Poem sonification: from words to notes

Caroline-Cristina Stere
University Politehnica of Bucharest
313 Splaiul Independentei, Bucharest, Romania
caroline.stere@gmail.com

Ștefan Trăușan-Matu
University Politehnica of Bucharest
313 Splaiul Independentei, Bucharest, Romania
and
Research Institute for Artificial Intelligence
and
Academy of Romanian Scientists
stefan.trausan@cs.pub.ro

ABSTRACT
Poetry and music are two art forms that complete each other, as is the case of songs, opera, and the Ancient Greek art of reciting poems with lyre accompaniment. Although it seems natural to extract one from the other, there has been little research to automate this process. The purpose of the research presented in this paper is to generate a pleasing and expressive melody that matches the rhythm and the emotions of a poem. The connection between verses and melody is vital to creating music that is evoking and consistent. For this reason, we extracted certain properties of the poem (the accent of the syllables and the sentiment of the verses) and mapped them to musical properties (rhythm, scale, and tempo). To mimic the natural pattern of human communication, that mainly consists in requests and replies, we applied the musical technique of call and response, where every musical phrase comes as an answer to the previous one.

Author Keywords
Poem sonification; sentiment analysis; lexicon; poem rhythm; call and response.

ACM Classification Keywords
H.5.5. Information interfaces and presentation (e.g., HCI): Sound and Music Computing.

General Terms
Experimentation; Algorithms.

INTRODUCTION

The profound connection between poetry and music is not a novelty, as the ability to evoke feelings, images and actions through music has been put into practice for centuries. The Italian composer Antonio Vivaldi published his famous violin concerts *Four Seasons* together with a series of four sonnets, one for each season. Each verse describes in words the events of the musical passages: the song of the birds, the murmur of the streams, the roaring of the thunderstorms, the dance of the peasants, etc. This kind of highly expressive and evocative music, accompanied by descriptive verses, is called *programme music*, i.e. music with a strong narrative element, that closely follows a certain theme [1].

In this paper we try to get one step closer to automatically creating programme music based on a poem. Without having the vanity to think that our music comes close to Vivaldi's concerts, we created a functional prototype that takes a poem, analyzes it and generates a piano melody.

This paper has, first of all, an artistic purpose, to bring joy to those who love poetry and wish to enjoy it under a different form. Since the feelings and ideas transmitted by a poem are universal, they can be transposed to music, as well as to other art forms, as painting or even dancing.

Our paper also has a more practical purpose, as it can lead to a wide range of applications, e.g. generating soundtracks for movies (based on the script) or for video games (based on the player's actions).

Our approach starts by analyzing the poem's rhythm, represented by the sequence of stressed and unstressed syllables. The rhythm of the poem determines the rhythm of the melody and the duration of the notes. Then it is determined the emotion of each verse, by using a lexicon-based method, which, in its turn, determines the scale of its corresponding musical phrase, either a major or a minor one. Major scales sound more cheerful, while minor scales have a melancholic vibe. Finally, based on these features of the poem, the musical technique of call and response is applied to compose a pleasing melody that tries to capture and evoke the poem's pulse.

While the approach presented in this paper is an original one, there are a few other attempts to sonify other kinds of literary texts. Among these we mention *TransProse* that sonifies a novel based on the connection between emotions and musical elements (tempo, scale) [2].

LEXICON-BASED SENTIMENT ANALYSIS

Each verse of a poem may express a different emotion, and the corresponding musical phrase must transmit that change accurately. To determine the polarity of each verse, we used the linguistic rules described by Cambria and Hussain [3, pg. 80] and implemented a lexicon-based algorithm that computes the polarity of a phrase. Lexicon-based methods

use a lexicon (a dictionary of word polarities) to compute the emotion of a text, based on some rules.

