Journal of Ambient Intelligence and Humanized Computing
https://doi.org/10.1007/s12652-020-01822-5

ORIGINAL RESEARCH

Integration of a music generator and a song lyrics generator to create Spanish popular songs

María Navarro-Cáceres[1] · Hugo Gonçalo Oliveira[2] · Pedro Martins[2] · Amílcar Cardoso[2]

Received: 7 December 2018 / Accepted: 19 February 2020

Abstract
The automatic generation of music is an emerging field of research that has attracted wide attention in Computer Science. However, most works are centered in classical music. This work develops ETHNO-MUSIC, an intelligent system that generates melodies based on popular music. ETHNO-MUSIC generates melodies with Markov models, which learns from a corpus of Spanish popular music. Then, given the importance of the lyrics in this context, ETHNO-MUSIC was integrated with Tra-La-Lyrics, an existing system that generates lyrics following a melody, which has been specifically adapted to suit this purpose. Several experiments were carried out to evaluate the quality of the results, based on human opinions towards generated pieces of music and lyrics. Overall, results are positive. Briefly, they reflect that, on the one hand, the melodies transmit a feeling of Spanish popular music, and on the other hand, the text of the lyrics is related to the topics analyzed, and the rhythm follows the melodic aspects of the music.

Keywords Computational creativity · Music generation · Lyrics generation

1 Introduction

Popular music has had a great influence in society, since it has been transmitted orally from generation to generation. From an ethnomusicological point of view, this music has a great interest due to its peculiar sound characteristics and the social context that they represent or have represented. This kind of music has some particular features that makes them very different to other genres like classical or jazz music. They are commonly composed of melodies without musical accompaniment, with complex rhythms and uncommon scales (from a classical perspective), rich in ornaments.

Unlike classical music, Spanish popular music is always linked to a functionality, meaning the purpose for which the melody was conceived. In this sense, lyrics are essential for identifying this purpose, i.e. if it is a work song or a love song. Consequently, most of the existing repertoire include the use of lyrics, which in fact, is one of the most important factors in the popular culture. Although the popular melody usually follows the lyrics structure and rhythm, it also happens to the contrary, lyrics that are adapted to a melody created beforehand. This case is specially interesting to analyze what kind of words are chosen to be part of the lyrics, what topics are more common and how they adapt to the popular tone.

Years of ethnomusicological education are often required to master the idiosincracies of the modal system and understand the vocal music in a popular context. Consequently, to be able to create songs that combine music and lyrics is considered a challenge, moreover if the creator is based on Artificial Intelligence. Currently, there are some works centered in the application of artificial intelligence to generate music from lyrics. Singh et al. (2017) create narrative-based lyrics, which serve as basis to generate music. Similarly, ALYSIA (Ackerman and Loker 2017) is able to generate music from given lyrics. M.U. Sicus-Apparatus (Toivanen et al. 2013) generates both lyrics and music for art songs. However, most of the literature reviewed does not construct a whole system that incorporates music and lyrics generation, and are centered in one of the process, either music generation or lyrics creation. Additionally, these proposals usually generate music following the lyrics structure, whereas in

✉ María Navarro-Cáceres
maria90@usal.es

[1] Expert Systems and Applications Laboratory (ESALab), Department of Computer Sciences, Universidad de Salamanca, Salamanca, Spain

[2] CISUC, Department of Informatics Engineering, University of Coimbra, 3004-504 Coimbra, Portugal

Published online: 11 March 2020

this work we pretend to adapt lyrics once the melody is constructed and cannot be changed, as it commonly occurs in the popular tradition. Moreover, the initiatives have usually focused on the generation of music of a rather tonal/classical character and, to date, there has been no relevant studies addressing the generation of Spanish popular music. As we mentioned, this genre of music differs from the classical music in many aspects, including the sonority, the sounds disposition or the rhythmic formulas used.

For this purpose, we present a system capable to generate music and lyrics whithin a popular context. The contribution is precisely how the popular music is integrated in two systems and adapted to generate popular songs successfully. For this purpose, new melodies are generated following the style of original Spanish popular songs. These songs have extracted from multiple sources of traditional music and, after the analysis, encoding and storage of relevant features, are used as a training corpus.

Among the different learning models that can be applied to generate music from a previous corpus (Ajoodha et al. 2015), Markov models (MMs) have been selected due to their successful application in other related works (Papadopoulos et al. 2016; Whorley and Conklin 2016; Williams et al. 2019; Pachet and Roy 2011). MMs are trained in a corpus and then used to generate a new melody that fits the style of popular songs. The novelty of the present proposal is that the MMs will allow the participation of users that will guide the melody and therefore, guide the probabilities that the MMs calculate to select the notes that will be part of the melody.

Subsequently, we worked on an integration of Tra-la-Lyrics (Gonçalo Oliveira 2015), which generates lyrics for a given rhythm in our previous system. This is in line with recent collaborations between different creative systems [see, e.g., Znidarsic et al. (2016), Concepción et al. (2018)].

Tra-la-lyrics has been adapted to the Spanish language through an augmented semantic network and new line templates, collected automatically from the lyrics of Spanish popular songs. In most songs, the sentences are connected through a common subject. To imitate this behavior and set the generation domain, seed words were carefully selected according to the common sets that appear in popular songs.

With those initial parameters established, generation of lyrics can start once the melody is created. Simple heuristics are applied for setting a suitable division of the music into phrases, and then lines of text are generated for each part, while trying to maximize two main constraints: (i) one syllable per note; (ii) stressed syllables matching strong beats of the melody.

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.

Additionally, among the vocal songs generated, some were selected for another human evaluation. The evaluators had to score on the one hand, the quality of the melody, sound and rhythm according to the Spanish popular music standards. On the other hand, they are also asked whether the text of the lyrics was within the target domains, and about the rhythm and its meaning of the lyrics.

The remainder of this paper is structured as follows. Section 2 starts with a brief overview of related work on music and lyrics generation. Section 3 describes the generation process to create popular melodies. Section 4 gives an overview about the system understood as an integration of ETHNO-MUSIC and TraLaLyrics, detailed in Sect. 4.1. Section 5 provides an analysis about the experiments, examples and evaluation of the system developed. Finally, last section describes the final conclusions and future work.

2 Related work

Artificial Intelligence (AI) has been widely applied to problems in very different contexts, such as Biology, History or Mathematics. With the recent evolution of Artificial Intelligence, the field of Computational Creativity (CC) (Colton and Wiggins 2012) has attracted the interest of researchers in Computer Science, Cognitive Sciences and Arts.

Creative systems were developed for different artistic fields, including the generation of visual arts (Machado and Cardoso 2002; Colton et al. 2015), narrative (León and Gervás 2014; Pérez y Pérez 2015; Ontanón et al. 2012), poetry (Gonçalo Oliveira 2017; Lamb et al. 2017), or music (Zhu et al. 2018; Ajoodha et al. 2015).

Different techniques have been applied for generating music automatically. For instance, evolutionary computing has been widely used for generating interesting musical products, including a chord sequencer (Scirea et al. 2016), a system to detect motifs Serrà and Arcos (2016), or an accompaniment system based on emotions (Kuo et al. 2015).

However, statistical methods were a preferable tool due to their ability to learn from the context or the corpus, which is the case of our proposal. The deep analysis performed by Whorley and Conklin (2016) about different statistical models is particularly interesting, as it demonstrates how these methods are quite succesful in the generation of music and art. We should also remark that, specifically, Markov models are very commonly used for generating different kinds of music (Anton and Trausan-Matu 2018; Pachet and Roy 2011; Papadopoulos et al. 2016). However, they are usually thought to generate classical or jazz music automatically, and not as a guide to generate popular songs. Recently, companies as Google have been significantly developed projects

such as Magenta (Project 2018), which uses convolutional neural networks to create melodies. Although the deep learning techniques have supposed a wide improvement in the community, it commonly requires plenty of data for the training corpus, and cannot be applied in our problem, as the amount of melodies that can be used as a corpus are quite low.

Music generation systems can be adapted and used as tools to provide new music that accompany given lyrics. For instance, Monteith et al. (2012) analyze the stresses of text for obtaining the rhythm, generate several sequences of pitches for the rhythmic progression, and finally select the best, according to a set of aesthetic and singability measures (e.g., low repetition, low octave range, small interval jumps). ALYSIA (Ackerman and Loker 2017) learns the relation between music and lyrics from a song corpus, with a machine learning algorithm such as Random Forests, and then suggests suitable melodies for a given text. From the suggestions, the user selects the melody to use.

