A GENERATIVE SYSTEM FOR THE CREATION OF NEW SONGS FROM PORTUGUESE PROSODY

Rodrigo Schramm, Helena de Souza Nunes[1], Aurélien Antoine, Eduardo Reck Miranda[2]

[1] *Department of Music, Universidade Federal do Rio Grande do Sul, Brazil*

[2] *ICCMR, University of Plymouth, UK*

Correspondence should be addressed to: rodrigos@caef.ufrgs.br

Abstract: **Depending on the prosodic choices of the reader, using various acoustic parameters to apply emphasis, the meaning of the text may change. In songs, where the text and music work together, the prosody ensures coherence between the intentions of both languages, reducing possible ambiguities. This paper presents a generative system for automatic composition of melodies from lyrics using the prosody of the Portuguese language. This approach is divided into two principal stages. First, the prosody information is extracted from the reading of the lyrics and the captured audio is aligned with the text. The inflection lines from the expressive intonation by the reader associated with the temporal alignment is used to generate a probabilistic set of note transitions. A chain of constraints is applied sequentially in order to define musical scale, harmonic field, prosody and stylistic properties. The second stage focuses on the rhythmical structure. A set of rules based on the Portuguese prosody is used to generate compatible rhythm structures along the text. These two stages generate songs ensuring the correct prosody, which is the main goal of the system. Tests were performed to evaluate this approach with aim to evaluate the capability of the algorithm to generate songs without prosody errors, and also to evaluate how pleasant the songs generated by the machine are, in comparison with songs generated by human composers using the same set of text. The results showed that the proposed system can create pleasant songs while keeping a low number of prosody errors.**

1. INTRODUCTION

This paper introduces a new tool able to generate sets of rhythmic-melodic structures that arise from the expressive reading of a text. These sets of possibilities form the basis for future elaborations and refinements by the composer, sculpting the musical ideas in a song format. The main assumption of this approach is that the generative process performed by an algorithm should be coherent with those obtained from the manual procedures following the CDG methodology [1, 2, 3].

The composition process of a song following this methodology starts with the reading aloud of a selected text. The text is read and recorded using several manners of voice intonation. After, which the composer chooses one option from the whole set of audio recordings. Through the composer's labor, musical structures should emerge from this psychoacoustic phenomenon. Successive reformulations and adjusts are applied concomitantly to the text and the music, transforming and reducing both to a concise musical structure. This final structure is a song composition. It has totality, while keeps the spirit of improvisation, the lightness and pleasant outlook. During the compositional process, the composer should put some considerable effort to produce an extract of "what can be manifested". i.e. the composer must be able to identify the artistic possibilities emerged from the chaos and then organize them. Thus, a tool that brings this organizational process up and reveal artistic ideas, becomes an important part of the creative process.

The CDG compositional approach [3] believes that the intention, communicated by the expressive voice, may be explored as an initial start point in the process of song creation, since the pronunciation does not violate the prosody. In Portuguese, the true intentions from a speaker are revealed in a more eloquent way when using the voice inflections, hand gestures and facial expressions, than by a simple enunciation of a sequence of words. In this language, the sequence of tones (intonation) allows distinct interpretations from a same text phrase, changing its meaning by moving the prosody focus [4]. Pronunciation in different ways implies distinct meaning even on equal syntax of words and phrases. Such a phenomenon becomes more relevant in songs, where, on one hand, the text and the music are undissociated parts and, on the other hand, they can carry their own meaning.

This study gives more details about the rhythmic contribution of the expressive reading. However, aspects connected to intonation and pitch variation are also relevant to compound the rhythmic structure of the Portuguese language [4]. In addition, the pitch variation constitutes the main acoustic correlation to the perception of the intonational structure from sentences [5]. For these reasons, rules related to melody and harmony progressions are also mentioned in this paper. From these two aspects arise prosodic questions.

A panorama built by Mateus [4] about the prosody of Portuguese language refers back to the XVI century. Since then, this subject has included elements as duration, frequency, dynamics, speed and emphasis (stress). These are common elements of the musical language.