To illustrate the algorithm we have implemented, we will use the following example [3, pg. 81]: *This car is very old, but it is rather not expensive*. The only words here that convey polarity are *old* and *expensive*, and the valence modifiers (constructions that modify the polarity of a word) are *very, but, rather* and *not.*

One of the basic lexicon-based methods for computing polarity is to identify the words that convey polarity, separately sum their positive and negative polarities and, based on these values, decide the polarity of the phrase [4, pg. 391]. But this means to ignore both the structure of the phrase, and the words that don't have intrinsic polarity, which leaves only the words *old* and *expensive*. As these are both negative words, the polarity of the above phrase would be negative, which is clearly wrong.

However, if the structure of the phrase (i.e. the syntactic dependency tree) is considered, this can be reinterpreted as an electrical circuit, where the signal (i.e. the polarity) passes from one element to another and in the end outputs the phrase polarity. If the words that do not contribute to the polarity are removed, we end up with a circuit like the one in Figure 1 [3, pg. 81].

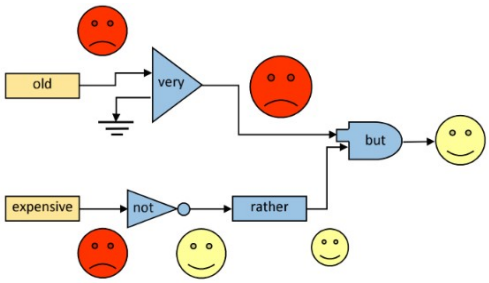

Figure 1. The polarity flow through the sentence's "circuit"

Algorithm description

To determine the polarity (a value between -1 and 1) of a phrase, only the polarities of nouns, verbs, adjectives and adverbs are considered. Negations are also identified in the dependency tree, to invert the polarity of negated words. For example, in the sentence *The dog is not sick*, the polarity of *sick* is multiplied by -1.

Everything seems alright, but let us consider the sentence *She never feared adversities*, that contains the following dependencies:

- The negation (feared, never)
- The direct object (feared, adversities)

Clearly the sentence is positive, but only the word *feared* is negated and receives a positive polarity. The word *adversities* will still have a negative polarity and so, the sentence might end up as negative. To deal with these cases, the polarity of words that depend on negated words is also inverted.

Another element that receives special treatment is the adversative conjunction *but*: everything before it is ignored and only the part that comes after it is considered. For example, the previous sentence *This car is very old, but it is rather not expensive* is positive because the first part is ignored and only *it is rather not expensive* is considered.

Figure 2 shows the resulting algorithm.

```
phrasePolarity = 0
foreach word in phrase:
    if (word is verb, substantive, adjective or adverb):
        if (word is negated) or (word depends on another negated word):
            phrasePolarity += -1 * polarity(word)
        else:
            phrasePolarity += polarity(word)
    else if (word is but):
        phrasePolarity = 0
```

Figure 2. The algorithm for determining the phrase polarity

The proposed algorithm correctly interpreted phrase *She never knew happiness* as negative, and *They are poor, but they never complain about it* as positive.

Implementation details

To get the polarity of words, the *SenticNet* API was used [5]. A call to the API looks like this: http://sentic.net/api/LANGUAGE/concept/CONCEPT_NAME/polarity/intensity, where:

- LANGUAGE encodes the concept's language.
- CONCEPT_NAME is either a word or an expression.
- The returned value is a floating number between -1 and 1, where -1 means completely negative and 1 means completely positive.

For the text analysis (splitting the text into sentences, extracting the words and their lemmas, part of speech tagging, generating the syntactic dependency tree) *Simple CoreNLP* was used, a simplified version of Stanford CoreNLP [6].

THE LYRICS-MELODY CONNECTION

To create expressive and captivating music, it is mandatory to build a tight connection between the melodic phrases and the lyrics, so that the melody both follows and introduces the words. One way of doing this is by associating basic properties of the words with certain musical properties. For example, the famous composer of movie and theatre music, Stephen Sondheim, chooses the rhythm of a melody based on the natural inflections of speech [8]. This strategy is also supported by books on musical composition [9].

A research conducted by Indiana University in collaboration with Microsoft Research studied a collection of popular songs and identified several patterns of the lyrics-melody connection [7]. The authors investigate the idea that prominent and meaningful words are correlated with prominent musical passages, by analyzing the lyrics-

melody connection from a corpus of hundreds of popular songs, belonging to various genres (pop, rock, jazz, etc.).

