Navarro-Cáceres M., Olarte-Martínez M., Amílcar Cardoso F., Martins P. (2019). "User-Guided System to Generate Spanish Popular Music". En: Novais P. et al. (eds), *Ambient Intelligence – Software and Applications. Advances in Intelligent Systems and Computing* (pp. 24-39), vol 806. New York: Springer, Cham.` ISBN: 978-3-030-01745], https://doi.org/10.1007/978-3-030-01746-0_3

"AbstractThe automatic generation of music is an emerging field of research that has attracted wide attention in Computer Science. Additionally, the interaction between users and machines is nowadays very present in our daily lives, and influences fields such as Economy, Sports or Arts. Following this approach, this work develops an intelligent system that generates melodies based on Spanish popular music and some indications of the users through an interface. The system creates a melody by learning from the corpus selected through a Markov model, which is also influenced by the users' preferences. Several experiments were carried out to evaluate the musical quality and the usefulness of the system to interact with the user and generate music. The results of the evaluation shows that the proposal is able to generate music adapted to the style standards of Spanish popular music and to the users' indications.
Keywords: Computational; Creativity; Melody generation; Popular songs

Paulo Novais, Jason J. Jung, Gabriel Villarrubia González, Antonio Fernández-Caballero, Elena Navarro, Pascual González, Davide Carneiro, António Pinto, Andrew T. Campbell and Dalila Durães (eds.), *Ambient Intelligence – Software and Applications –, 9th International Symposium on Ambient Intelligence*, Advances in Intelligent Systems and Computing 806
https://doi.org/10.1007/978-3-030-01746-0_3

User-Guided System to Generate Spanish Popular Music

María Navarro-Cáceres[1] ✉, Matilde Olarte-Martínez[1], F. Amílcar Cardoso[2] and Pedro Martins[2]

(1) University of Salamanca, Patio de Escuelas, 37001 Salamanca, Spain
(2) CISUC, Department of Informatics Engineering, University of Coimbra, Rua Sílvio Lima, Pólo II da Universidade de Coimbra, 3030 Coimbra, Portugal

✉ **María Navarro-Cáceres**
Email: maria90@usal.es

Abstract
The automatic generation of music is an emerging field of research that has attracted wide attention in Computer Science. Additionally, the interaction between users and machines is nowadays very present in our daily lives, and influences fields such as Economy, Sports or Arts. Following this approach, this work develops an intelligent system that generates melodies based on Spanish popular music and some indications of the users through an interface. The system creates a melody by learning from the corpus selected through a Markov model, which is also influenced by the users' preferences. Several experiments were carried out to evaluate the musical quality and the usefulness of the system to interact with the user and generate music. The results of the evaluation shows that the proposal is able to generate music adapted to the style standards of Spanish popular music and to the users' indications.

Keywords Computational Creativity – Melody generation – Popular songs

1 Introduction

The different computational advances in the field of Artificial Intelligence that have occurred in recent years have attracted the attention of researchers of multiple origins and motivations, creating innovative fields that unite apparently disparate concepts such as Artificial Intelligence and Art. From an interdisciplinary research field that sits at the intersection of the areas of AI, Psychology, Cognitive science, Linguistics, Anthropology and other human-centered sciences, the area of Computational Creativity (CC) was born. CC can be defined as a "philosophy, science and engineering of computational systems which, by taking on particular responsibilities, exhibit behaviors that unbiased observers would deem to be creative" [4].

The area of Computational Creativity has recently been significantly developed with the entry of companies as important as Google, with projects such as DeepDream, which uses a convolutional neural network that transforms images or, more recently, Magenta [3], the equivalent for the generation of music. Although many different techniques have been applied to generate music automatically, statistical methods were a preferable tool due to its ability to learn from the context or the corpus. The deep analysis performed in [11] about different statistical models, is particularly interesting. We should also remark that Markov Models are used very commonly to generate different kinds of music [8–10].

However, to the best of our knowledge, the initiatives have usually focused on the generation of music of a rather tonal character and, to date, there has been no study addressing the generation of Spanish popular music. This genre of music differs from the classical music in many aspects, including the sonority, the sounds disposition or the rhythmic formulas used. The automatic generation of this kind of music also depends on multiple factors that are intrinsically connected, such as the representation of the tonality, the melodies and the rhythm.