Music organizes sounds by given meaning; the prosody organizes the sound *continuum* of a language by using specific pronunciation to communicate ideas and by building characteristic audible patterns. Furthermore, the etymology of the word prosody come from the greek words *pros* (in Latin *ad*) and *odos* (in Latin *cantus*), which conducts to associate text and music, particularly when combined in the format of a song. Music and speech are the essential elements of a song. But the coherence between the lyrics and the rhythm/melody, harmony, style and other formal structures of music has not always been considered essential throughout the song composition in Brazilian music. There are many typical brazilian songs from the folk and popular repertoire that often have no correspondence between the stressed points in the text and the stressed points in the musical phrase. Nevertheless, the authors of the CDG method always have advocated that the consistency between the tonic accents of the text and the accents in the music are important for the music education. Corroborating with this point of view, recent studies about networks in the brain have shown that there is a robust link between music and speech perception and that this link can be mediated by rhythmic cues (time and stress) [6]. The intersection points between music and speech are interdependent aspects, especially from the point of view of the intonation and emotional prosody.

Based on this set of considerations, a composed musical piece will not have internal coherence if there are some divergence between the lyrics and the respective music. As consequence, the understanding from this inconsistency will be at least ambiguous. Thus, in this context, a prosody error refers to this phenomenon, where there are mismatches between prominence points from the expressive reading of the lyrics and the intrinsic prominence points of the correspondent musical structure. These aspects motivated the development of a new tool able to generate a set of potential songs in a draft version from a lyric text in Portuguese language. These versions should not contain prosody errors and serve as the basis to the future labor of the composer.

2. RELATED WORKS

Generative systems have been extensively applied to music composition. Despite the diverseness among several approaches, the most well known algorithms are based on cellular automata [7],

generative grammars [8], genetic algorithms [9, 10] and stochastic models [11]. The approach described in this paper follows the guidelines of a rule based method [12]. But it also contains some characteristics of a stochastic model, for the reason that it implements a set of nondeterministic transitions between musical notes, where the probabilities are adjusted by using constraints based on the harmonic field (chord progression). A good overview about computer models for algorithmic music composition can be found in [7] and [8].

The writing of lyrics and the composition of melodies are linked together in the process of song creation. Hausen et al. [6] showed that melody (intonation) and rhythm (stress and timing) are central elements between the music and the speech prosody. In fact, it is possible to find several works in the computer music field which aim to use the prosody information to generate music.

Miranda developed the *Prose* system [13] for aiding the task of composing melodies from given texts. *Prose* is not an algorithmic composition system, but a tool to extract the prosody information that can be plotted for visual assessment and/or mapped onto musical parameters. The prosody information were also used in automatic generative systems, as developed by Fukayama et al [14]. The proposed system, named by the authors as *Orpheus*, generates melodies from lyrics using prosody of the Japanese language. Their algorithm is divided in two parts, one related to the rhythm and other focused to determine the pitch of each note. *Orpheus* uses dynamic programming to allocate one note for each syllable of the lyrics. The authors introduce a set of rules to cluster groups of notes into segments, given preference to the generation of segments with similar number of syllables. In addition, a probabilistic inference method is used to determine the pitch of each note of the melody. Despite the good results obtained by the system, Orpheus [14] (working on Japanese language) and its extension [15] (working on Mandarin Chinese) are only focused on "pitch accent" languages, differing significantly from the approach presented in this paper, which uses a "pitch and stress accent" language, such as Portuguese.

Another system, named *T-Music* [16], introduces the idea of lyric-note correlation, where the changes between the pitch of the melody notes are correlated with the changes in the pitch from the syllables belonging to the lyrics.

Considering the related work, we designed a new generative system to aid the process of song composition, extracting the Portuguese prosody from the reading of the lyrics and applying an algorithm for melodic generation and rhythmic shaping. In the section 3, we describe the model used in our generative system. Then, results of various experiments performed with the developed approach are presented in section 4. We conclude the paper in section 5 with discussions and remarks on the system and future work.

3. MODEL

This approach focuses on the compliance of the generated songs with the prosodic choices from the reading of the lyrics. The system must extract the acoustic parameters from the respective audio recording and generate a set of plausible musical notes without prosodic errors. This work focuses on the prosody of the Portuguese language and its motivation is based on an empirical method developed for the composition of songs [3]. After a wide analysis of this compositional method, the system architecture was designed into two main stages: (1) extraction of melodic information, and (2) rhythmic structure shaping.

The first stage has the task to generate the pitch of notes in coherence with the lyric phrase intonation. In this case, the fundamental frequency, extracted from the recording of the reading, is used to estimate the pitch variations. It is worth noting that the system does not use the extracted pitch direct to generate the musical notes. Instead, the pitch variation and a chain of constraints are used for modeling a probability distribution over time, ensuring creativeness of the system and also independence of any tonality.

