Automatical Composition of Lyrical Songs

Jukka M. Toivanen and **Hannu Toivonen** and **Alessandro Valitutti**
HIIT and Department of Computer Science
University of Helsinki
Finland

Abstract

We address the challenging task of automatically composing lyrical songs with matching musical and lyrical features, and we present the first prototype, M.U. Sicus-Apparatus, to accomplish the task. The focus of this paper is especially on generation of art songs (lieds). The proposed approach writes lyrics first and then composes music to match the lyrics. The crux is that the music composition subprocess has access to the internals of the lyrics writing subprocess, so the music can be composed to match the intentions and choices of lyrics writing, rather than just the surface of the lyrics. We present some example songs composed by M.U. Sicus, and we outline first steps towards a general system combining both music composition and writing of lyrics.

Introduction

Creation of songs, combinations of music and lyrics, is a challenging task for computational creativity. Obviously, song writing requires creative skills in two different areas: composition of music and writing of lyrics. However, these two skills are not sufficient: independent creation of an excellent piece of music and a great text does not necessarily result in a good song. The combination of lyrics and music could sound poor (e.g., because the music and lyrics express conflicting features) or be downright impossible to perform (e.g., due to a gross mismatch between pronunciation of lyrics and rhythm of the melody).

A crucial challenge in computational song writing is to produce a coherent, matching pair of music and lyrics. Given that components exist for both individual creative tasks, it is tempting to consider one of the two following sequential approaches to song writing:

- First write the lyrics (e.g. a poem). Then compose music to match the generated lyrics.

Or:

- First compose the music. Then write lyrics to match the melody.

Obviously, each individual component of the process should produce results that are viable to be used in songs. In addition, to make music and lyrics match, the second step should be able to use the result from the first step as its guidance. Consider, for instance, the specific case where lyrics are written first. They need to be analyzed so that matching music can be composed.

Several issues arise here. The first challenge is to make such a modular approach work on a surface level. For instance, pronunciation, syllable lengths, lengths of pauses, and other phonetic features related to the rhythm can in many cases be analyzed by existing tools. The composition process should then be able to work under constraints set by these phonetic features, to produce notes and rhythmic patterns matching the phonetics. Identification of relevant types of features, their recognition in the output of the first step of the process, and eventually generation of matching features in the second step of the process are not trivial tasks.

The major creative bottleneck of the simple process outlined above is making music and lyrics match each other at a deeper level, so that they jointly express the messages, emotions, feelings, or whatever the intent of the creator is. The pure sequential approach must rely on analysis of the lyrics to infer the intended meaning of the author. Affective text analysis may indicate emotions, and clever linguistic analysis may reveal words with more emphasis. However, text analysis techniques face the great challenge of natural language understanding. They try to work backwards from the words to the meaning the author had in mind. In the case of composing music first and then writing corresponding lyrics, the task is equally challenging.

Fortunately, in an integrated computational song writing system, the second step can have access to some information about the creative process of the first step, to obtain an internal understanding of its intentions and choices. Figuratively speaking, instead of analyzing the lyrics to guess what was in the mind of the lyricist, the composer looks directly inside the head of the lyricist. We call this approach *informed sequential* song writing (Figure 1). In this model, information for the music composition process comes directly from the lyrics writing process, as well as from text analysis and user-given input.

In this paper we study and propose an instance of the informed sequential song writing approach. The presented system, M.U. Sicus-Apparatus, writes lyrics first and then composes matching music. Since lyrics generation is in this approach independent of music composition, our emphasis will be on the latter. Empirical evaluation of the obtained results is left for future work.