Additionally, more and more humans work in collaboration with other humans and/or computational entities, both directly and indirectly, to meet a series of objectives that, individually, would be impossible. Currently, the support of new technologies are key to generating a collaborative organization in the educational or business framework. This new paradigm of virtual collaboration through electronic devices has great potential in areas as diverse as Home automation, Energy saving, Medicine or even Visual Arts or Music.

There have been several approaches in which the interaction of the users is essential for the achievement of music [2, 7, 8, 10]. However, many of them are not focused on users who are not familiar with musical education, making hard for them to generate melody in a semi-automatic way. Drawing on this phenomenon, a user-guided melody generation is proposed, with the goal of bringing the Spanish popular music to those ones without a specific musical education.

For this purpose, new melodies are generated from original Spanish popular songs, extracted from multiple sources of popular music. After the analysis, encoding and storage of relevant features, original melodies are used as a training corpus. Among the different learning models that can be applied to generate music from a previous corpus [1], Markov Models (MMs) have been selected due to their successful application in other related works [9–12]. MMs are trained in a corpus and then used to generate a new melody that fits the style of popular songs. However this melody generation is always guided by the users, who can influence the final result through a hardware device. For the purposes of this work, a mouse is selected as the tool to guide the melody generation. The user interacts in a screen moving the mouse in two axes: horizontal axis to control the rhythm of the melody, and the vertical axis to control the pitches.

Once the melodies are generated, a listening test was developed to evaluate the musical quality according to the Spanish popular music standards and to collect the users' opinion about the usefulness of the system to interact with them and generate music.

The remainder of this paper is structured as follows. Section 2 details the generation process of music and the interaction with the users. Section 3 gives an analysis about the evaluation of the system developed. Finally, Sect. 4 describes the final conclusions and future work.

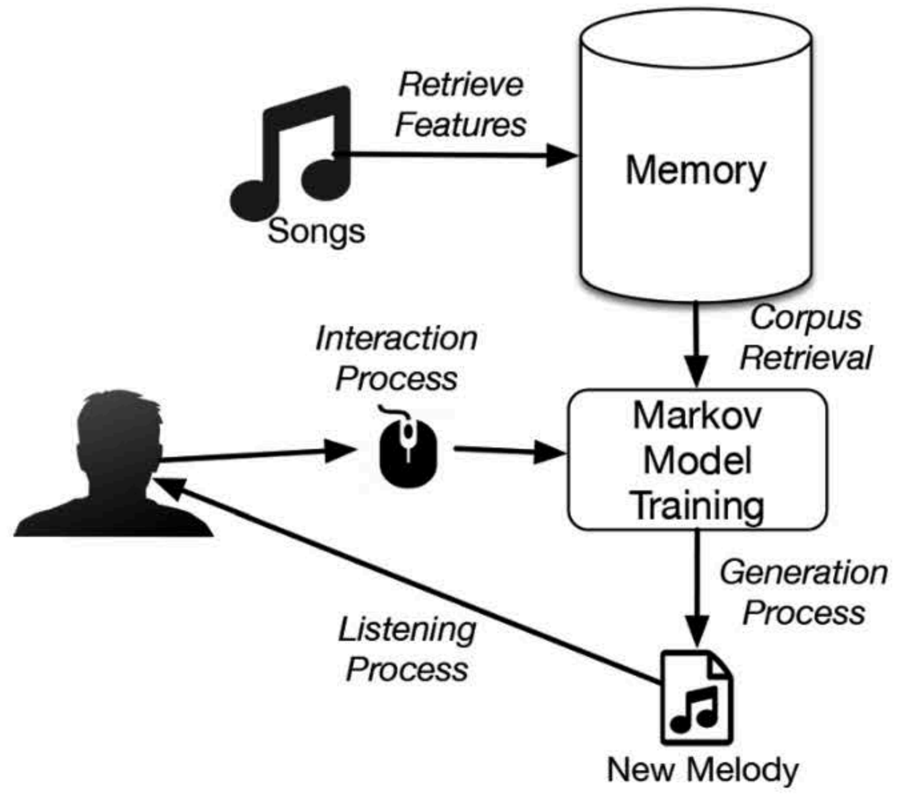

Fig. 1. Overview of music generation flow.

2 System Description

Figure 1 gives an overview of the overall system to create popular Spanish music.

The system retrieves the musical data from popular music songs. Consequently, it is provided with a memory to store different melodies that belong to the popular music style. This corpus has been manually analyzed and retrieved from song books and include a wide variety of popular music.