The rhythmic structure of the song is defined on the second stage, where the note values are organized on duple and/or triple subdivisions of the beat time. This stage uses a set of rules based on the stressed syllables position to align the lyrics with each beat time and musical bar, keeping the correct prosody. The Figure 1 shows the algorithm pipeline of the two aforementioned stages, which internal procedures will be explained in details throughout this section.

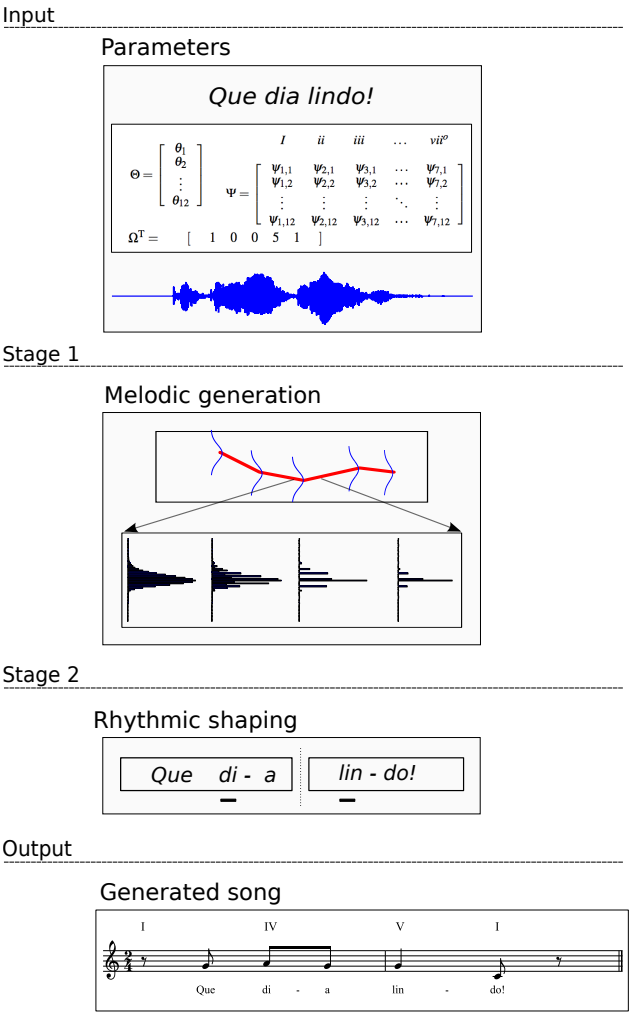

Figure 1: Algorithm pipeline.

3.1. Input Parameters

Some parameters need to be inputted in the model to configure the algorithm. The parameters L (lyrics) and τ (phrase anchor) are mandatory, as they are a precondition to the composition of a song in the context of the CDG [3] method. There are also optional parameters, which in the case of omission, default values are used in the model. The optional parameters are: Θ (musical scale), and $\Lambda = \{\Psi, \Omega\}$ (harmonic field). Details of these parameters will be explained further in the paper. Despite the fact that these parameters are optional, it is interesting to include them into the model because they offer a significant influence on the style.

The parameter L is a sequence of poetic syllables representing the text phrase used as the lyrics of the song. For example, the Portuguese version of the phrase "What a beautiful day!" is written as

Que dia lindo! (1)

and is expressed by the list of syllables (strings) $L = \{$'*Que*', '*di*', '*a*', '*lin*', '*do*!'$\}$.

The parameter $\tau \in \{1, ..., |L|\}$ is the index (position) of the stressed syllable that supports the prosody of the phrase. For instance, in the previous example 1, if $\tau = 2$, then there is an anchor over the syllable '*di*', and the focus of the lyrics is on the meaning of the word *dia* (day); if $\tau = 4$, then the anchor is over the syllable 'lin', changing the expressiveness (and maybe the meaning) of the phrase,

because the focus of the lyrics in now on the meaning of the word *lindo* (beautiful). The anchor τ will always be at a stressed syllable and it is used by the rhythmic shaping algorithm.