The generation of song lyrics can also be seen as a kind of poetry generation, for which there are several computational approaches, based on broad range of techniques. There are some works that make use of neural networks for generating melodies from lyrics in Chinese Bao et al. (2019), learn higher-level rhyming constraints with generative adversarial networks Jhamtani et al. (2019) or generate song lyrics in Korean, with a neural network with a LSTM for maintaining context Son et al. (2019) [see Gonçalo Oliveira (2017) or Lamb et al. (2017) for surveys]. Even though poetic text typically follows metrical constraints, song lyrics must address these more tightly because lyrics aim to be sung. In addition, lyrics should follow syntactic rules and, ideally, be meaningful under some topic. Works on the generation of song lyrics for given melodies include the original version of Tra-la-Lyrics (Gonçalo Oliveira et al. 2007), which tries to match the stress of the words with the strong beats in the melody, further adding syntactic constraints and repetition, but without any concerns on semantics. In order to generate more meaningful text, other systems add semantic constraints to the previous. This can be made, for instance, with the help of a semantic network (Ramakrishnan A and Devi 2010), also the case of the most recent version of Tra-la-Lyrics (Gonçalo Oliveira 2015), word associations (Toivanen et al. 2013), or as semantic relatedness constraints in a Markov process (Barbieri et al. 2012). In the latter work, a constraint model is learned from a set of lyrics of the same author for generating new lyrics in that author's style. In recent work (Watanabe et al. 2018), a Recurrent Neural Network Language Model is learned from a large collection of lyrics aligned to their melodies. As it happens for music generation, there are also systems that do not generate song lyrics independently, but may assist in the process of writing, as they suggest sentences that suit the metre, the rhythm and rhymes [see, e.g., Abe and Ito (2012); Watanabe et al. (2017)].

When it comes to the generation of both music and lyrics, there is not as much work. A common approach is to start by creating the lyrics and then generating suitable music for accompaniment. Known systems generate lyrics first and only then music. One example is the work by Singh et al. (2017), who integrate a plot generation system and a lyrics generator in the production of narrative-based lyrics. In that case, lines are generated with a Markov model and a generate-and-test approach is followed for selecting the most suitable lines to use in the lyrics, according to rhyme and sentiment. Suitable melodies may then be generated with a system such as ALYSIA (Ackerman and Loker 2017). M.U. Sicus-Apparatus (Toivanen et al. 2013) generates both lyrics and music for art songs. Lyrics are generated first, by replacing words in fragments of human-created poems, resorting to word associations for semantic coherence. Music generation considers the number of syllables, their length and punctuation signs for the rhythm, and a user-specified mood for the harmony, created based on second-order Markov chain selections of regular harmonic patterns expressed in a given key. A relevant feature, which may improve the connection between music and lyrics, is that the music generation process has access to the parameters that guided the generation of the lyrics.

As in the previous, in this work, both music and lyrics are produced. However, in this case, music is produced first, and independently of the lyrics. The resulting process is analogous to having two different people composing a song: one that composes the melody and another the lyrics. Although other techniques could have been adopted for both music and lyrics generation, we decided to tackle this goal with two systems that were familiar to us — ETHNO-MUSIC and Tra-la-Lyrics 2.0 — and focus on their integration for the generation of Spanish popular songs. Other reasons that lead to this decision include the nature of ETHNO-MUSIC, developed with Spanish popular music in mind, and the ease of integrating and adapting Tra-la-Lyrics 2.0 to this domain, given that it is built on top of PoeTryMe, a flexible poetry generation platform, already adapted to many different scenarios, including the generation in three languagues (Gonçalo Oliveira et al. 2017): Portuguese, Spanish and English.

3 System description

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

The system that generates music, ETHNO-MUSIC, is provided with a memory to store beforehand different melodies in the popular music style. These music files are retrieved from a wide variety of popular music melodies

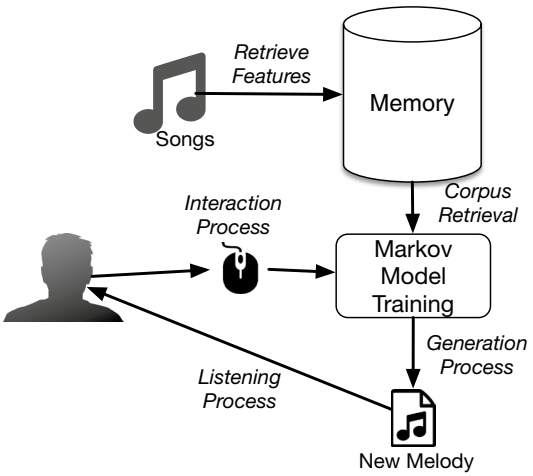


Fig. 1 Overview of ETHNO-MUSIC

from song books. The selection process is described in Sect. 3.1. Secondly, these melodies should be correctly classified, analyzed and encoded to extract the main features of each song, as explained in Sect. 3.2. The results of the analysis are used as the training corpus by the Markov models, which will generate music in the popular style, further described in Sect. 3.3.

3.1 Music retrieval

The resulting platform aims at the generation of music based on the features of the Spanish popular music, where, according to ethnomusicologists (Manzano Alonso 2001), three main factors should be considered:

- Melodic behavior: Unlike the tonal music, which is based in minor and major modes, Spanish popular music makes use of seven modes, each built from the seven notes of the natural scale, starting from different notes. The modes are: Do Mode, Re Mode, Mi Mode, Fa Mode, Sol Mode, La Mode, Ti Mode. The melodic behavior consists of continuous chromatizations and instabilities.
- Rhythm: Popular music usually resorts to uniform beat due to the syllabic text used in the lyrics, or their functionality, created to dance or to follow rituals.
- Genre: The genre is related to the functionality of the music, namely the context in which the music has been conceived. For example, "sevillanas" are created according to a specific dance, "lullabyes" share some common features in lyrics and music to make children sleepy, etc. In this sense, we can classify music in different genres, each one representing their own musical features, such as sonority, lyrics and rhythm. Some examples are work songs, love songs, lullabies, wedding songs, sacred songs, dance music, among others.

Generally, it is very difficult to understand the specific features that some songs can share in popular music, since this type of music has been in constant evolution due to its characteristic grammar and its oral transmission from generation to generation. Therefore, folk songs in Spain are as varied as its regions.

In order to obtain this information and preserve popular music, some ethnomusicologists have visited different geographical zones, transcribing folk songs that people sing in rituals or traditional holidays. This information was analyzed, classified and published in songbooks, which often pass a scientific process to guarantee the quality of the ethnomusicologic research. Therefore, the songs used as the training corpus are extracted from songbooks, recordings or digital scores published.

By 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. According to some authors (Schindler 1991), in Spanish popular music, work, love songs and lullabies share features related to the key signature, rhythmic patterns and general sonority or tonality, which makes them very interesting to use as a corpus in the development of the learning model. Therefore, in order to train the MM, melodies related to these genres were collected (Manzano 1990).

3.2 Encoding melodies

In order to represent the information about rhythm and sonority, the sources must be digitized. But to identify the popular music, we should not only analyze the particular duration or pitch. There are other relevant features, such as the duration of the musical phrases, the degrees in which the melody reposes (notes with a long duration), and the particular cadences. Unlike classical music, in popular music the harmonic tension and the use of the chords degrees 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 by Whorley et al. (2013), the following features of the popular songs were selected:

- Pitch: musical note. It is encoded with a MIDI Musical Instrument Digital Interface) number from 0 to 127.
- Duration: rhythmic formula of one note. Each number represents a whole, a half, a quarter, etc.
- Degree: position of the note within the musical scale. It can take values from 1 to 7, where 1 means the first degree, 7 means the last note in the scale.
- First in bar: boolean value which indicates if the note is the first in a bar or not.

- Time signature: it is a rate that represents the time signature of the melody. it consits of two numbers, one for the numerator, which represents the total number of rhythmic beats and another for the denominator, which represents the rhythmic value (whole, half, quarter, etc.) for each beat.
- Musical Phrase: it represents the position of a note in a musical phrase. It can take three possible values: 1 if the note is the first in the phrase, -1 if it is the last one, and 0 otherwise

Currently, there is no standard format that only addresses these features. However, a wide number of songs is already encoded as MIDI [files (Jungleib 1996)], 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. These data are finally stored in a spreadsheet file.

3.3 Music generation

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. The time of training was about 3 min.

The Markov models represents a special type of stochastic process in which the probability of an event occurring (in this case a note or silence with a given duration) depends only on the n previous events. This feature of "limited memory" is what is known as Markov property. In our case, the popular music does not have so many musical resources as classical music, for example, which makes the number of states of the MM to decrease. We made empirical experiments to study the variability of the music, and according to that, n was set to 3.

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:

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

where k is the weight of each probability.

Typically, musical and text phrases in Spanish popular music are quite short in order to make them easier for people to learn and start singing right away. Therefore, we added a constraint to keep the musical phrases shorter than 3 bars. In particular, the system has a fuzzy threshold parameter, where the probability that long notes appear is increased as more notes are added to the same phrase while the MM is running.