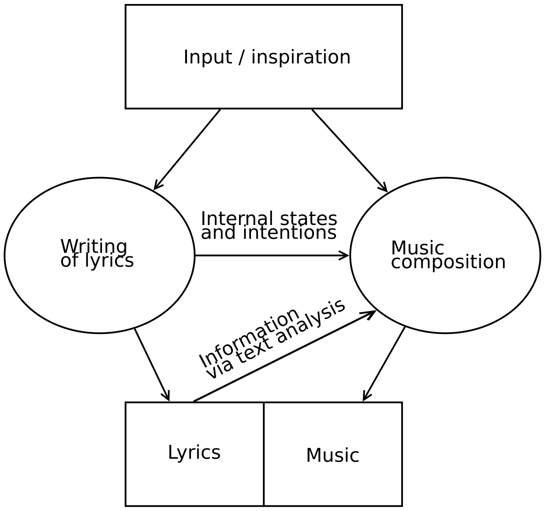

Figure 1: Schema of the informed sequential song generation

Art Songs

Songs can be divided in rough categories like art, folk, and pop songs. This paper concentrates on the genre of so called art songs which are often referred to as lieds in the German tradition or mélodies in the French tradition. Art songs are a particularly interesting category of compositions with strong interaction of musical and lyrical features. Finest examples of this class include the songs composed by F. Schubert. Art songs are composed for performance, usually with piano accompaniment, although the accompaniment may be written for an orchestra or a string quartet as well. [1]

Art songs are always notated and the accompaniment, which is considered to be an important part of the composition, is carefully written to suit the overall structure of the song. The lyrics are often written by a poet or lyricist and the music separately by a composer. The lyrics of songs are typically of a poetic, rhyming nature, though they may be free prose, as well. Quite often art songs are through-composed which means that each section of the lyrics goes with fresh music. In contrast, folk songs and some art songs are strophic which means that all the poem's verses are sung to the same melody, sometimes possibly with little variations. In this paper, we concentrate on through-composed art songs with vocal melody, lyrics, and piano accompaniment.

Related Work on Music and Poetry Generation

Generation of music and poetry on their own right have been studied separately in the field of computational creativity and there have been a few attempts to study the interaction of textual and musical features (Mihalcea and Strapparava 2012). Some attempts have also been made to compose musical accompaniments for text (Monteith et al. 2011; Monteith, Martinez, and Ventura 2012). Interestingly however, generation of lyrical songs has received little attention in the past. Because of the lack of earlier work on combining music and lyrics in a single generative system, we next briefly review work done in the area of music and poetry/lyrics generation separately.

Song Generation

Composing music algorithmically is an old and much studied field. Several different approaches and method combinations have been used to accomplish this task (Roads 1996). One of the most well-known examples, usually known as Mozart's Musikalisches Würfelspiel, dates back to the year 1792, long before modern computers. Many musical procedures such as voice-leading in Western counterpoint can be reduced to algorithmic determinacy. Additionally algorithms originally invented in other fields than music such as L-systems, fractals, constraint based methods, Hidden Markov Models, and conversion of arbitrary data like electro-magnetic fields into music, have been used as the basis for music composition. A review of the approaches used in algorithmic music composition is outside the scope of this paper. For example, Roads (1996) presents a good overview of different methodologies.

Monteith et al. (2012) have proposed a model of generating melodic accompaniments for given lyrics. This approach concentrates on the extraction of linguistic stress patterns and composition of a melody with matching note lengths and fulfilment of certain aesthetic metrics for musical and linguistic match. Differently from this approach, our system composes all aspects of a song including the lyrics, harmony, and melody, and thus it is not limited to the musicalization of existing lyrics. It also employs an informed-sequential architecture and thus the integration of lyrics writing and music composition subprocesses is tighter.

Poetry or Lyrics Generation

A number of approaches and techniques exist for automatic generation of poetry (Manurung, Ritchie, and Thompson 2000; Gervás 2001; Manurung 2003; Toivanen et al. 2012). Some systems have also been proposed to be used for generating song lyrics (Ramakrishnan, Kuppan, and Devi 2009) and not only pure poetry. Again, a review of the approaches used to produce poetry or lyrics automatically is outside the scope of this paper.

Informed Sequential Song Generation

The lyrics part of the song contains the denotational content of the song and partly some connotational aspects like word choices and associations. In the current implementation, the lyrics are written about a user-specified theme (e.g. life) (Toivanen et al. 2012). The music composition module, on the other hand, conveys only connotational information: in the current implementation mood and intensity of the song. The mood is a user-specified input parameter, currently sad or happy, respectively corresponding to positive or negative

[1] Sometimes songs with other instruments besides piano are referred to as vocal chamber music and songs for voice and orchestra are called orchestral songs.