The authors only considered songs with a 4/4 meter (four quarter-note beats per measure). They parsed each sheet, coded in MusicXML, containing the mappings between each syllable and the corresponding musical note, and built a table with these note-syllable pairs.

A note is described by the following properties:

- *Position* in the 4/4 measure: downbeat (1st beat), half-beat (3rd beat), quarter-beat (2nd and 4th beat) and eighth-beat (half of each beat).
- *Melodic peak*: whether the note is higher than the notes surrounding it.
- *Relative duration*: whether the note is longer than the average duration of notes in that song.

A syllable is described by the following properties:

- *Accent*: primary, secondary or none.
- *Stop word*: whether the syllable is a monosyllabic stop word (words that do not carry content, for example "the", "to", "a", etc.).
- *The contained vowel*: long, short or diphthong (two vowels).

The study found that the syllables have the following effects on the note properties.

The accent of the syllable

Stressed syllables have higher chances of being mapped to stronger beats (downbeat and half-beat), while unstressed syllables receive weaker beats (quarter-beat, eighth-beat and off beat). Also, stressed syllables are often mapped to melodic peaks and longer notes.

Stop words

As about stop words, the authors started from the idea that words that convey little semantic information get associated with weaker, less prominent notes. And indeed, the corpus showed that, just as unstressed syllables, stop words are more likely to be mapped to weaker beats and lower notes.

The contained vowels

The vowel of a syllable influences the note duration, as it has a direct effect on the artist's performance. The authors studied the link between the phonetic length of a vowel and the average duration of the corresponding note and discovered that shorter vowels are mapped to shorter-than-average notes, while longer vowels and diphthongs receive longer-than-average notes.

CALL AND RESPONSE IN MUSIC

Music, just like speech, is more appealing and entertaining as a conversation, rather than a monologue. Just as dialogues have calls (questions) and responses (answers), music also has phrases acting as calls, that induce an incomplete feeling and the necessity of an answer, and responses that bring the satisfying resolution. Without even realizing, we are in a constant search for this pattern in music, and it is only natural, as dialogue is such a great part of our lives.

Call and response were used in the Western classical music of the 16th and 17th centuries, to solve the problem of delays between the choirs seated in opposing lofts of the basilica. Composers got around this by making the choirs sing contrasting phrases, one at a time [10].

The call and response technique may be also found in the African culture, where it was used as a means for the public to participate in rituals or civic meetings. It is also used in gospel music, when the pastor sings a verse (the call) and the church members respond. This has a great effect on people's mood, as it brings them joy and energy [10].

Call and response is a beautifully simple and yet powerful technique. No wonder that it is used in almost any musical style, from gospel and classical to folk, pop, rock, electronic and many others.

There are various methods of creating call and response phrases. Some of these are [11]:

- The response is a repetition of the call.
- The response is a repetition of the call but with a lower last note. Let's take as example the sentences *Music is a cure?* and *Music is a cure.* We notice that even though the words are the same, the meaning is different because our voices raised the pitch at the end of the question.
- The call acts as a statement and the response is a short phrase that confirms it.

But music is not the only place where we see this technique. If some verses are considered, such as the verses of the *Strangers in the night* tune of Frank Sinatra, pairs of calls and responses are noticed, of statements that leave us hanging, followed by statements that provide more information:

Call: Strangers in the night
Response: Exchanging glances
Call: Wondering in the night
Response: What were the chances

POEM SONIFICATION

Any musical composition is made up of two basic elements: rhythm and melody. Therefore, our approach to generate the composition follows two main steps, each of them dedicated to one of these elements.

The rhythm

First, the rhythm of each verse is generated. The rhythm is a sequence of musical durations, without associating musical notes (like A, B, C, etc.) to them. The verses of a poem have a well-defined rhythm, conveyed by the sequence of stressed and unstressed syllables, which is used to create a

direct link between the rhythm of the poem and that of the composition. Therefore, verses with a similar rhythm will generate similar musical phrases. Considering the results of the study [7] discussed in *The lyrics-melody connection* section, unstressed syllables are mapped to a shorter duration (eighth note), stressed syllables to a longer duration (quarter note), and the last syllable of the verse (that usually tends to be sung longer, whether or not it is stressed) to an even longer duration (half note).