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.

4 Integration with Tra-La-Lyrics

The generation of popular Spanish songs results from the integration of two creative systems: ETHNO-MUSIC, in charge of generating melodies, and Tra-la-Lyrics (Gonçalo Oliveira 2015), in charge of generating suitable lyrics. Figure 2 gives an overview of the resulting generation flow.

As we described in the previous section, music is generated based on the features of Spanish popular music, which include the rhythm and the sonority. Initially, the user selects the musical style (sonority) and the topic for the lyrics generation. Then, ETHNO-MUSIC generates a new melody with the user. Once the melody is available, Tra-la-Lyrics is used for generating lyrics on the desired topic that suit the rhythm of the melody. Finally, the results of Tra-la-Lyrics for the melody generated are shown to the user.

4.1 Lyrics generation

Tra-la-Lyrics (Gonçalo Oliveira 2015) generates text based on rhythm, and is currently built on top of PoeTryMe (Gonçalo Oliveira et al. 2017), a versatile poetry generation platform. Tra-la-Lyrics can thus be seen as an instantiation of

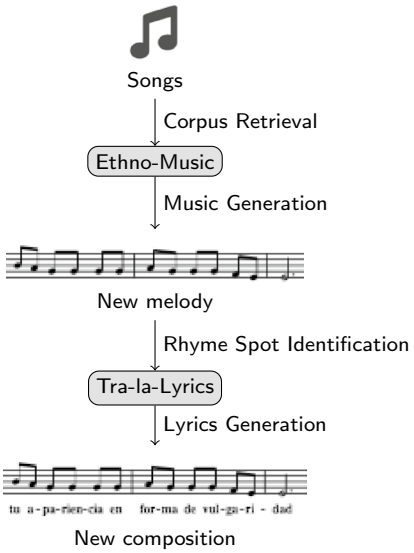


Fig. 2 Overview of music generation flow

the latter and shares most of PoeTryMe constituents, including the line generation module, which produces semantically-coherent lines, with the help of a semantic network, and a grammar with line templates. Furthermore, Tra-la-Lyrics may follow a generation strategy specifically adapted to the generation of suitable lyrics for a given melody. For this purpose, the target melody must be first analysed, in order to break the melody into phrases (see Sect. 4.1.1), in case the latter division is not provided, and for representing the rhythm as sequence of beats (see Sect. 4.1.2).

Input melodies must be represented in the ABC music notation[1], which is easily obtained from a MIDI file[2]. For producing text that matches the rhythm, the Generation Strategy (see Sect. 4.1.3) should consider the number of beats as well as their strength. It will then use the set of seed words for constraining line generation (see Sect. 4.1.4) and compose the lyrics, which will consist of a selection of the best lines for each phrase, out of those generated by the Line Generation. The following sections describe the previous modules.

4.1.1 Phrase breaker

In PoeTryMe, a generation strategy selects and organizes lines, produced by the Line Generator module, according to a given form. These lines are coherent textual fragments end their end is a suitable place for rhymes. For poems, the form is given by the number of stanzas, lines and target lengths. Yet, given a melody, the limits of lines have to be provided or identified automatically. In this work, simple heuristics were applied for the latter purpose, based on the lengths of the longest pause (*LP*) and longest note (*LN*): the melody is split after pauses that are *LP*/3-long or longer and notes that are *LN*/2-long or longer, if resulting phrases have at least *minP* notes. Phrases with more than *maxP* notes go through the same splitting process until they have less than *maxP* notes. These heuristics are similar to the LBDM algorithm for music segmentation (Cambouropoulos 2001), but only applied to rhythm, rather than pitch segmentation.

4.1.2 Rhythm analyzer

The rhythm analyzer is the same as in the original version of Tra-la-Lyrics (Gonçalo Oliveira et al. 2007). It is based on the dot system (Lerdahl and Jackendoff 1983), which sets the metrical accents of each beat inside a bar, and thus the strengths of each note, according to their position. The first beat of each bar is always the strongest. Strengths are then distributed according to the beats (stronger) and downbeats (weaker), depending on the time signature. In fact, Tra-la-Lyrics only considers those that are substantially stronger, when compared to the remaining, as actual strong beats. For instance, in a 4/4, those will be the first and the third crotchets, in a 6/8 the first and the fourth quaver, but in a 3/4, only the first crotchet is used as the only strong beat.

4.1.3 Generation strategy

Similarly to other instantiations of PoeTryMe, lines are produced with a generate-and-test approach. Briefly, for each line in the poem structure, textual fragments are retrieved from the Line Generation module. This strategy is responsible for selecting the fittest fragments for each line, from a maximum of n retrieved fragments per line. The fitness function is based on features that are relevant to the rhythm. More specifically, it has penalties for:

- the difference between the number of syllables in a textual line and the number of notes in the musical part (α);
- unstressed syllables in strong beats (β);
- stressed syllables in weak beats (γ);
- words interrupted by a pause (δ).

In order to increase the chance of rhymes, there is a bonus for each pair of nearby lines with the same termination (ϵ), unless they end with the same word, which results in another penalty (ζ). This function is summarized in Eq 2.

$$Fitness(line) = \epsilon - (\zeta + \alpha + \beta + \gamma + \delta) \quad (2)$$

[1] http://abcnotation.com/.

[2] Using, e.g., a UNIX binary such as `midi2abc`.

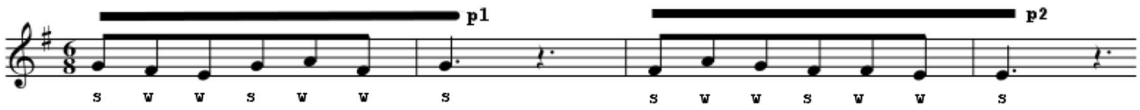


Fig. 3 Melody with two 7-note phrases (p1 and p2) with the strength of each beat ('s' for stress, 'w' for weak)

4.1.4 Line generation

PoeTryMe has line generation modules for producing text fragments in Portuguese, English and Spanish, based on a semantic network and a grammar with templates, tightly connected to the relation types in the network. Towards generation within a semantic domain, the network is first constrained by a set of seed words, which will select a subnetwork that sets the generation domain. Line generation involves selecting an edge in this subnetwork (e.g., a-relation-b), which represents a relation instance, and using its arguments for filling the placeholders of a suitable textual template. For instance, the following relations:

– *computer* usedFor *work*; *guitar* usedFor *music*; *hat* usedFor *shade*)

Could be used with the following rules:

– usedFor → *choose the best [arg1] for your [arg2]*
– usedFor → *if you need [arg2], i'll be your [arg1]*

Given the kind of text to generate, it made sense to use the Spanish instantiation of PoeTryMe (Gonçalo Oliveira et al. 2017), though with some additions that enable the generation of more varied lines. In this instantiation, semantic relations were originally acquired from the Spanish wordnet, in the Multilingual Central Repository (Gonzalez-Agirre et al. 2012), which covers mainly synonyms and hypernyms, while lines of the generation grammar had been extracted from an anthology of about 400 Spanish poems. For this work, the semantic network was enriched with relations between two Spanish words in the most recent version of ConceptNet (Speer et al. 2017). Moreover, for the creation of the grammar, a set of about 280 Spanish popular song lyrics, transcribed for this purpose, was exploited, in addition to the 400 poems, thus increasing variation in the produced text. To complete the adaptation, lyrics had to be generated with seed words related to concepts typically invoked in Spanish popular lyrics.

4.2 Example

Figure 3 shows a simple melody, with a 6/8 time signature, which would have two parts (marked with p1 and p2) with seven notes each, split by a rest. For each note, the strength of the corresponding beat is also marked, based on the 6/8 time signature – `s` stands for stressed and `w` for weak beats.

Now suppose that:

– We would like to generate a line for the first part with the topic *amor* (love);
– With the semantic network in Figure 4 and a grammar that we omit (used rules can be inferred from network and the produced lines);
– Using the following parameters for the fitness function: $\alpha = 1$, $\beta = 0.5$, $\gamma = 0.1$

The Line Generator could start, for instance, by producing the following line:

- *una calentura y un amor*

To obtain its rhythm, this line is split into syllables:

- *u-na ca-len-tu-ra y un a-mor* → `s w w w s w s w w s`

When matching the rhythm of this line with the rhythm of the musical phrase, there are three syllables more than the target rhythm, plus one weak syllable in a strong beat and one strong syllable in a weak beat. Not considering potential rhymes, the fitness would thus be: $2 + 0.5 + 0.1 = 3.6$. The Line Generator would then keep on producing lines, for which fitness would be computed. The line with the lowest score is always kept to be used in the lyrics.