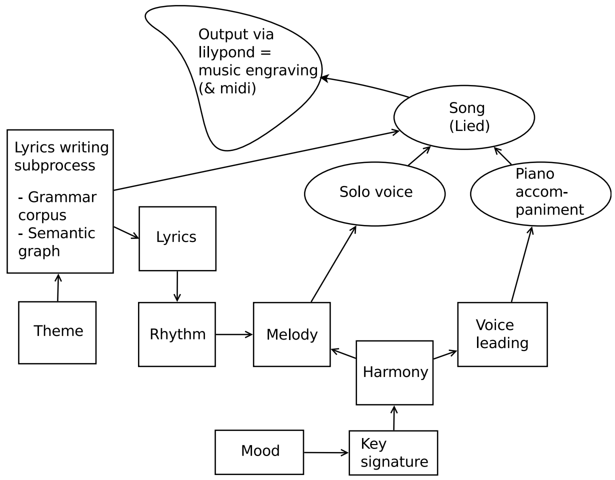

Figure 2: Detailed structure of M.U. Sicus-Apparatus

emotional valence. Intensity corresponds to the emotional arousal to be expressed in the song. It comes from the lyrics writing process and illustrates how internal information of creative processes can be passed between the subprocesses. It is also used as a way to direct the attention to the words expressing the input theme.

We employ the informed sequential song generation scheme with the overall flow of Figure 2. First, the user provides a theme (e.g., snow) and mood (e.g., happy) of the song. M.U. Sicus-Apparatus then generates lyrics for the song that tell about the given theme. The rhythm of the melody is composed by a stochastic process that takes into account the number of syllables, syllable lengths, and punctuation of the lyrics. The harmony of the song is generated either in a randomly selected major (for happy songs) or minor (for sad songs) according to the user's input. Next we discuss each of these phases and the overall structure of M.U. Sicus-Apparatus in more detail.

Lyrics Generation

The lyrics for a new song consist of a verse of automatically generated poetry. Typically a theme for the song is given by the user, and the method then aims to provide a new and grammatically well structured poem with content related to the theme. For lyrics generation, we use the method of Toivanen et al. (2012). We give a short overview of the methodology here.

The lyrics generation method is designed to avoid explicit specifications for grammar and semantics, in order to reduce human effort in modeling natural language generation. Instead of explicit rule systems, the method uses existing corpora and statistical methods. One of the reasons behind this approach is also to keep language-dependency of the methods small. The system automatically learns word associations to model semantic relations. An explicit grammar is avoided by using example instances of actual language use and replacing the words in these instances by words related to a given theme in suitable morphological forms.

As the lyrics writing module is writing lyrics for the song it subsitutes varying proportions of words in a randomly selected piece of text by new words (Toivanen et al. 2012). This proportion can vary between 0% and 100% for every individual line of lyrics although we required the overall substitution rate to be over 50% in the experiments for this paper. The arousal level of the song in a particular place is determined by this substitution rate as discussed in the Dynamics section.

Music Generation

As an overview, M.U. Sicus-Apparatus works as follows. The system first generates a rhythm for the melody, based on the phonetics of the lyrics already written. A harmonical structure is then generated, followed by generation of a melody matching the underlying harmony. A piano accompaniment is generated directly from the harmony with additional rules for voice leading and different accompaniment styles. Finally the resulting song is transformed to a music sheet and a midi file. We next discuss each of the phases in some more detail.

Affective Content Affective connotation has a central role in the overall process. It is provided by the combination of two elements. The first one is the emotional valence, expressing the input mood via harmony and melody. The second element is intensity, expressing emergent information of the lyrics writing process (i.e. word replacement rates, see below).