In addition to the elements constituents of the rhythm, the model allows for considerable flexibility in regards to the melodic generation by varying the parameter

$$\Theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_{12} \end{bmatrix}, \tag{2}$$

which is a vector containing twelve elements representing one whole octave on the chromatic scale. The variable $\theta_i \in \mathbb{R}$ is in the range $[0,1]$ and means the probability of the respective semitone. Usually, the values are set to the extrema, where $\theta_i = 0$ will disable the ith semitone in the chromatic scale and $\theta_i = 1$ will include it deterministically[1]. The composer can easily specify distinct musical scales just by setting new combinations of values in Θ. For the sake of illustration, the composer could define a C major scale as

$$\Theta^{\mathrm{T}} = \begin{bmatrix} \overset{C}{1} & \overset{C\#}{0} & \overset{D}{1} & \overset{D\#}{0} & \overset{E}{1} & \overset{F}{1} & \overset{F\#}{0} & \overset{G}{1} & \overset{G\#}{0} & \overset{A}{1} & \overset{A\#}{0} & \overset{B}{1} \end{bmatrix}. \tag{3}$$

The last parameter Λ is the harmonic field and it is defined by two variables: Ψ and Ω. Ψ is a matrix which each column is the probability of the twelve semitones (one octave) belonging to a chord over one degree of the musical scale (harmonic field). Each column $\boldsymbol{\psi}_i \in \mathbb{R}$ has its values in the range $[0,1]$ and enables several chord combinations. For example, using the major scale (7 tone chord degrees: $I, ii, iii, IV, V, vi, vii^o$), Ψ has the following form:

$$\Psi = \begin{array}{c} \begin{matrix} I & ii & iii & \dots & vii^o \end{matrix} \\ \begin{bmatrix} \psi_{1,1} & \psi_{2,1} & \psi_{3,1} & \cdots & \psi_{7,1} \\ \psi_{1,2} & \psi_{2,2} & \psi_{3,2} & \cdots & \psi_{7,2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \psi_{1,12} & \psi_{2,12} & \psi_{3,12} & \cdots & \psi_{7,12} \end{bmatrix} \end{array}. \tag{4}$$

In order to simplify the notation, the vector $\boldsymbol{\psi}$ with one integer subscript represents the tone degree. Thus, using the example in 3 (C major scale), the probabilities of semitones belonging to the chord over the I (tonic) and V (dominant) could be expressed, respectively, by

$$\boldsymbol{\psi}_1^{\mathrm{T}} = \begin{bmatrix} \overset{C}{1} & \overset{C\#}{0} & \overset{D}{0} & \overset{D\#}{0} & \overset{E}{1} & \overset{F}{0} & \overset{F\#}{0} & \overset{G}{1} & \overset{G\#}{0} & \overset{A}{0} & \overset{A\#}{0} & \overset{B}{\frac{1}{2}} \end{bmatrix},$$

$$\boldsymbol{\psi}_5^{\mathrm{T}} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & \frac{1}{2} & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

This example uses maximum probabilities for the triad notes and half values for the interval of a third above the fifth of the chord (tetrachord), prioritizing the first three chord notes. The last parameter is Ω, which is a list of scale degrees and has the same cardinality as L. Each value in this list represents a chord constraint and it is related to a syllable in L. Thus, the composer can apply constraints based on a set of chords over the harmonic field. The default value is zero, which means no constraints ($\boldsymbol{\psi}_0 = \mathbb{I}_{12\times 1}$). For sake of illustration, the example 1 could have the following chord progression:

$$\begin{matrix} I & - & - & V & I \end{matrix}$$

$$\begin{array}{lc} L = & \{ \; Que \quad di \quad a \quad lin \quad do! \; \} \\ \Omega^{\mathrm{T}} = & \begin{bmatrix} 1 & 0 & 0 & 5 & 1 \end{bmatrix} \\ & \boldsymbol{\psi}_{\omega_1} \quad \boldsymbol{\psi}_{\omega_0} \quad \boldsymbol{\psi}_{\omega_0} \quad \boldsymbol{\psi}_{\omega_5} \quad \boldsymbol{\psi}_{\omega_1} \end{array}$$

[1] Other representations with different ways of tuning can be applied using the same logic and fewer adaptations on the algorithm.