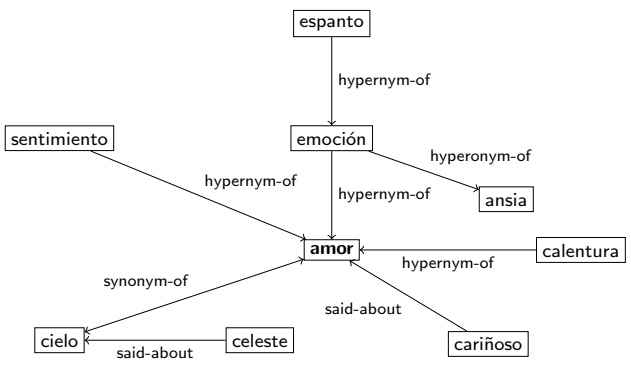


Fig. 4 Part of a semantic network, with the words and relations used for generating the example line

Here are a few examples of other lines that could be generated, their rhythm pattern and fitness score:

- con emociones temo y con espanto suspir
  → s w w s w s w w s w s w w s w → 8.6
- como un amor de sentimiento
  → s w w s w w w s w → 2.5
- celeste prenda de cielo
  → w s w s w w s w → 1.6
- olalla cariñosa en lo amor
  → w s w w s w w w w s → 5.8

The generation of lines would stop once the maximum number of generations is reached or when a line with a score of 0 (or lower, if rhymes are considered) is produced. The following is an example of such a line for this example:

- ansias de mi emoción
  → s w w s w w s → 0.0

The generation of lines for the remaining phrases would follow the same procedure. In the end, there will be a line for each phrase of the melody.

5 Results and discussion

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

The evaluation of its musical results is the remaining challenge because of the inherent subjectivity of human listeners (Pearce et al. 2002). However, the involvement of users in the evaluation is essential to demonstrate the usefulness of our work, as the system is thought as a tool to generate music with lyrics. Consequently, the musical results should also be acceptable for our potential users or listeners from a popular musical point of view. In fact, for an evaluation of the final results, most works use objective measures and some optimization method without human evaluation at all, while a few use a human expert to evaluate the results. López-Ortega and López-Popa (2012) discuss the quality of the system theoretically applying some creative concepts, such as deliberation and spontaneity. Herremans and Sörensen (2013) presented an exhaustive statistical experiment to validate the efficiency according to the rules established in the system. Delgado et al. (2009) conducted a preliminary listening test to evaluate their application. Pearce and Wiggins (2007) proposed a listening test and a study with statistical measures. Collins et al. (2016) also performed a listening test and a comparative study between the present system and a previous work. Therefore, in this work, we will also perform a listening test in which expert users are involved, followed by a statistical analysis and a comparative study to demonstrate the goals previously exposed.

For the generation of the music, 280 popular songs were selected. All of them make use of similar rhythm patterns, covering time signatures 2/4, 3/4, 4/4 and 6/8. Each song consisted of 3 or 4 musical phrases with similar length, and they are constructed following two possible scales. Therefore, we divided the corpus in Mi mode with possible modifications of this mode in its evolution to E minor and La mode with possible modifications to evolve to A minor.

The melodies have been encoded according to the Sect. 4 and saved in an Spreadsheet with all the properties. These features were the training corpus for the Markov model. The memory of the MM, which corresponds to the number of states that it can remember for the future generation of the melody, was empirically set to 3. During the generation process, each iteration of the system consists of adding a new note to the melody. It is iterated until the system decides to stop.

In this section, we first describe the evaluation about the performance of the system to generate music in Sect. 5.1. In a second stage, we produced several songs to perform a listening test and retrieve a subjective evaluation of the popular music and the lyrics generated. This evaluation and the discussion is described in Sect. 5.2. Finally, in order to know the popular features that they contain and how they were generated, some of the examples produced are shown and analyzed in Section 5.3.

5.1 Analyzing the interaction with the user

In the original ETHNO-MUSIC, 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.

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 La mode and 5 melodies of Mi mode. Additionally, 4 melodies have a time signature of 6/8, 3 of 3/4 and 3 of 2/4.

We follow Berenson et al. (2012) theory to measure the statistical conditions of our listening test. The confidence level refers to the probability of success in the estimation, and is set to the standard for most statistical analysis, 95%. The standard error is related to the distribution and the deviation of the parameters measured in the test. In percentages, for this listening test we have a relative standard error of 21.18%, which can be considered admissible for such subjective tests.

Table 1 Statistics resulting from the listening test results

	χ^2	M_e	M_o
Melody	$6.238e-04$	4	4
Sound	$2.136e-02$	4	4
Rhythm	$5.641e-03$	4	3

Table 2 Final statistics after the users finished testing the system

	Easy to use	Interface	Control Quality	Overall Ratings
Mode	4	3	4	4
Median	4	3	4	3

The listening tests always include some level of subjectivity, as usually depends on the culture or even the mood of the people involved. To minimize the subjectivity of our evaluation, we looked for expert users in Spanish popular music. They have studied Ethnomusicology or Musicology, or are very familiar with the Spanish tradition due to their background (for example, teachers of traditional instruments, or musicians of Spanish folk music). Their background makes the results more reliable, as they are asked to evaluate more "objective" features (considering that music and art in general, are very subjective fields). These objective features include similarity between the music generated and the Spanish popular music style, similarity of lyrics, or significance, always from the popular music perspectives.

Finally, 21 expert users in ethnomusicology were selected. They had at least four years of experience working in Ethnomusicology or Musicology, or have finished their music studies in those degrees less than one year ago. Alternatively, we also contacted people that worked as a musician of Spanish folk music or teachers that teaches Spanish traditional instruments, and they are familiar with the Spanish tradition. These experts were asked if they think the melodies generated follows the standards of popular music according to the following items:

- Melody: how pleasant is the melody?
- Sound: how well does the melody, in some way, give a feeling of the popular style of the songs?
- Rhythm of the melody: how well does the rhythm suit the popular music style?

They should evaluate the quality in a 5-point Likert scale, 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 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 can generate melodies with a good musical quality and that 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 in both items. The mode for the rhythmic aspects shows most users thought the rhythm is average. The rhythm is quite difficult to capture, even when the user tries to make changes throught the device, as the the melodies in popular music often use very regular figures and it is difficult to extract new ones. However, the results are also quite positive, as half of the scores obtained are above "Fair "according to the median.

In this part, we also aim to validate the usefulness of the system. For this purpose, a total of 15 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.

The Table 2 shows that the general satisfaction degree is quite high, with a mode of 4 and median value of 3. The users consider the system to be very easy to use even for people without any musical training (with a median and a mode of 4), although the interface could be improved (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.

5.2 Analyzing the melody and lyrics

For the generation of lyrics, melodies were split into phrases using $minP = 8$ and $maxP = 16$, because Spanish popular songs typically use lines of 8 or 16 syllables. Given their commonality among Spanish popular music, the domains of *sleep* (common in lullabies) and *love* (used in love songs) were used for lyrics generation, and represented by the following groups of seed words:

- Sleep: *dormir, cuna, bebé, pañal, noche, mamá, coco, sueño, soñar, estrellas, luna* (in English, to sleep, cradle, baby, diaper, night, mummy, poo, dream, to dream, stars, moon);

Table 3 Comparison between our proposal and other proposals

	ETHNO-MUSIC +TraLaLyrics	ALYSIA	Singh et al. (2017)
Lyrics and music generation	X	X	X
Includes folk songs	X	–	–
Generating music before lyrics	X	–	–
Learning Machines	X	X	X
Use of Markov models	X	–	X
Interaction with the users	–	–	X
Use of Spanish language	X	–	–

– Love: *amor, novia, moza, mozo, bella, belleza, feliz, alcoba, morena, guapa, sonrisa, ojos, bonito, bonita* (in English, love, girlfriend, girl, lad, beautiful, beauty, happy, bedroom, brunette, pretty, smile, eyes, pretty).

For each selected melody-domain pair, 20 lyrics were generated. Yet, in the end, the system only selected the lyrics with higher scores on the rhythm for inclusion the validation set. Each line was the fittest of 5000 ($n = 5000$) and the following parameters were empirically set for the fitness function: $\alpha = 1, \beta = 0.5, \gamma = 0.1, \delta = 0.3, \epsilon = 2, \zeta = 1$

Finally, as we stated in Sect. 2, the system proposed share multiple features with other works, but also present some differences that are important to highlight. There are two specific proposals, AlYSIA (Ackerman and Loker 2017) and Singh et al. (2017) that have points in common, such as the application of learning techniques and the generation of lyrics and music as a process. To compare the features of the three systems and highlight the differences among them, a qualitative comparison is shown in Table 3.