Rhythm of the Melody The rhythm generation procedure takes into account the number of syllables in the text, lengths of the syllables, and punctuation. Words in the lyrics are broken into syllables and the procedure assigns for every word a rhythmic element with equally many notes as there are syllables in the word. These rhythmic elements are randomly chosen from a set of rhythmic patterns usually found in art songs so that in addition to the number of syllables also the syllable lengths constrain the set of possible candidates. Longer syllables get usually longer time values and shorter syllables get usually shorter time values. The punctuation mark is often stressed with a rest in the melody rhythm.

Harmony The harmony is composed according to the user-specified mood. If the valence polarity of the mood is positive the key signature is constrained to major and then randomly selected from the set of possible major keys. In the opposite case the key is selected from the set of minor keys.

The system database contains different sets of harmonic patterns regularly found in diatonic western classical music for major and minor keys. The construction of harmony is based on a second-order Markov-chain selections of these harmonic patterns and expression of these as chord sequences in a given key. A typical harmonic pattern is, for instance, the chord sequence *I*, *IV*, *V*. When dealing with minor keys, harmonic minor scale is used. The harmony generation procedure also assigns time values for each of the chords in a probabilistic manner so that the length of the generated harmonical structure matches the length of the melody rhythm generated earlier. Usually each chord is assigned a time value of either half note or a whole note. After generating the sequence of chords, the method moves on to determine pitches of the melody notes.

Melody The melody note pitches are generated on the basis of the underlying harmony and pitch of the previous note by a random walk. Firstly, the underlying chord defines a probability distribution for pitches which can be used. For example, if the underlying chord is C major as well as the key signature, the notes *c*, *e*, and *g* are quite probable candidates, *a*, *f* and *d* are less probable and *h* is even less probable. Secondly, the pitch of the previous note affects the pitch of the next note in a way that small intervals between these two notes are more probable than large intervals. Finally, the note pitch is generated according to a combined probability distribution that is a product of the probability distribution determined by the underlying chord and the probability distribution determined by the previous melody note.

Accompaniment and Voice Leading The harmonical structure provides the basic building blocks of the accompaniment but the chord sequence can be realised in many styles. Currently, we have implemented several different styles like Alberti bass and other chord patterns.

In order to have smooth transitions between chords in the accompaniment, we apply a simple model of voice leading. For a given chord sequence our current implementation chooses chord inversions that lead to minimal total movement i.e. smallest sum of intervals, of simultaneous voices.

Dynamics The arousal level of the song is expressed as dynamic marks in the music. Higher arousal is associated with higher loudness (e.g. forte) and lower arousal is associated with more peaceful songs (e.g. piano). For every line of lyrics this proportion of substituted words (S) in a line of poetry is expressed in the music either as piano (*p*, $S < 25\%$), mezzo-piano (*mp*, $25\% < S < 50\%$), mezzo-forte (*mf*, $50\% < S < 75\%$), or forte (*f*, $75\% < S < 100\%$).

Output The system outputs both sheet music to be performed by musicians and a midi file to be played through a synthesizer. The music engraving is produced with the LilyPond music score language (Nienhuys and Nieuwenhuizen 2003).

Examples

Figures 3 and 4 contain two example songs generated by M.U. Sicus-Apparatus[2]. The song in Figure 3 is a sad one about life, and the one in Figure 4 is a happy song about flower buds. The words that have been emphasised by the lyrics writing process are marked in bold in lyrics.

The proposed methodology seems to provide relatively good combinations of text and music. As explained above, the transmission of information on song dynamics comes directly from the lyrics writing process. This is interesting because that particular information would be impossible to extract directly from the lyrics itself. For instance, in the

[2]These and other songs are available also in midi form at http://www.cs.helsinki.fi/discovery/mu-sicus

Figure 3: Excerpt of a sad song composed with the theme “life” (in Finnish “elämä”).

song of Figure 4, first two phrases have a very high arousal (forte) due to the high emphasis on the overall song theme whereas after that the arousal calms down.

Taking the syllable lengths and punctuation into account in the rhythm seems to lead to quite singable melodies (our subjective view which needs to be evaluated objectively in later work). However, taking the syllable stress into account as well could lead to further improvements.