Initially, the system selects the collected melodies and learns a model for generating a new composition. For the purposes of this paper, Markov Models (MM) were used as the learning algorithm. The Markov Models are influenced by the users' indications through the position of the mouse in the interface of the system.

2.1 Music Generation

The platform intends to generate music based on the features of the popular music, such as the rhythm or the sonority.

In order to generate music according to the popular music (or popular songs) standards, a set of musical sources has been retrieved. The sources should reflect the most common sonorities of the popular music. In order to make the process easier, the songs are extracted from songbooks, recordings or digital scores.

Folk songs in Spain are as varied as its regions. However, generally it is very difficult to understand the origin of the sonority in popular music, since this type of music has been in constant evolution due to its characteristic grammar and its transmission from generation to generation orally. However, analyzing many of the melodies collected over the years by different ethnomusicologists and in different songs, we can draw some conclusions in this regard that allow us to discern the popular music, selecting those ones that better represent the popular musical features according to [6].

The sources should be digitalized to represent the information about rhythm and sonority. To identify the popular music we should not only analyze the particular duration or pitch, but also the duration of the musical phrases, the degrees in which the melody reposes (notes with a long duration), and the particular cadences. Unlike the classical music, in popular music the harmonic tension and the use of the chords degree are not particularly relevant, as it does not follow harmonic rules; they are only used according to the melodic course.

Drawing on these properties and also inspired by the concept of viewpoints exposed in [12], the following features of the popular songs were selected:

– Pitch: Musical note
– Duration: Rhythmic formula of one note
– Degree: Position of the note within the musical scale
– First in bar: Boolean value which indicates if the note is the first in a bar or not.
– Time Signature: Number that represents the time signature of the melody
– Musical Phrase: It represents the position of a note in a musical phrase.

Currently, there is no standard format that only addresses these features. However, there are a wide number of songs already encoded as MIDI (Musical Instrument Digital Interface) files [5], due to its availability throughout the network, the low difficulty in creating such files based on digital scores, and its structure, which allows easy access to notes and durations. The files do not contain the sounds. Instead, they include instructions that allow the reconstruction of the song by using a sequencer and a synthesizer that work with MIDI specifications. Therefore, the files are quite light since they allow to encode a complete song in a few hundred lines. The mathematical data inside these files along with a manual analysis have been used to encode the features above.

Once the files are available, the next step is to extract the necessary information for the project: notes, durations, bars, time signature, etc. These data are considered the training set for the learning model. In the first experiments, we combine different musical features and check which ones works better with the Markov Models. The Markov model determines the possible transitions for each state and the initial probabilities of each one.

Consequently, the position of the mouse must be translated into a note and a duration. To do so, the mouse works as an indicator with two dimensions, and the user can move it through the screen provided. In this sense, the *Y* axis was set to indicate the reference note (higher or lower pitches), since we intuitively associate "climbing" with higher notes and "lowering" with lower notes. Likewise, the *X* axis indicates a reference duration. On the *Y* axis, 0 represents silence (lowest position of the mouse; while 2 reference octaves are represented, from the lowest to the highest. On the *X* axis, 1 represents the shortest note (sixteenth note) and 16 represents the longest note (round). These delimitations of the space of reference facilitate the training process of the model.

In this case, the probability $P(t)$ of each note t_i to be selected as the next note in the melody is a linear combination between the probability $P_M(t_i)$ given by the Markov Model and the position of the mouse $P_D(t_i)$ given by the user:

$$P(t) = k \cdot P_D(t_i) + (1 - k) * P_M(t_i), \tag{1}$$

where k is the weight of each probability and is empirically set to 0.55.

Once we selected the notes of the melody, the music generated by the system will be encoded in MIDI format to use a standard synthesizer that could be incorporated in the own computer. We used a standard piano as the synthesis sound due to its versatility.

3 Results and Discussion

A system which generates popular songs is built. For the generation of the music, 180 popular songs were selected, 85 dance melodies and 95 work songs. All of them make use of similar rhythm patterns, with time signature of 2/4, 3/4 and 6/8. Each song consisted of 3 or 4 musical phrases with similar length, and uses songs with the two different sonorities, meaning we divided the corpus in Frigian mode with possible modifications of this mode in its evolution to E minor and Eolian mode with possible modifications to evolve to A minor.