3.2. Melodic Generation

After setting all input parameters in the system, the algorithm extracts the prosodic information of the audio recording from the reading of the lyrics defined in the parameter L. To achieve this, the system uses the Praat tool [17] (with the plugin EasyAlign [18]) to segment and align the phonetics (poetic syllables) with the lyrics from L. The segmentation and alignment allow the system to use the onset times as well as the duration of each syllable L_k to estimate the respective average pitch P_k. This preprocessing step generates a new list P' with same cardinality of L, denoting the normalized derivatives of the pitch signal, computed using the first order differences. P' is used rather than the absolute pitch because it gives a measure of variability, which is more suitable to the proposed melodic generation. In this approach, the goal is not to map the pitch from the voice to melodic notes over a predefined musical scale. Instead, the main idea is to associate the large pitch variation between two syllables to large probabilities of jumps between two notes in the melody. Thus, the melody will carry the expressiveness from the intonation line recorded by the reader, without imitating the melody reproduced by the voice.

As the pitch differences P' are normalized in the $[0,1]$ range, it is necessary to create a map between this interval and the relative musical scale defined by Θ. The proposed algorithm uses a nondeterministic mapping based on the normal probability density function. The normal distribution was chosen because it is symmetric about its mean. For each syllable L_k, a normal probability density function

$$p_{L_k} \sim \mathcal{N}(\mu_k, \sigma^2) \tag{5}$$

is estimated using the Gaussian function, where the centre of the peak is expressed by $\mu_k = T_c + (P'_k - P'_{k-1})\eta_1$, with $P'_0 = 0$, T_c is the tonal centre, and η_1 is a scale factor. The standard deviation σ is calculated from the entire set P' and controls the width of the "bell". In other words, μ_k relates the maximum probability (centre of the normal "bell") with the pitch variation P'_k and the spreading around the peaks is given by the pitch variance throughout the reading.

The distribution p_{L_k} is continuous in the interval $[-\infty, +\infty]$. The probability density function is then discretized into 87 regular bins (related to the standard midi note numbers from 21 to 108) by

$$pM_{L_k}[m] = \int_{B_m \times \frac{1}{\varepsilon}}^{B_m \times \varepsilon} pL_k(x)dx \tag{6}$$

where $m = 21, ..., 108$ (the midi note index), and B_m is the frequency center in Hz of each bin, defined by $B_m = 2^{\frac{m-69}{12}} \times 440$. The variable $\varepsilon = 2^{\frac{20}{1200}}$ includes the probability of a range around the frequency center, equivalent to 20 cents.

The final discrete probability function pN_k is then estimated using Equation 6 and the probabilities from ψ_k and θ_k. Interpreting these variables as independent probabilities, pN_k is estimated by

$$pN_{L_k}[m] \sim \frac{\Phi(pM_{L_k}[m], \hat{\psi}_{\Omega_k}, \hat{\theta}_k)}{\sum_{b=21}^{108} \Phi(pM_{L_k}[b], \hat{\psi}_{\Omega_k}(b), \hat{\theta}_k(b))}, \tag{7}$$

where

$$\Phi(a_1, a_2, a_3) = a_1 a_2 a_3 + \alpha. \tag{8}$$

Notice that $\hat{\psi}_{\Omega_k}$ and $\hat{\theta}_k$ are extended versions of the ψ_{Ω_k} and θ_k, respectively, such that the semitones are replicated along the adjacent octaves, matching with the 87 bins (midi notes) of pM_{L_k}.

The variable α ensures the final probabilities will not be set to zero. In our experiments we used $\alpha = 0.001$. The denominator in Equation 7 ensures the final probability distribution sums are equal to 1. For each new syllable, a new pitch note is generated drawing one sample following the pN_{L_k} discrete distribution. This approach carries the concept of creativeness, as the system tends to generate distinct melodies even when the same input parameters are used, while also avoids a pure random behaviour applying the chain of

constraints on the probability distribution over time[2]. Finally, it also allows a stylistic influence by the parameters θ_k and ψ_k, previously configured by the composer.

3.3. Rhythmic Structure

This work is motivated by the relevance of song compositions with correct prosody.The system was devised to prioritize simple and straightforward songs without prosodic errors. To achieve this goal, the developed algorithm uses a set of rules to conform the rhythmic structure with the lyrics. The set of rules was defined after an exhaustive analysis on the process of songs composition developed by [3] and aims to find a minimal and sufficient set of constraints.