The melody

After generating the rhythm for each verse, it is time to map these bare durations to actual musical notes. To this aim, we apply the call and response technique. For every call, a new sequence of musical notes is generated, while the response is a transposition of the call, as it was discussed in the *Call and response in music* section.

The polarity of each verse is analyzed using the algorithm described in the *Lexicon-based sentiment analysis* section. Based on the verse's polarity, the scale of that verse's musical phrase (be it call or response) is chosen. If the verse is positive, a major scale is selected, as major scales sound joyful. If the verse is negative, a minor sale is chosen, as minor scales sound more melancholic.

When the call is generated, a few basic rules of music theory are followed. The first note of the call will always be the tonic (i.e. the first note of the scale). The next notes are randomly chosen from the notes of the scale, but in a stepwise motion, i.e. the move is done by adjacent scale notes instead of leaps. For example, let us say we are in the C major scale (composed of notes C, D, E, F, G, A and B) and the last note generated was E. The next note is selected only from D and F.

The response is a transposition of the call, but what kind of transposition depends on whether the call and response have the same polarity. Before generating the melody, two scales are chosen: a major one (for happy verses) and a minor one (for sad verses).

- If the call and response have different emotions, then the response will be a repetition of the call, but with the notes transposed in the other scale.
- If the call and response have the same emotion, then the response would be an exact repetition of the call. In this case, the call is repeated in the same scale, but lowering the last note, as described at the end of the *Call and response in music* section.

Apart from the polarity of each verse, the polarity of the whole poem is also considered, as it influences the tempo of the melody. A negative polarity will lead to a slower tempo and a positive polarity to a faster one.

CONCLUSION

In this paper we aimed at extracting the song from a poem, that song that, without words, speaks everything the poem speaks. But such a song is tied to the poem with bonds that don't just reveal themselves. In our approach we managed to find a couple of them, but many still remain.

We extracted the rhythm of the poem from the sequence of stressed and unstressed syllables and transformed this rhythm into that of the song. We saw that a poem is a series of calls and responses and applied this pattern to the song. We computed the polarity of the whole poem and decided on the tempo. We also computed the polarity of each verse and decided on the scale of that musical phrase.

As a future development, we intend to strengthen the connection between the song and the poem by applying more of the patterns described in *The lyrics-melody connection* section. We also plan to consider the role of each note in a scale when generating the call, and also diversify the techniques of generating responses.

REFERENCES

1. Niecks, F. (2009). *Programme Music in the last four centuries*. Ardent Media.
2. Davis, H. and Mohammad, S. M. (2014). Generating music from literature.
3. Cambria, E. and Hussain, A. (2015). *Sentic Computing: A Common-Sense-Based Framework for Concept-Level Sentiment Analysis* (1st edition). Springer.
4. Jurafsky, D. and Martin, J.H. (2018). *Speech and Language Processing: An Introduction to Natural Language Processing, Computation Linguistings, and Speech Recognition* (3rd edition draft).
5. Sentic API. Retrieved from https://sentic.net/api/
6. Simple CoreNLP. Retrieved from https://stanfordnlp.github.io/CoreNLP/simple.html
7. Nichols, E., Morris, D., Basu, S. and Raphael, C. (2016). Relationships between lyrics and melody in popular music.
8. Secrest, M. (1998). *Stephen Sondheim. A Life.* New York, New York: Dell Publishing.
9. Peterik, J., Austin, D. and Lynn, C. (2010). *Songwriting for Dummies (2nd edition).* Indianapolis, Indiana: Wiley Publishing.
10. What Is Call and Response in Music? (2019, February 7). Retrieved from https://www.masterclass.com/articles/what-is-call-and-response-in-music#what-is-call-and-response-in-african-music
11. Treseler, S. (2016, January 18). 5 types of Call and Response Phrases. Retrieved from https://stevetres.com/2016/01/5-types-of-call-and-response-phrase