The melodies have been encoded according to the Sect. 2 and saved in an Excel file with all the properties. These features were the corpus for the Markov Model training. The memory of the MM, meaning the number of estates that it can remember for the future generation of the melody was empirically set to 4.

During the generation process, each iteration of the system consists of adding of a new note in the melody assisted by the mouse position and the MM, and it is iterated until the user decides to stop. Figure 2 shows the interactive screen of the user while he is generating the melody.

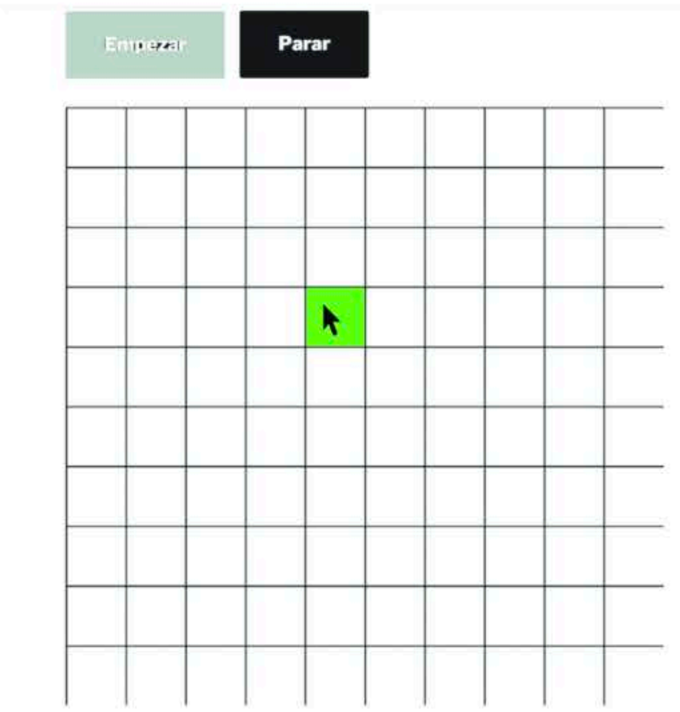

Fig. 2. Interface of music generation system.

The upper buttons indicate to start or stop the generation process. The interaction interface is divided into sections in order to frame the pitches and durations of the musical compositions. The cell where the mouse is over is painted in green to highlight the musical reference in which the system is currently working to generate music and rhythm.

The evaluation of this system is twofold. On the one hand, we aim to validate the musical results, meaning the melodies generated can follow the style of the Spanish popular music. On the other hand, we aim to validate the usefulness and the interactivity of the overall system to generate such kind of music.

In order to assess the quality of the music, a listening test was designed in which 10 melodies generated by the system were chosen. There were 5 melodies of Eolian mode and 5 melodies of Frigian mode. Additionally, 4 melodies have a time signature of 6/8, 3 of 3/4 and 3 of 2/4. 10 expert users in ethnomusicology were selected and asked if they think the melodies generated follows the standards of popular music according to the sonority and the rhythm. They should evaluate the quality in a scale from 1 to 5, where 1 means Very poorly and 5 means Perfectly.

For such categorical data, we use the chi-square test to prove that the results reflect the quality of the melodies according to the subjective ratings given by the listeners. It is important to note that we can get very subjective ratings since all listeners can have different interpretations of the musical pieces and different musical tastes. Thus, we considered the analysis of the median M_e and the mode M_o a useful step.

Table 1. Statistics resulting from the listening test results.

χ^2	M_e	M_o
$6.8381e-04$	Good	Good

Table 1 collects statistical values calculated from the data, namely the p-value for the chi-value χ^2 the median M_e and the mode M_d. The statistical analysis suggests that the system captures the style of the Spanish popular melodies. The majority of listeners think the values are "Good" (according to the M_d) and the median indicates at least half of the ratings obtained are scored as "Good" or even better. The statistical results lead us to conclude that the fitness function captures perceptual musical quality quite well.

The second part of the test consists of validating the usefulness of the system. For this purpose, a total of 8 users were selected to test the system. Each user generated 10 melodies and then were asked for their experience with the system. They answered a questionnaire about whether the system is easy to use, the interface conforms to the real movements of the device, as well as the overall score for the system and possible suggestions. All the questions could be rated from 1 ("Completely disagree") to 5 ("Completely agree"). Table 2 shows the mean scores for all these questions.

Table 2. Final statistics after the users finished testing the system.