The three systems are centered in the generation of lyrics and music altogether. However, only the proposal system generated lyrics according to a predefined melody, instead of using the lyrics as the basis to create a new melody. Additionally, both ETHNO-MUSIC and Singh et al. (2017) proposal make use of MMs with successful results according to the evaluation performed. However, our proposal includes important novelties, such as the use of Spanish language and the use of a new corpus of folk songs, unlike the rest of the works, which are centered in English classical and pop music. The techniques applied are related to machine learning.

5.2.1 Listening test

Once designed and implemented, we got insights on the validity of the new automated composer. Due to the inherent subjectivity of human listeners, this kind of evaluation remains a challenge for music composed automatically (Pearce et al. 2002). The same happens for creative text. Following other examples of subjective evaluation by a group of human listeners [e.g. Delgado et al. (2009); Pearce and Wiggins (2007); Collins et al. (2016)], who perform listening tests, we follow a similar approach for evaluating the musical results and the lyrics of the system.

The listening test was designed with 8 melodies generated by the system. Although the melodies can follow seven diatonic modes, we selected the two modes which appear in the Spanish culture more frequently, namely Mi Mode and La Mode. It is important to note that both modes show a similar behavior, making pauses in 4th, 3rd or 2nd notes of the scale, or containing only a limited tessitura around 5–6 notes. Therefore, the melodies selected should show this behavior. Additionally, we have to take into account that the Mi mode is more frequent than La mode in Spanish popular music. Therefore, for the listening test, we selected 2 melodies in La mode and 6 in Mi mode, both with limited tessitura and similar behavior.

Additionally, the melodies can have different time signatures. Among the possible options, Spanish popular music commonly has times of 6/8, 4/4, 3/4 and 2/4. Therefore, we selected melodies with those time signatures to analyze possible variations of lyrics according to different rhythmics. Finally, we selected one in 6/8, three in 4/4 and two in 3/4 and 2/4.

Finally, the melodies should reflect different aspects of the topics "love" and "sleep". For each of the melodies, two lyrics were generated, one in the topic love and another on sleep. Among them, 8 songs were selected for the listening tests, 4 of them related with love, and the rest with sleep.

A total of 17 users with musical knowledge about popular music (more than 4 years of experience in popular music, or students of Musicology) were asked to answer and online form with questions about the melodies generated and how they follow the standards of popular music according to the sonority and the rhythm. There were questions targetting the melody, the lyrics, and both. More precisely, the melodic line could be played by a piano, the lyrics were shown below in text, and there was also a picture with the score with lyrics included, as shown in Fig. 5.

The following questions were made for each song and were answered in a scale from 1 to 5, where 1 means Very poorly and 5 means Perfectly.

1. Melody: how pleasant is the melody?
2. Sound: how well does the melody, in some way, give a feeling of the popular style of the songs?
3. Rhythm of the melody: how well does the rhythm suit the popular music style?
4. Rhythm of the lyrics: how well does the text suit the rhythm of the original melody?

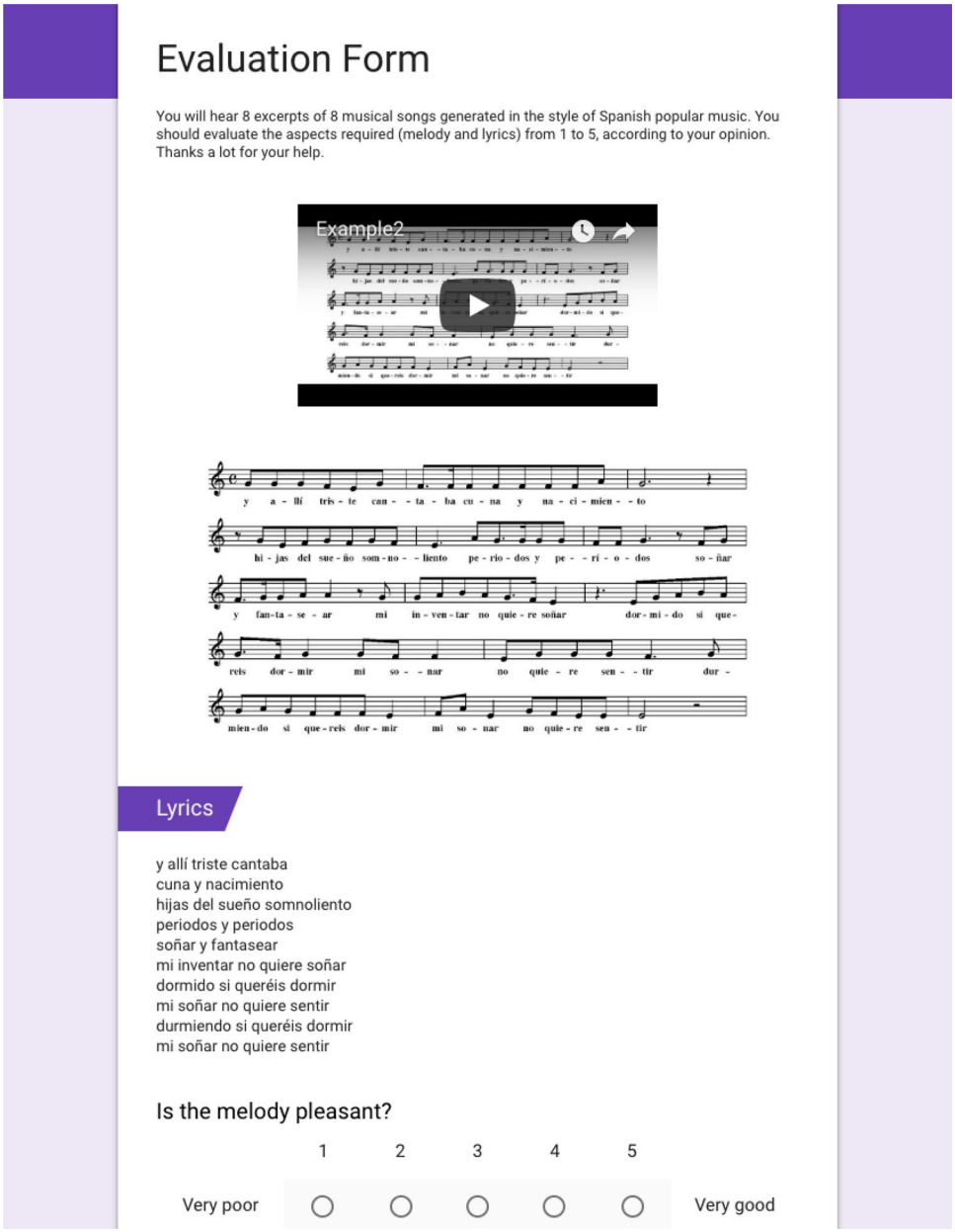


Fig. 5 Screenshot of the validation form to analyze melody and lyrics

5. Subject: how is the text related to any of the following topics: work, love?
6. Meaning: how much sense does the text of the lyrics make? Is it possible to, somehow, interpret it?
7. Overall quality: in general, what is the quality of the melody plus lyrics?

For such categorical data, we use the median and mode values to show 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 subjective ratings since all listeners can have different interpretations of the musical pieces and different musical tastes.

We expect the system to reflect the perceptual quality of the melodies according to the popular songs style. Likewise, we wanted to verify the quality of the lyrics according to the positive ratings given by the listeners. Table 4 presents the answers given for each assessed item (question), as well as their mode (M_o) and median (M_d).

The table shows a varied range of ratings, which confirms that assessing the resulting songs is a subjective task. Yet, for each assessed items, there is majority of 4 s and a few 3 s, which means that, overall, the results are positive.

More than the 54% of the responses evaluated the melody as "Good" or even better. That means the users found the melody pleasant when they listened to it. The 60% of the

Table 4 Overall validation results for the 10 assessed songs

Item	Rating					M_o	M_d
	1	2	3	4	5		
Melody	0	13	42	47	18	4	4
Sound	3	4	41	59	13	4	4
Rhythm (Melody)	0	8	47	52	13	4	4
Rhythm (Lyrics)	2	8	45	41	24	3	4
Subject	5	12	25	53	24	4	4
Text Meaning	20	6	32	50	11	4	4
Overall Quality	1	14	52	41	12	3	3

users considered that the music adapts to the popular music standards well or very well, as does the melodic rhythm, where the 54% of the users evaluated this item as good or very good.

The item more frequently rated with 1 (very poor) is the text meaning, which we consider to be the most challenging. Yet, overall, this item still got a positive median and mode. It was by a low margin, but the mode of the rhythm of the lyrics is in the middle of our scale, which shows that not all the lyrics match the rhythm perfectly. In fact, while, in some cases, it is matched almost perfectly, in other cases many words that do not match the rhythm. We will show this phenomenon with the examples in Sect. 5.3.

