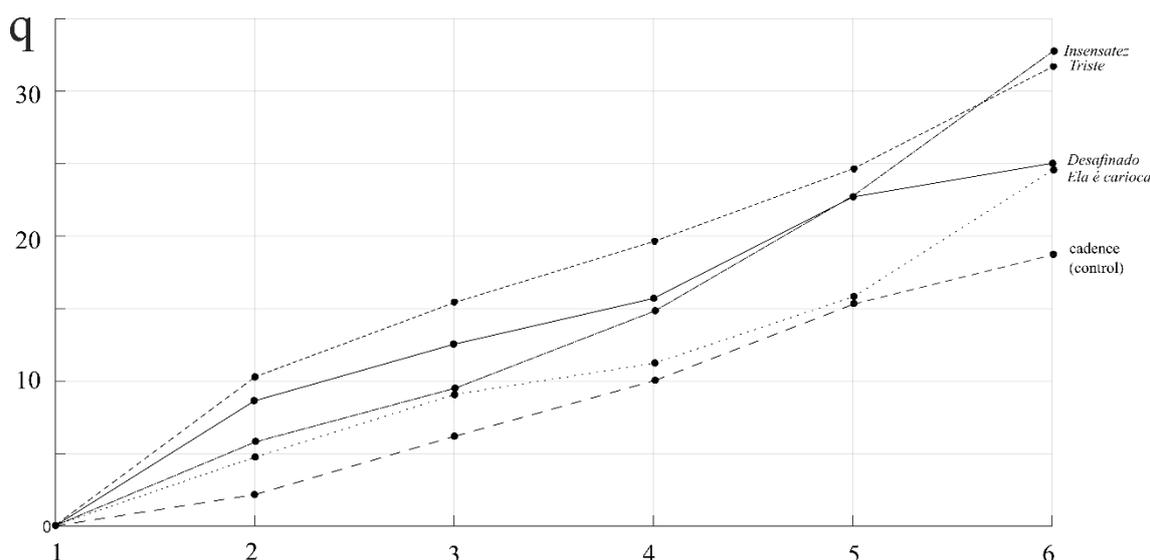

Figure 16: Comparison of hexagrams of the starting measures of four Jobim's songs: *Ela é carioca*, *Desafinado*, *Triste*, and *Insensatez*. An idiomatic cadence plays the role of control

	1	2	3	4	5	6
Ela é carioca	CM7.9	Am7	D7.13	D7(♭13)	Dm7	B♭m6
Desafinado	CM7	D7(♯11)	Dm7	G7	E$^\varnothing$	A7(♭9)
Triste	CM7	A♭M7(♯11)	CM7	C6	Em7	A7(♭13)
Insensatez	Cm7	B°7	B♭m6	E♭7	F7	A♭6
Control	C7.13	C7(♭13)	F7.9	F7(♭9)	B♭7	E♭M7

Table 5: The hexagrams considered in Fig. 16; all initial chords were transposed to C

---

[42] *Insensatez* and *Ela é carioca* are co-authored by Vinicius de Moraes. *Desafinado* is co-authored by Newton Mendonça.

Maybe the most interesting aspect in this view concerns the fluctuations of probabilities in the change of chords, depicted graphically by the slopes of the respective edges. If we compare, for example, the initial transitions of the five trajectories, we observe that *Triste* has the most unexpected beginning (a mediantic chromatic relation $Z_0(8Z_3)$), while in the control-cadence it corresponds to a very common internal flattening of the tension 13 in a dominant chord ($Y_4(0Y_{4.1})$). By contrast, in the transition between chords 3 and 4, the control solution (F7.9 | F7($\flat$9)), again an internal voice-leading adjustment in a dominant chord, is comparatively *more* expected than the two-five formula Dm7 | G7 that occurs at the same point in *Desafinado*. In sum, entropy fluctuates in different rates in any case, but the final values for q at the matching point (the sixth chord, in this example) establish an average ranking that allows us to make comparisons. Thus, we can say, for example, that *Insensatez* – among this restricted group of songs – has *in average* and considering the sequence of the initial six chords the highest entropy. We could even venture to say that this song is more idiosyncratic than its exemplified counterparts (again, considering only the six fix chords), which suggest that the q scale could be roughly seen as a possible measure for style. Higher q values (when compared with compatible ones) may indicate that a song (or a selected passage) has a more *personal* harmonic construction, or near of a "dialectal" usage, using a linguistic analogy. Conversely, low values for q could be associated with a shared practice (as the cadential formula of our example), having a more "idiomatic" nature.

7. Concluding remarks and future work

This article explored a new and very promising field of investigation on the realm of popular-music harmony.[43] Based on the studies of Shannon, Meyer, Huron, Temperley, among others, the strong correlations between probability, expectancy, and entropy formed a solid framework from which the ideas of the Probabilistic Harmonic Space and the probabilistic index emerged. The application of these concepts on the data obtained from the statistical analysis of

---

[43] Evidently, we are considering the context inhabited by Jobim and his counterparts. Basically, this universe is characterized by the use of complex, dense chords (normally enriched with one to three tensions) connected by expanded or diluted functional relations, frequently involving remote tonal regions.

Jobim's *corpus*, mediated by the augmented transition matrix (whose mathematical formalization will be the subject of a future article), provided the necessary means for qualifying and quantifying harmonic progressions through an entirely original bias, namely, as paths on an infinite space of possibilities.