Algorithm: RHYTHMIC SHAPING

Input:	L	Lyrics sequence
	τ	Phrase anchor index
	b_1	Number of beats per bar
	b_2	Beats division
Output:	Υ	Ordered list with duration times related to the input L

(1) Create two lists A^L and A^R containing the index positions of the stressed syllables at the left and right side of τ, respectively:
$A^L = \{a_k\} | k \in \{1, ..., \tau - 1\}$ and
$A^R = \{a_k\} | k \in \{\tau, ..., |L|\}$, where
$a_k = \begin{cases} k & \text{if } L_k \text{ is a stressed syllable} \\ -1 & \text{otherwise.} \end{cases}$

(2) Create a ordered list with the sizes of the words starting by a stressed syllable:
$\Delta = \{\delta_k = (\pi_{k+1} - \pi_k)\} | \pi_k \in \Pi \text{ and } k \in \{1, ..., |\Pi| - 1\}$,
where $\Pi = \{a_k \in \{A^L, A^R, |L|\} | a_k \geq 0\}$

(3) Create a ordered list Υ such that each element $\upsilon_j^{\delta_k} | \delta_k \in \Delta$ represents a note duration:
$\Upsilon = \{\{\upsilon_1^1, \ldots, \upsilon_{\delta_1}^1\}, \{\upsilon_1^2, \ldots, \upsilon_{\delta_2}^2\}, \ldots, \{\upsilon_1^{\delta_k}, \ldots, \upsilon_{\delta_k}^{\delta_k}\}\}$,
where $\upsilon_j^{\delta_k} = \frac{b_2}{\delta_k} | \forall j$.

(4) Insert a quantity of p_l rest notes with duration b_2 (one beat time) each, at the left side of Υ, such that $|A^L| + p_l \mod b_1 = 0 | p_l \in \mathbb{N}$ and $p_l < b_1$.
Analogy, insert a quantity of p_2 rest notes with duration b_2 each, at the right side of Υ, such that $|A^R| + p_r \mod b_1 = 0 | p_r \in \mathbb{N}$ and $p_r < b_1$.

(5) Any final sequence Υ, representing the musical rhythm structure, is a valid output if it satisfies the constraint:
$(A^L| + p_l \mod b_1) \mod (|A^R| + p_r \mod b_1 = 0) = 0$

The next definition formalizes the rules which allowed the development of an algorithm for semiautomatic song composition using the portuguese prosody.

Definition 1. *A song written in Portuguese language and composed following the method of [3] is a sequence of musical notes attached to the respective lyrics under expressive reading, satisfying the following conditions:*

i Each syllable L_k from the lyrics is related to a musical note.

ii The phrase anchor τ is the first beat time on a music bar.

iii Each stressed syllable represents the beginning of one beat time.

iv The sequence of syllables is broken into subsequences containing only one stressed syllable, which it is always the first syllable of the respective set.

v The total duration of all syllables from a subsequence should be exactly one beat time.

vi Once defined the time signature, rests should be added at beginning or end of the rhythmic phrase to complete the measure.

The rules in Definition 1 are a minimal set which provides a simple and straightforward way to avoid prosody errors. Results have shown that the song compositions generated by the system are a good start point for composers, who can use the initial output as a draft version for their work.

Thus, given the lyrics, the parameter τ, the time signature $\frac{b_1}{b_2}$ and the set of rules from the Definition 1, the rhythmic structure is generated by the algorithm *rhythmic shaping*.

In this algorithm, step (1) breaks the lyrics sequences into two parts, putting all syllables before the anchor point in the set A^L and putting all remaining syllables in the set A^R. This procedure ensures the item *ii* of Definition 1. The second step is related to the item *iv* of Definition 1, and it rearranges the lyrics in words[3] with only one stressed syllable at the beginning. The step (3) generates a preliminary sequence of rhythm notes from the previous set of words, such that the first syllable of each word starts at a new beat time (item *iii* of Definition 1). Furthermore, the note values are subdivided such that all syllables belonging to one word fit inside one beat time. Steps (4) and (5) append rest notes at the beginning and at the end of sequence Υ, ensuring the correct position of the syllables inside the bars.

4. EXPERIMENTS

To realize the experiments, five songs were composed by a musician expert (composer with more than 10 years of experience), following the methodology of [3]. All songs were made using characteristics of metrical rhythm and tonal music. With the lyrics of each of these songs, two new versions were generated by the proposed system. The Figure 2 illustrates the music score of one of these examples.