The worst rated item ended up being the overall quality, which, on the one hand, is a bit surprising, because it we would expect it to be roughly an average of all the other items. Yet, we may also see this as a result of summing all the issues identified in the previous items.

In order to check the impact of the seed words used for the generation on the lyrics, Table 5 presents the validation numbers according to the semantic domain. Results are comparable for all items but for the rhythm of the lyrics. According to these results, the rhythm of the lyrics generated in the 'sleep' semantic domain is worse than those generated in the 'love' domain. This suggests that the selection of the seeds does not have an impact on the meaning only, but also on the rhythm, even if indirectly, because seeds will constrain the semantic network, which is tightly related to the grammar. For instance, if selected seeds are not present in many relations, are only present in relations of a few types, or in relations of types for which the grammar does not have many patterns, the generation space might become too limited, making it difficult to find suitable fragments for every possible rhythmic sequences.

5.3 Examples

Figures 6 and 7 show examples of two generated songs, selected from the validation samples. We selected those ones that present different popular features, like different time signatures, scales and lyrics.

For each example, we present the score, the Spanish lyrics assigned to the corresponding notes, the same lyrics alone and a rough English translation for them. All the melodies generated share representative features of popular music, such as the constant repetition of pitches and the limited tessitura. Additionally, to some extent, lyrics match the rhythm and use a varied range of words related to the selected topics.

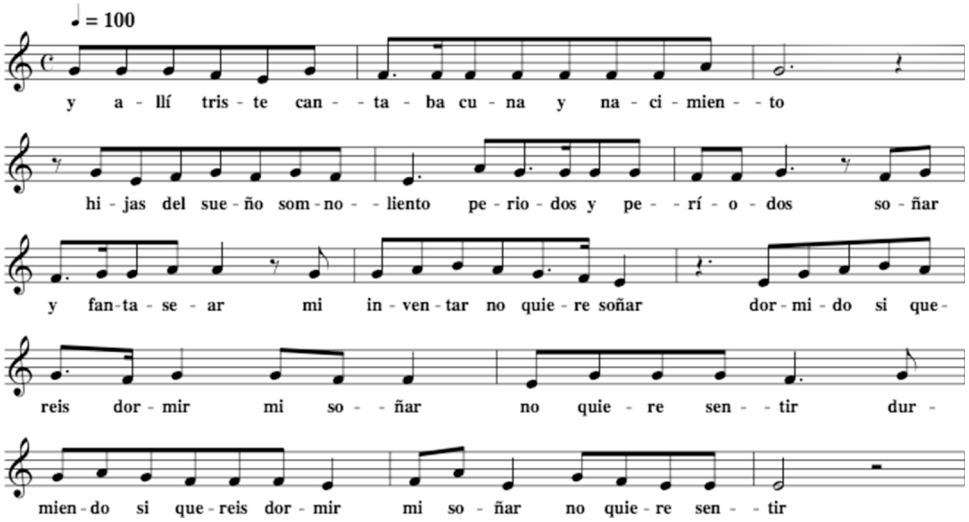
The melody in Fig. 6 shows a limited tessitura, as the notes only go from E to B. The rhythmic formulas are very repetitive and centered on the syllabic text that accompanies the melody, instead of showing complex rhythm figures. The melody usually makes pauses on the second and third degrees, which is typical of modal music, unlike tonality, where pauses are usually produced on fifth and tonic degrees. The melody represents a good example of popular music, as the median and modes obtained in the listening test of 4, even if it is not very pleasant according to the evaluations, with a mode and a median of 3.

Lyrics for this song were generated with the sleep-related seeds, which is clear by the presence of words like *cuna* (cradle), *dormir* (to sleep), *sueño* (dream) or *soñar* (to dream), from the seed set, and also other related words, such as *somnoliento* (sleepy), *dormido* (asleep), *durmiendo* (sleeping), or *sentir* (to feel, an hypernym of to dream). The previous are used in semantically-coherent textual fragments. This item was very well rated by the listeners, with a median and a mode of 4.

On the rhythm, all but two notes are assigned to a single syllable. Exceptions occur in the word *som-no-lien-to* and in one of the occurrences of *so-ñar*, for which the two last

Table 5 Validation of the lyrics for the 10 assessed songs, according to the semantic domain

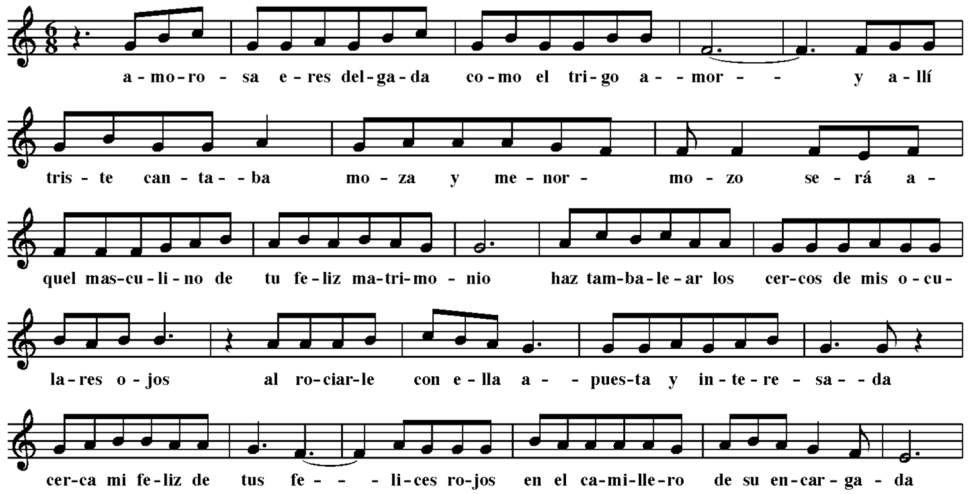
Item	Sleep		Love	
	M_o	M_d	M_o	M_d
Rhythm (lyrics)	3	3	4	4
Subject	4	4	4	4
Text meaning	4	4	4	4
Overall quality	3	3	3	3



y allí triste cantaba, cuna y nacimiento
hijas del sueño somnoliento
periodos y periodos
soñar y fantasear
mi inventar no quiere soñar
dormido si queréis dormir
mi soñar no quiere sentir
durmiendo si queréis dormir
mi soñar no quiere sentir

and there (he) was sadly singing, cradle and birth
daughters of the sleepy dream
periods and periods
of dream and fantasy
my inventing does not want to dream
asleep if you want to sleep
my snoring does not want to feel
sleeping if you want to sleep
my snoring does not want to feel

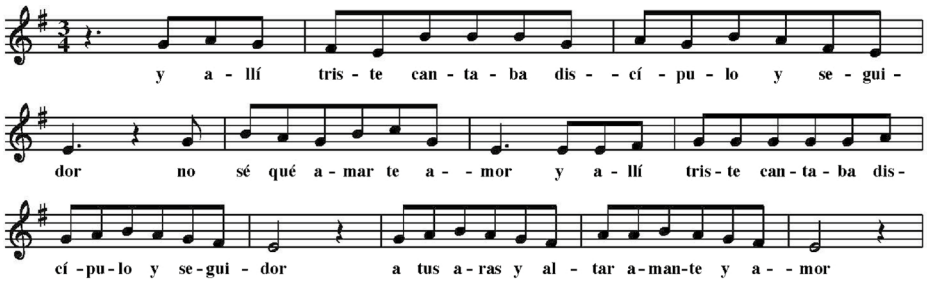
Fig. 6 Song in Mi mode, with Spanish lyrics on the domain 'dormir' (sleep), including rough English translation



amorosa eres delgada como el trigo amor
y allí triste cantaba, moza y menor
mozo será aquel masculino de tu feliz matrimonio
haz tambalear los cercos de mis oscuros ojos
al rociarle con ella apuesta e interesada
cerca mi feliz de tus felices rojos
en el camillero de su encargada

loving you are thin like wheat, my love
and there sadly singing, girl and minor
boy will be that male of your happy marriage
shake the edges of my dark eyes
by spraying her with it pretty and interested
(you're) close, my happy, to your happy reds
in your attendant's stretcher.

Fig. 7 Song in Mi mode, with Spanish lyrics on the domain 'amor' (love), including rough English translation



y allí triste cantaba, discípulo y seguidor
no sé qué amarte, amor
y allí triste cantaba, discípulo y seguidor
a tus aras y altar
amante y amor

and there was sadly singing
disciple and follower
and there was sadly singing
disciple and follower
to your altars and altar
lover and love

Fig. 8 Song in La mode, with Spanish lyrics on the domain 'amor' (love), including rough English translation

syllables are assigned to the same note. Furthermore, only a minority of unstressed syllables matches strong beats, namely the last syllable of *nas-ci-mien-to* and the first syllable of *in-ven-tar*. This is partially compensated by the presence of rhymes in *nacimiento/somnoliento*, *fantasear/soñar* and *sentir/dormir*. These issues are also perceived by the listeners, therefore a lower score has been obtained in the rhythm of the text, in particular, a mode and a median of 3.