The melody, harmony, and rhythm seem to constitute quite a coherent whole. The major weakness in the music is a lack of clear phrase structures, which has also been a problem in many music generation systems before. The lyrics writing method has been evaluated earlier with good results by Toivanen et al. (2012).

Conclusions and Future Work

We have proposed the task of generating lyrical songs as a research topic of computational creativity. This topic has received only little attention in the past although both music composition and poetry/lyrics generation have been studied on their own.

As a first step towards a generative music-lyrics model we have implemented a system, M.U. Sicus-Apparatus, that

Figure 4: Example of a happy song composed with the theme "flower buds" (in Finnish "nuput").

generates simple lyrical art songs. The current system composes happy and sad songs about a given theme by writing first lyrics of the song and composing then music with a melody rhythm that matches the phonetic structure of the lyrics. The system works in an informed-sequential manner. This means that writing of the lyrics and composition of the music are not performed separately but the lyrics writing module can convey part of its internal data structures directly to the music composition system.

An automatical generation procedure of lyrical and musical content also offers interesting possibilities for musicalization of data (Tulilaulu et al. 2012). For example, converting news stories automatically into songs could be an interesting application of the presented methodology.

In the future, we would like to carry out an empirical evaluation of the methods and results using recorded performances of a collection of songs. As a further step, we would like to study how emotional states can be transferred in songs, partly by conveying the affective state in the lyrics and partly by modifying the tempo, loudness, modality, melody movements, and rhythm of the music. A wide body of research exists on correlation of musical features with perceived affective states. For example, varying the note rate in the accompaniment and using dissonance and consonance as well as different instrumental techniques like staccato to convey intensity, could be used to improve the system.

The ultimate goal is to break the inherently sequential structure of the architecture, and to develop a song generation system with a much tighter integration or interaction between the lyrics writing and music composition processes.

Acknowledgements

This work has been supported by the Algorithmic Data Analysis (Algodan) Centre of Excellence of the Academy of Finland under grant 118653.

References

Gervás, P. 2001. An expert system for the composition of formal spanish poetry. *Journal of Knowledge-Based Systems* 14(3–4):181–188.

Manurung, H. M.; Ritchie, G.; and Thompson, H. 2000. Towards a computational model of poetry generation. In *Proceedings of AISB Symposium on Creative and Cultural Aspects and Applications of AI and Cognitive Science*, 79–86.

Manurung, H. M. 2003. *An Evolutionary Algorithm Approach to Poetry Generation*. Ph.D. Dissertation, University of Edinburgh, Edinburgh, United Kingdom.

Mihalcea, R., and Strapparava, C. 2012. Lyrics, music, and emotions. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2012)*, 590–599.

Monteith, K.; Francisco, V.; Martinez, T.; Gervás, P.; and Ventura, D. 2011. Automatic generation of emotionally-targeted soundtracks. In *Proceedings of the Second International Conference on Computational Creativity*, 60–62.

Monteith, K.; Martinez, T.; and Ventura, D. 2012. Automatic generation of melodic accompaniments for lyrics. In *Proceedings of the International Conference on Computational Creativity*, 87–94.

Nienhuys, H. W., and Nieuwenhuizen, J. 2003. Lilypond, a system for automated music engraving. In *Colloquium on Musical Informatics (XIV CIM 2003)*.

Ramakrishnan, A.; Kuppan, S.; and Devi, S. L. 2009. Automatic generation of tamil lyrics for melodies. In *Proceedings of the Workshop on Computational Approaches to Linguistic Creativity (CALC'09)*, 40–46.

Roads, C. 1996. *The Computer Music Tutorial*. The MIT Press.

Toivanen, J. M.; Toivonen, H.; Valitutti, A.; and Gross, O. 2012. Corpus-based generation of content and form in poetry. In *International Conference on Computational Creativity*, 175–179.

Tulilaulu, A.; Paalasmaa, J.; Waris, M.; and Toivonen, H. 2012. Sleep musicalization: Automatic music composition from sleep measurements. In *Eleventh International Symposium on Intelligent Data Analysis (IDA)*, volume 7619 of *LNCS*, 392–403.