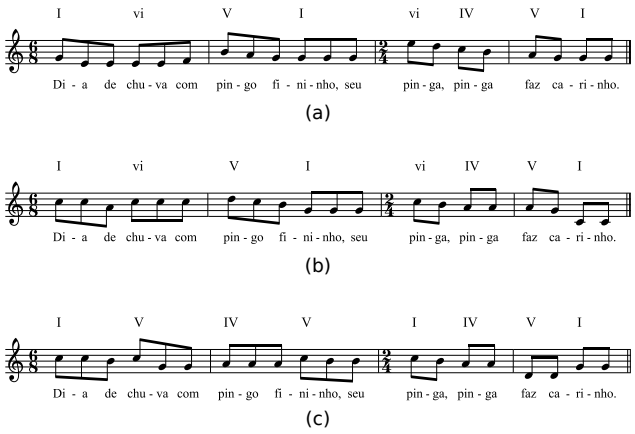

Figure 2: Example of songs composed with the lyrics: "*Dia de chuva com pingo fininho, seu pinga, pinga faz carinho.*": (a) song composed by the musician expert. (b) and (c) versions generated by the proposed algorithm using the same lyrics.

Tests were performed with the proposed approach to evaluate the capability of the algorithm to generate songs without prosody errors. On section 1, it was defined that prosody error is a consequence of the presence of mismatches between prominence points from the expressive reading of the lyrics and the intrinsic

[2]The concept of time is implicit as the sequence of syllables happen along the time in an increasing and monotonic way.

[3]Here, the term "word" means a sequence of syllables which only one of them is stressed, and it does not have any connection with the originals words in the lyric text.

prominence points of the correspondent musical structure. The developed algorithm avoids these mismatches by using a set of rules to align the musical structures with the stressed syllables regarding the lyrics. The test aims to evaluate how robust the algorithm is regarding the prosody errors. No other algorithm that could be used to compare the effectiveness of our approach (benchmarking) has been found. To overcome this issue, an external committee of experts (post graduate students in music) was used to identify possible prosody errors in the generated songs. Thus, for each generated song, the members of this committee answered with yes/no to the question: Is there any prosody error in this song?

The Table 1 shows, for each song, the percentage of votes of yes and no for the aforementioned question. As it can be seen, the number of positive votes in favor of prosody errors in the generated songs is as small as the number of votes in favor of prosody errors in the compositions made by the expert. Also, it is worth noting that in both cases, the number of votes does not mean the number of prosody errors. The percentages in the Table 1 indicate the proportion of votes from distinct evaluators on the entire set of samples in a specific category (human composition or algorithm composition). In fact, the evaluators had no consensus about the presence of the prosody errors. The analysis of the data showed that none of the evaluations over one same song had all votes equal "yes" for prosody error.

Table 1: Percentage of votes considering the prosody errors on the entire set of evaluated samples.

Prosody Errors			
Human Composition		Algorithmic Composition	
yes	no	yes	no
2.85%	97.15%	7.15 %	92.85 %

It was also verified how coherent the system outputs were with regards to the input parameters defined in the experiments. This second aspect of the evaluation process addresses the aim of measuring how pleasant the resulted songs are. This aspect is important because songs that are not in coherence with the tonal music (and the metrical rhythm) are not desired. Again, the same committee of experts were used to evaluate this point. They were asked to answer two questions. First, they had to answer yes/no to the question: Is the song in compliance with the tonal music and the metrical rhythm? Secondly, they had to classify each composition in one of two categories: (1) composed by human; (2) generated by machine.

Table 2: Percentage of votes considering the compliance with tonal music and metrical rhythm on the entire set of evaluated samples.

Compliance with tonal music and metrical rhythm			
Human Composition		Algorithmic Composition	
yes	no	yes	no
80.0%	20.0%	81.43%	18.57%

From the Table 2, it is possible to note that the committee had similar impressions about both set of compositions. Moreover, either of the songs composed by the human or the generated songs by the machine had similar high scores in the yes label, indicating that the songs are in compliance with tonal music and the metrical rhythm. Also, the confusion matrix in Table 3 shows the results from the last question. These values corroborate with the previous analysis, indicating that the evaluators were not able to discriminate the songs generated by machine from the songs written by the human composer.

Table 3: Confusion matrix.

		Output Class		
		Human	Algorithm	Doubtful
Target Class	Human	35.0%	30.0%	35.0%
	Algorithm	38.57%	32.85%	28.58%

5. Conclusion and Future Work

This paper presented a generative system for automatic composition of songs in Portuguese language. The system architecture combines a set of probabilistic constraints with an alignment process to shape the music rhythmic structures in compliance with the prosody from the lyrics. Experiments showed that the generated songs keep a coherent structure as desired and defined by the composer. Regardless of the subjectiveness on the judgment of how pleasant a song can be, the experiments had shown that the output generated by the proposed system are in agreement with the concept of tonal music among several musicians. Despite the wide range of parameters and options, the system follows a natural and easy way to create new songs without prosody error. The described algorithm can be used as a powerful tool to aid the compositional labor, exploring quickly and automatically numerous melody combinations. Further research is planned that will expand the rhythmic possibilities of the actual model, allowing song compositions with more complex structures.