In the melody of Fig. 7 the rhythm which accompanies the lyrics is quite repetitive. This fact is also perceived by the user, and consequently both the median and the mode for the rhythm of this song are 3. However, the melody behaves like the popular music, with rests in the second and third degrees, a typical feature of modal (and popular) music, in this case, of the Mi Mode. In this excerpt, the evaluation of this item obtained a median and a mode of 4.

Lyrics for this song were generated with the love-related seeds, which results in the presence of words like *amor* (love), *moza* (girl), *mozo* (boy), *ojos* (eyes), *feliz* (happy), both in the seed set, and other related words such as *amorosa* (lovely), *matrimonio* (marriage) or *apuesta* (pretty, synonym of *bella*). The relationship between the lyrics and the seed is quite accurate, as the results of the listening test demonstrated, with a median and a mode of 4.

On the rhythm, there are no notes with more than one syllable but there are three with none. Of those, two are tied to the previous, and thus considered by the Tra-la-Lyrics as their extension. Therefore, the only issue on this aspect is the word *me-nor*, which has two syllables that spread by three notes. We do not see this as critical, as it could by fixed at singing time, for instance, by repeating the last syllable in the third note. On the other hand, for this melody, the system struggled to match strong beats with stressed syllables, as there are several unstressed syllables matching strong beats, namely, the the first and third syllables of *a-mo-ro-sa*, the first of *del-ga-da*, *me-nor*, *se-rá*, *ma-tri-mo-nio*, the third of *tam-ba-le-ar*, the first of *ro-ciar-te*, *a-pues-ta*, *in-te-re-sa-da*, *fe-liz*, the first and third of *fe-li-ces*, the third of *ca-mi-lle-ro* and the fourth of *en-car-ga-da*. The number of issues is also due to the time signature of this music, 6/8, for which, apparently, it was more difficult to find textual fragments with matching rhythm. Despite the previous issues, three rhymes are present in *amor/menor*, *teresada/encargada* and *ojos/rojos*.

The melody of the third example, in Fig. 8 is generated in La mode, unlike the Examples 1 and 2. Therefore, even though the behavior is similar, the sonority is different and more similar (auditively) to tonality. It shows a limited tessitura but wider than the other examples, as the notes go from E to C. The melody usually makes pauses in the first degrees but avoids 5th and 4th degrees, which are frequently used in tonality. The behavior of the melody is quite good according to the listening test, with a median and a mode of 4. However, the rhythm obtained a median of 4, which means that at least half of the listeners considered the rhythm good or very good. However, the mode value is 3. Some users could have perceived that the rhythmic formulas are very repetitive and centered on the syllabic text that accompanies the melody, instead of showing some complex rhythm figures.

Lyrics for this song were again generated with the love-related seeds, which results in the presence of words like *amor* (love) and other related, such as *amarte* (loving you), *amante* (lover), or *discípulo* (disciple, hyponym-of *mozo*). Again the evaluation of this item demonstrate the good relationship between lyrics and seeds, with a median and a mode of 4.

On the rhythm, there are no notes with more than one syllable nor syllables that need to spread among more than one note. Rhythm is matched more tightly than the previous example, as all the syllables matching strong beats are stressed. The results in this particular case are slightly better, with a median of 4 and a mode of 3.

Furthermore, as both the previous examples, these lyrics use the pattern "*y allí triste cantaba...*". This happens for a series of reasons, including that this pattern has eight

syllables, a common length for musical phrases in Spanish popular music and half the value of *minP*, which becomes closer to *minP* with its variable part (in this case, *moza y menor*); and it can be used for three different semantic relations. More precisely, the original text the pattern was extracted from was *y allí triste cantaba, noche y dia* and, in our semantic network, the words *noche* (night) and *dia* (day) are connected by three different relations, namely:

- *dia* hypernym-of *noche*
- *noche* part-of *dia*
- *noche* antonym-of *dia*

6 Conclusions

This paper presents ETHNO-MUSIC, 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. Additionally, it presents an integration of ETHNO-MUSIC and Tra-La-Lyrics to add lyrics to the melodies also following the style of the Spanish popular music.

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 with the mouse.

Evaluation aimed to demonstrate that both the melody and the lyrics share common features with the original Spanish popular music. For this purpose, a new listening test was performed to retrieve users' opinion on the quality of musical features such as the sound, the rhythm and the melodical behavior, and the quality of the rhythm and semantics of the lyrics. The results revealed positive aspects about the melody and the lyrics.

Although there are some open issues in current state of the art, the main goal of this work was the application of different techniques to create popular songs, which is a novelty by itself. Additionally, the process of generating lyrics from melodies is quite novel, and has been carried out with interesting results. Finally, the use of MMs as a complement to help the users through a musical composition has been scored as quite interesting, as the users feel included in the composition process, and they are an essential part to create new melodies, preserving the Spanish popular music style.

The results also assisted us to identify some limitations that should be addressed in the future, for a new version of the system.

In order to improve the automatic generation of music, it would be interesting to capture other important characteristics of the extracted melodies as they can be either musical phrases or harmony. We can study the possibility of introducing this information in the system to generate more phrasal music. There are few approaches that can be adapted, for example the use of *n*-grams(Whorley and Conklin 2016), or the generation of structural trees that control the tension surface and the phrases (Lerdahl and Jackendoff 1983).

Regarding the music with lyrics, it would be interesting to test singing voice synthesis software for singing the generated songs, which could possibly make the validation clearer. The results have shown the rhythmic patterns in the lyrics could be improved. Therefore, alternative strategies could be tested for matching the rhythm. In a future, an evolutionary strategy, available in the same platform, could be tested for lyrics. However, its fitness function would require new work, as it requires new parameters about music and lyrics, and a new configuration. Moreover, user interaction for the composition of the lyrics may be provided with Co-PoeTryMe (Gonçalo Oliveira et al. 2019), a user interface for the co-creation of poetry and song lyrics with PoeTryMe. It would be just a matter of adding the generated melodies to those available and, possibly, changing the Spanish grammars. By the way, being based on PoeTryMe, Co-PoeTryMe already enables the generation in other languages, namely Portuguese and English.

Given the importance of lyrics in popular music for ethnomusicologist studies, a deeper analysis of popular songs and their lyrics should be addressed in order to improve the vocabulary and semantics of the lyrics generator.

References

Abe C, Ito A (2012) A Japanese lyrics writing support system for amateur songwriters. In: Proceedings of 4th Asia-Pacific signal and information processing association annual summit and conference, Hollywood, CA, USA, APSIPA 2012

Ackerman M, Loker D (2017) Algorithmic songwriting with ALYSIA. In: Computational intelligence in music, sound, art and design - 6th international conference, EvoMUSART 2017,