	Easy to use	Interface	Control quality	Overall ratings
Mean ratings	4.22 ± 0.86	3.23 ± 1.53	4.09 ± 0.94	4.01 ± 0.79
Mode	4	3	3	3
Median	4	3	3	3

The Table 2 shows that the general satisfaction degree is quite high, with a mean of 4.01 and a mode and median values of 3. The users consider the system to be very easy to use even for people without any musical training (4.22, and a median a mode of 4), although the interface could be improved (3.23, with mode and median of 3). Some users have suggested the addition of a complete score with notes and rhythms instead of only listening to the final sounds and see the score at the end of the generation process.

4 Conclusions

This paper presents an intelligent system to compose melodies using a common device such as a mouse to control the duration and the pitch of the generated notes. The melodies adapt to user preferences through the indications given by the position of the mouse in an interface. As a first step, the proposed approach retrieves a set of MIDI files from which some musical features are extracted. A Markov model is then trained with the data of the collection of music, and the transition probabilities of this model are modified according to the control device to generate a melody that respects these "controlled" probabilities.

To deploy the system, an application has been developed that could be described as a controllable intelligent sequencer, since on the one hand it learns to generate sequences from sets of examples and on the other it admits the direct intervention of the user to guide the process of generation of the melody.

The results of the different experiments carried out emphasize the importance of user preferences in the melody generation. However, despite the users' indications, we do not avoid to follow the standards of the popular music in the generation process. In fact, the user has an essential role to guide the generation process and adapt the melody to his preferences through a mouse.

To improve the interaction of the users, we would like to analyze the incorporation of more standard devices, such as a keyboard or a joystick, to indicate a more accurate feature related to pitch and rhythm duration. In order to improve the automatic generation of music, it is needed to analyze the melodies with new properties, including a more general view of the composition. In popular music, the vocal songs are really important for ethnomusicologist studies. Consequently, a deeper analysis of popular songs and a study to incorporate popular lyrics to the music generated will be addressed in a future work.

Acknowledgments

This work was supported by the Spanish Ministry of Economy and FEDER funds. Project SURF: Intelligent System for integrated and sustainable management of urban fleets TIN2015-65515-C4-3-R.

References

1. Ajoodha, R., Klein, R., Jakovljevic, M.: Using statistical models and evolutionary algorithms in algorithmic music composition. In: Encyclopedia of Information Science and Technology, 3rd edn., pp. 6050–6062. IGI Global (2015)

2. Bergsland, A., Wechsler, R.: Composing interactive dance pieces for the motioncomposer, a device for persons with disabilities. In: Proceedings of the International Conference on New Interfaces for Musical Expression, pp. 20–23 (2015)

3. Casella, P., Paiva, A.: MAgentA: an architecture for real time automatic composition of background music. In: Intelligent Virtual Agents, pp. 224–232 (2001)
[Crossref]

4. Colton, S., Wiggins, G.A.: Computational creativity: the final frontier? Front. Artif. Intell. Appl. **242**, 21–26 (2012)

5. Jungleib, S.: General Midi. AR Editions, Inc. (1996)

6. Manzano, M.: El folklore musical en españa, hoy. Boletín Informativo de la Fundación Juan March, vol. 204, pp. 3–18 (1990)

7. Moreira, J., Roy, P., Pachet, F.: Virtualband: interacting with stylistically consistent agents. In: ISMIR, pp. 341–346 (2013)

8. Pachet, F.: The continuator: musical interaction with style. J. New Music. Res. **32**(3), 333–341 (2003)
[Crossref]

9. Pachet, F., Roy, P.: Markov constraints: steerable generation of Markov sequences. Constraints **16**(2), 148–172 (2011)
[MathSciNet][Crossref]

10. Papadopoulos, A., Roy, P., Pachet, F.: Assisted lead sheet composition using flowcomposer. In: International Conference on Principles and Practice of Constraint Programming, pp. 769–785. Springer (2016)

11. Whorley, R.P., Conklin, D.: Music generation from statistical models of harmony. J. New Music. Res. **45**(2), 160–183 (2016)
[Crossref]

12. Whorley, R.P., Wiggins, G.A., Rhodes, C., Pearce, M.T.: Multiple viewpoint systems: time complexity and the construction of domains for complex musical viewpoints in the harmonization problem. J. New Music. Res. **42**(3), 237–266 (2013)
[Crossref]