Acknowledgment

This work was supported by grant number BEX-2106/13-2 from the CAPES Foundation, Ministry of Education of Brazil. We also thank Plymouth University for the support and encouragement.

References

[1] H. de Souza Nunes: *A educação musical modalidade EAD nas políticas de formação de professores da Educação Básica.* In *Revista da ABEM*, volume 23:34–39, 2010.

[2] H. de Souza Nunes: *Fundamentos Pedagógicos de um Curso de Licenciatura em Música EAD*. In *ICTUS*, volume 12:1–5, 2011.

[3] H. de Souza Nunes, C. de Godoy Menezes, C. E. dos Santos, J. Leite, L. Nunes, and L. Serafim: *Microcanções CDG: Primeiros Registros.* In *9a Conferencia Latinoamericana y 2a Panamericana de la Sociedad Internacional de Educación Musical, ISME*, pages 641–649. Santiago, 2013.

[4] M. H. Mateus: *Estudando a melodia da fala: traços prosódicos e constituintes prosódicos.* In *Palavras - Revista da Associação de Professores de Português,*, (28):79–98, 2005.

[5] C. A. Gonçalves: *Transcrevendo a entonação*. In *Veredas: Revista de Estudos Lingüísticos.*, volume 3(2):9–19, 1998.

[6] M. Hausen, R. Torppa, V. R. Salmela, M. Vainio, and T. Särkämö: *Music and speech prosody: a common rhythm*. In *Frontiers in psychology*, volume 4, 2013.

[7] E. R. Miranda: *Composing Music with Computers (Music Technology)*. Focal Press, Oxford, 2001.

[8] G. Nierhaus: *Algorithmic Composition: Paradigms of Automated Music Generation*. Springer Publishing Company, Incorporated, 1st. edition, 2008.

[9] M. Marques, V. Oliveira, S. Vieira, and A. Rosa: *Music composition using genetic evolutionary algorithms*. In *Evolutionary Computation, 2000. Proceedings of the 2000 Congress on*, volume 1, pages 714–719 vol.1. 2000.

[10] P. Sheikholharam and M. Teshnehlab: *Music Composition Using Combination of Genetic Algorithms and Kohonen Grammar*. In *Computational Intelligence and Design, 2008. ISCID '08. International Symposium on*, volume 1, pages 255–260. 2008.

[11] W. Schulze and B. van der Merwe: *Music Generation with Markov Models*. In *MultiMedia, IEEE*, volume 18(3):78–85, 2011.

[12] M. Henz, S. Lauer, and D. Zimmermann: *COMPOzE-intention-based music composition through constraint programming*. In *Tools with Artificial Intelligence, 1996.,*

Proceedings Eighth IEEE International Conference on, pages 118–121. 1996.

[13] E. R. Miranda: *Computer-Aided Song Design: Prosody as Scaffolding*. In *VIII Brazilian Symposium on Computer Music*. Fortaleza, 2001.

[14] S. Fukayama, K. Nakatsuma, S. Sako, T. Nishimoto, and S. Sagayama: *Automatic song composition from the lyrics exploiting prosody of the Japanese language*. In *Proc. 7th Sound and Music Computing Conference (SMC)*, pages 299–302. 2010.

[15] S. Qin, S. Fukayama, T. Nishimoto, and S. Sagayama: *Lexical tones learning with automatic music composition system considering prosody of mandarin chinese*. In *Second Language Studies: Acquisition, Learning, Education and Technology*, pages 3–6, 2010.

[16] C. Long, R.-W. Wong, and R. K. W. Sze: *T-Music: A melody composer based on frequent pattern mining*. In *Data Engineering (ICDE), 2013 IEEE 29th International Conference on*, pages 1332–1335. IEEE, 2013.

[17] P. Boersma and D. Weenink: *Praat: doing phonetics by computer [Computer program].*, 2014.

[18] J. P. Goldman: *EasyAlign: An Automatic Phonetic Alignment Tool Under Praat.* In *INTERSPEECH*, pages 3233–3236. ISCA, 2011.