Although Jobim was chosen as the starting point for the present investigation (and we think that we could not find a better name for this!), the model is plainly generalizable and can be easily applied to other repertoires, since the methodological apparatus needs only to be fed by harmonic sequences (in midi format) to produce the related data (including a transition matrix). The possibility of comparison of distinct *corpora* with respect to the distribution of chord-type entropy is maybe the most obvious continuation for the research, as well as a very exciting perspective.

Finally, and especially, the present study reinforces one of the most important of Huron's and Meyer's formulations, that compositional styles are strongly dependent on statistical learning. In this sense, to compose could be seen not only as the art of making (good consequential) choices among uncountable options, but especially choices that have also a past sedimented in a terrain of other and other good choices. We believe, in sum, that the style of a composer (like Jobim) is slowly formed in this way, carving deeply the paths that will build his/her personal and unique probabilistic space.

References

1. Almada, Carlos. 2022. *A harmonia de Jobim*. Campinas: Editora da Unicamp.

2. Almada, Carlos; Cabral, Eduardo; Miccolis, Ana; Braga, Vinicius; Kühn, Max; Usai, Claudia; Chagas, Igor; Penchel, João. 2019. A Software for Computer-Assisted Analysis of Antônio Carlos Jobim's Songs. In *Symposium of the Brazilian Society of Computer Music*, 17., *Proceedings*…. São João del Rei: UFSJ.

3. Ames, Charles. 1989. The Markov Process as a Compositional Model: A Survey and Tutorial. *Leonardo*, v. 22, n. 2, pp. 175–87.

4. Carvalho, Hugo. 2019. An Introduction to Markov Chains in Music Composition and Analysis. *MusMat: Brazilian Journal of Music and Mathematics*, v.3, n. 2, pp.18–43.

5. Cover, Thomas; Thomas, Joy. 2006. *Elements of Information Theory*. 2nd Edition. Hoboken: Wiley-Interscience.

6. Degroot, Morris; Schervish, Mark. 2012. *Probability and Statistics*. 4th Edition. Boston: Addison-Wesley.

7. Barthet, Mathieu; Plumbley, Mark. D.; Kachkaev, Alexander.; Dykes, Jason; Wolff, Daniel.; Weyde, Tillman. 2014. Big Chord Data Extraction and Mining, In *International Conference on Interdisciplinary Musicology* (CIM). *Proceedings*….

8. Hiller, Lejaren; Fuller, Ramon. 1967. Structure and Information in Webern's Symphonie, Op. 21. *Journal of Music Theory*, v. 11, n. 1, pp. 60–115.

9. Hiller, Lejaren; Bean, Calvert. 1966. Information Theory Analyses of Four Sonata Expositions. *Journal of Music Theory*, v. 10, n. 1, pp. 96–137.

10. Hilmar, Rosemary. 1984. Alban Berg's studies with Schoenberg. *Journal of the Arnold Schoenberg Institute*, v. 8, n. 1, pp. 7–29.

11. Huron, David. 2006. *Sweet Anticipation*: Music and the Psychology of Expectation. Cambridge: The MIT Press.

12. Jaynes, Edwin. 2003. *Probability Theory*: The Logic of Science. Cambridge: Cambridge University Press.

13. Knopoff, Leon; Hutchinson, William. 1983. Entropy as a Measure of Style: The Influence of Sample Length. *Journal of Music Theory*, v. 27, n. 1, pp. 75–97.

14. Loy, Gareth. 2006. *Musimathics:* The Mathematical Foundations of Music (vol. I). Cambridge: The MIT Press.

15. Maia, Lucas. 2016. *Formalismos da Composição Algorítmica:* Um Experimento com Canções Folclóricas Brasileiras (M.Sc. Eng. Elétrica). Rio de Janeiro: UFRJ.

16. Maia, Adolfo; Maia, Igor. 2019. An Information Theory Based Analysis of Ligeti's Musica Ricercata: Movements I and II. *MusMat: Brazilian Journal of Music and Mathematics*, v. 3, n. 2, pp. 72–88.

17. Manzara, Leonard; Witten, Ian; James, Mark. 1992. On the Entropy of Music: An Experiment with Bach Chorale Melodies. *Leonardo Music Journal*, v. 2, n. 1, pp. 81–8.

18. Meyer, Leonard. *Emotion and Meaning in Music*. Chicago: University of Chicago Press, 1956.

19. _____. 1957. Meaning in Music and Information Theory. *The Journal of Aesthetics and Art Criticism*, v.15, n. 4, pp. 412–24.

20. Miccolis, Ana; Usai, Claudia; Cabral, Eduardo; Penchel, João; Chagas, Igor; Kühn, Max; Braga, Vinicius; Almada, Carlos. 2021. Composição algorítmica

de progressões harmônicas ao estilo de Antonio Carlos Jobim através de processos markovianos. *Musica Theorica*, v. 6, n. 1, pp. 218–38.

21. Müller, Meinard. 2015. *Fundamentals of Music Processing*: Audio, Analysis, Algorithms, Applications. New York: Springer.

22. Shanon, Claude. 1948. A Mathematical Theory of Communication. *The Bell System Technical Journal*, v. 27, pp. 379–423.

23. Ross, Sheldon. 2010. *A First Course in Probability*. 8th Edition. Upper Saddle River: Prentice Hall.

24. Temperley, David. 2007. *Music and Probability*. Cambridge: The MIT Press.

25. _____. 2009. A Statistical Analysis of Tonal Harmony. Available at: http://davidtemperley.com/kp-stats/

26. Youngblood, Joseph. 1958. Style as Information. *Journal of Music Theory,* v. 2, n. 1, pp. 24–35.