Amsterdam, The Netherlands, April 19-21, 2017, Proceedings, pp 1–16, https://doi.org/10.1007/978-3-319-55750-2_1
Ajoodha R, Klein R, Jakovljevic M (2015) Using statistical models and evolutionary algorithms in algorithmic music composition, 3rd edn. Encyclopedia of Information Science and Technology. IGI Global, Pennsylvania, pp 6050–6062
Anton TL, Trausan-Matu S (2018) Byzantine music composition using markov models. In: 15th international conference on human computer interaction, RoCHI 2018, Cluj-Napoca, Romania, September 3–4, 2018. Matrix Rom, pp 71–75
Bao H, Huang S, Wei F, Cui L, Wu Y, Tan C, Piao S, Zhou M (2019) Neural melody composition from lyrics. CCF international conference on natural language processing and chinese computing, Springer, LNCS 11838:499–511
Barbieri G, Pachet F, Roy P, Esposti MD (2012) Markov constraints for generating lyrics with style. In: Proceedings of 20th European conference on artificial intelligence, IOS Press, Montpellier, France, ECAI 2012, pp 115–120
Berenson M, Levine D, Szabat KA, Krehbiel TC (2012) Basic business statistics: concepts and applications. Pearson higher education AU
Cambouropoulos E (2001) The local boundary detection model (lbdm) and its application in the study of expressive timing. In: Proceedings of the 2001 international computer music conference, ICMC 2001, Havana, Cuba, September 17–22, 2001. Michigan Publishing
Collins T, Laney R, Willis A, Garthwaite PH (2016) Developing and evaluating computational models of musical style. Artif Intell Eng Des Anal Manuf 30(1):16–43
Colton S, Wiggins GA (2012) Computational creativity: The final frontier? Front Artif Intell Appl 242:21–26. https://doi.org/10.3233/978-1-61499-098-7-21
Colton S, Halskov J, Ventura D, Gouldstone I, Cook M, Perez-Ferrer B (2015) The Painting Fool sees! new projects with the automated painter. In: Proceedings of the 6th international conference on computational creativity, pp 189–196
Concepción E, Gervás P, Méndez G (2018) Afanasyev: a collaborative architectural model for automatic story generation. In: Proceedings of the 5th AISB symposium on computational creativity, University of Liverpool, UK
Delgado M, Fajardo W, Molina-Solana M (2009) Inmamusys: intelligent multiagent music system. Exp Syst Appl 36(3):4574–4580
Gonçalo Oliveira H (2015) Tra-la-lyrics 2.0: automatic generation of song lyrics on a semantic domain. J Artif Gen Intell 6(1):87–110 (Special Issue: Computational creativity, concept invention, and general intelligence)
Gonçalo Oliveira H (2017) A survey on intelligent poetry generation: languages, features, techniques, reutilisation and evaluation. In: Proceedings of 10th international conference on natural language generation, ACL Press, Santiago de Compostela, Spain, INLG 2017, pp 11–20, http://www.aclweb.org/anthology/W17-3502
Gonçalo Oliveira HR, Cardoso FA, Pereira FC (2007) Tra-la-Lyrics: an approach to generate text based on rhythm. In: Proceedings of the 4th international joint workshop on computational creativity, IJWCC 2007, London, UK, pp 47–55
Gonçalo Oliveira H, Hervás R, Díaz A, Gervás P (2017) Multilingual extension and evaluation of a poetry generator. Nat Lang Eng 23(6):929–967. https://doi.org/10.1017/S1351324917000171
Gonçalo Oliveira H, Mendes T, Boavida A, Nakamura A, Ackerman M (2019) Co-PoeTryMe: interactive poetry generation. Cogn Syst Res 54:199–216. https://doi.org/10.1016/j.cogsys.2018.11.012
Gonzalez-Agirre A, Laparra E, Rigau G (2012) Multilingual Central Repository version 3.0. In: Proceedings of the 8th international conference on language resources and evaluation, ELRA, pp 2525–9
Herremans D, Sörensen K (2013) Composing fifth species counterpoint music with a variable neighborhood search algorithm. Exp Syst Appl 40(16):6427–6437
Jhamtani H, Mehta SV, Carbonell J, Berg-Kirkpatrick T (2019) Learning rhyming constraints using structured adversaries. In: Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP), association for computational linguistics, Hong Kong, China, pp 6027–6033, https://doi.org/10.18653/v1/D19-1621, https://www.aclweb.org/anthology/D19-1621
Jungleib S (1996) General Midi. AR Editions, Inc, Middleton
Kuo PH, Li THS, Ho YF, Lin CJ (2015) Development of an automatic emotional music accompaniment system by fuzzy logic and adaptive partition evolutionary genetic algorithm. IEEE Access 3:815–824
Lamb C, Brown DG, Clarke CL (2017) A taxonomy of generative poetry techniques. J Math Arts 11(3):159–179. https://doi.org/10.1080/17513472.2017.1373561
León C, Gervás P (2014) Creativity in story generation from the ground up: nondeterministic simulation driven by narrative. In: 5th international conference on computational creativity, ICCC
Lerdahl F, Jackendoff R (1983) A generative theory of tonal music. The MIT Press, Cambridge
López-Ortega O, López-Popa SI (2012) Fractals, fuzzy logic and expert systems to assist in the construction of musical pieces. Exp Syst Appl 39(15):11911–11923
Machado P, Cardoso A (2002) All the truth about NEvAr. Appl Intell 16(2):101–119
Manzano M (1990) El folklore musical en españa, hoy. Boletín Informativo de la Fundación Juan March 204:3–18
Manzano Alonso M (2001) Cancionero popular de burgos. Diputacion de Burgos
Monteith K, Martinez TR, Ventura D (2012) Automatic generation of melodic accompaniments for lyrics. In: Proceedings of the 3rd international conference on computational creativity, Dublin, Ireland, May 30 - June 1, 2012., ICCC 2012, pp 87–94
Ontanón S, Arcos JL, Puyol-Gruart J, Carasusán E, Giribet D, de la Cruz D, Brito I, del Toro CL (2012) Gena: a case-based approach to the generation of audio-visual narratives. In: International conference on case-based reasoning, Springer, pp 297–311
Pachet F, Roy P (2011) Markov constraints: steerable generation of Markov sequences. Constraints 16(2):148–172
Papadopoulos A, Roy P, Pachet F (2016) Assisted lead sheet composition using flowComposer. In: International conference on principles and practice of constraint programming, Springer, pp 769–785
Pearce MT, Wiggins GA (2007) Evaluating cognitive models of musical composition. In: Proceedings of the 4th international joint workshop on computational creativity, Goldsmiths, University of London, pp 73–80
Pearce M, Meredith D, Wiggins G (2002) Motivations and methodologies for automation of the compositional process. Musicae Scientiae 6(2):119–147
Pérez y Pérez R (2015) A computer-based model for collaborative narrative generation. Cogn Syst Res 36:30–48
Project GM (2018) Google magenta project. https://magenta.tensorflow.org/, Accessed 30 Nov 2018
Ramakrishnan A A, Devi SL (2010) An alternate approach towards meaningful lyric generation in tamil. In: Proceedings of NAACL HLT 2010 2nd workshop on computational approaches to linguistic creativity, ACL Press, Los Angeles, CA, USA, CALC '10, pp 31–39, http://dl.acm.org/citation.cfm?id=1860649.1860654
Schindler K (1991) Música y poesía popular de España y Portugal. Centro de Cultura Tradicional

Scirea M, Togelius J, Eklund P, Risi S (2016) Metacompose: a compositional evolutionary music composer. In: International conference on evolutionary and biologically inspired music and art, Springer, pp 202–217

Serrà J, Arcos JL (2016) Particle swarm optimization for time series motif discovery. Knowl Based Syst 92:127–137

Singh D, Ackerman M, y Pérez RP (2017) A ballad of the mexicas: automated lyrical narrative writing. In: Proceedings of the 8th international conference on computational creativity, ICCC 2017, pp 229–236

Son SH, Lee HY, Nam GH, Kang SS (2019) Korean song-lyrics generation by deep learning. In: Proceedings of the 2019 4th international conference on intelligent information technology, ACM, New York, NY, USA, ICIIT '19, pp 96–100, https://doi.org/10.1145/3321454.3321470,

Speer R, Chin J, Havasi C (2017) Conceptnet 5.5: An open multilingual graph of general knowledge. In: Proceedings of 31st AAAI conference on artificial intelligence, San Francisco, California, USA, pp 4444–4451

Toivanen JM, Toivonen H, Valitutti A (2013) Automatical composition of lyrical songs. In: Proceedings of 4th international conference on computational creativity, The University of Sydney, Sydney, Australia, ICCC 2013, pp 87–91, http://www.computationalcreativity.net/iccc2013/download/iccc2013-toivanen-toivonen-valitutti.pdf

Watanabe K, Matsubayashi Y, Inui K, Nakano T, Fukayama S, Goto M (2017) LyriSys: an interactive support system for writing lyrics based on topic transition. In: Proceedings of the 22nd international conference on intelligent user interfaces, ACM, New York, NY, USA, IUI '17, pp 559–563, https://doi.org/10.1145/3025171.3025194

Watanabe K, Matsubayashi Y, Fukayama S, Goto M, Inui K, Nakano T (2018) A melody-conditioned lyrics language model. In: Proceedings of the 2018 conference of the North American chapter of the association for computational Linguistics: human language technologies, Volume 1 (Long Papers), ACL Press, pp 163–172, https://doi.org/10.18653/v1/N18-1015, http://aclweb.org/anthology/N18-1015

Whorley RP, Conklin D (2016) Music generation from statistical models of harmony. J New Music Res 45(2):160–183

Whorley RP, Wiggins GA, Rhodes C, Pearce MT (2013) 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

Williams D, Hodge V, Gega L, Murphy D, Cowling P, Drachen A (2019) Ai and automatic music generation for mindfulness. In: Audio engineering society conference: 2019 AES international conference on immersive and interactive audio, audio engineering society

Zhu H, Liu Q, Yuan NJ, Qin C, Li J, Zhang K, Zhou G, Wei F, Xu Y, Chen E (2018) Xiaoice band: A melody and arrangement generation framework for pop music. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining, ACM, pp 2837–2846

Znidarsic M, Cardoso A, Gervás P, Martins P, Hervás R, Alves AO, Gonçalo Oliveira H, Xiao P, Linkola S, Toivonen H, Kranjc J, Lavrac N (2016) Computational creativity infrastructure for online software composition: a conceptual blending use case. In: Proceedings of the 7th international conference on computational creativity, ICCC 2016, pp 371–379