The Muses of Poetry - In search of the poetic experience

Diana Arellano and Simon Spielmann and Volker Helzle¹

Abstract. This paper presents the current advances in "The Muses of Poetry", an on-going project that combines interaction, emotions and poetry. The goal of the project is to create an interactive installation where a virtual character not only recites poetry, but also manifests the affective content of the poem through facial expressions and voice rhythm. The novelty of our work is the combination of different elements like human computer interaction, semantic analysis, affective computing and character animation, which intend to create a unique poetic experience. These elements are presented in the first part of the paper. In the second part, we do a more in-depth analysis of all the aspects that would be necessary to have virtual characters that read poems out loud, while conveying their intrinsic emotional meaning.

1 INTRODUCTION

There is no doubt that poetry is one of the most creative forms of literary expression. A good poem is not only capable of eliciting mental images and awake feelings in the reader, but its rhythm and melody can transport this reader to the world created by the poet.

This is reflected in the words of Emily Dickinson, who defined poetry in the following way "If I read a book and it makes my whole body so cold no fire can warm me I know that is poetry. If I feel physically as if the top of my head were taken off, I know that is poetry. These are the only way I know it. Is there any other way?" (L342a, 1870).

However, one thing is to read poetry for oneself and another thing is to read it out loud. The latter presents a major challenge because the reader needs to be aware of the style of the poem, the pauses, the melody and the ideas that this poem intends to convey.

Having these elements in mind, we posed the question: is it possible to have a virtual character that reads poetry? Furthermore, can this virtual character make an audience connect with a poem, with its words and meanings?

Based on our previous experience with interactive virtual characters, we decided to tackle these questions by creating an interactive and audiovisual poetry installation. The work we present in this paper has been carried out within the framework of the on-going project "The Muses of Poetry", a media installation which main objective is to bring people closer to poetry.

One of the novelties of our work is the way poetry is addressed. It is not our intention to automatically generate poems, or just to have a character reading poetry. Instead, the installation allows users to interact with a virtual character who expresses the underlying affective content of each poem. Moreover, the project would also give poets from around the world the opportunity to expose their poems more publicly and to a wider audience. Another contribution is the mixture of different areas like semantic analysis, real-time computer graphics, voice generation and humancomputer interaction, to create very differentiated virtual characters that engage the user in a poetic experience.

Nevertheless, the quest for more intelligent, affective and creative characters raises a number of issues from a technological and somehow philosophical point of view. In the following sections, we present the current state of our project as well as the aspects that need to be taken into account to provide our characters with intelligence and creativity.

2 RELATED WORK

The majority of the existent works that combine poetry and artificial intelligence deal with "how to generate written poetry" instead of "how to automatically read poetry". Examples of these are Colton et al. [4], who came up with a corpus-based poetry generation system that constructs poems according to a given rhyme, sentiment, word frequency and similarity; David Cope, who created the program "Alena" (Artificial Life Evolving Natural Affinities) to automatically write haiku, which were subsequently published in the ebook "Comes the Fiery Night" [5]; Pablo Gervás, who created WASP, a reasoning rule-based system that takes as input a set of words in Spanish and verse patterns and returns a set of verses [7]; or Toivanen et al. [19], who made used of text mining methods, morphological analysis, and morphological synthesis to generate poetry in Finnish.

Other works, more related to our research, are the ones of Tizhoosh and Dara [17] and Tizhoosh et al. [18], which focused on the analysis of text in order to distinguish between poem and prose, without understanding or interpreting the underlying poetic meaning. Similarly, there have been a number of researchers working on assessing the style of a poem, either to use it as a tool for the study of different types of poems and see how they affect the reader's perception of the poem [10], or to figure out what makes a poem beautiful [9].

As for the use of interactive virtual characters, Naoko Tosa was one of the pioneers in this area with her interactive installation "MUSE" [20]. In MUSE, poems are created by exchanging poetic phrases between the user and the system, represented by a character which facial features are eyes, eyebrows and mouth. In Tosa's work, the facial expressions of the character change according to the emotions conveyed in the phrases uttered by the user. In the same direction of interactive systems, Kwiatek and Woolner [12] merged poetry into interactive storytelling based on still and video panoramas. The aim of their application was to develop interest not only in the life of the poet Charles Causley but also in his literary output.

In the field of emotional speech synthesis, works like [2], [15], or [1] have tried to achieve expressiveness in the synthetic voice. Nevertheless, it still remains an open issue.

¹ Institute of Animation, Filmakademie Baden-Wuerttemberg, Germany, email: diana.arellano@filmakademie.de

3 OVERVIEW OF THE SYSTEM

As we mentioned in Section 1, The Muses of Poetry is an interactive installation where a virtual character manifests the intrinsic emotionality of a poem through facial expressions and affective speech.

In this section, we present a brief technical overview of the implementation of the modules that are part of the system behind The Muses of Poetry (see Figure 1).



Figure 1. Modules of the system behind The Muses of Poetry

3.1 Words Visualization and Voice Recognition

The interaction with The Muses of Poetry begins when the user says three words aloud, which are recognized by the system allowing it to select a poem that contains those three words.

The Words Visualization and Voice Recognition module uses and implements the elements that allow the interaction of the user with the system: a Microsoft Kinect, the Word Cloud Display submodule, and the Voice Recognition submodule.

The Kinect array microphone is the device that captures the voice of the user, which is sent as input to the Voice recognition submodule. The reason for using Kinect and not other type of microphone is to provide the user with a very natural way of interaction (i.e. free of cables as if in a person-to-person conversation).

In the current implementation of The Muses of Poetry, we use the Microsoft Speech Platform as the speech recognition engine. Therefore, we are able to define in the Kinect the *grammar*, or set of words that should be recognized by the system. The grammar consists of all the words from all the poems in the repository with a length over 3 characters. Thus words like "a", "the", "or", and so on, would not be recognized.

In order to let the user know which words he or she can say, these are displayed in a "word cloud" arrangement, which is generated dynamically in every interaction. The first time the user approaches the installation a word cloud of the 30 most frequent words is shown. After a word is said and recognized, the cloud shrinks showing the most frequent words in those poems that contain the previously recognized word. This process is repeated once more, after which the control is passed to the Poem Selector module. Figure 2 shows an example of interaction with the installation.



Figure 2. Example of interaction

3.2 Poem Affect Analysis

An important step that is performed prior to the interaction with the installation is the affective analysis of the poems. In principle, this is done just once for each poem that is added to the repository.

To extract the affective information of each poem we carry on a semantic analysis, aided with the Whissell's Dictionary of Affect in Language (WDAL) [6]. This dictionary includes 10,368 English words with affective connotations, where each one is described with regard to the dimensions of Activation (or Arousal) and Evaluation (or Pleasantness). Two reasons led us to decide for the Whissell's dictionary and not for others like ANEW [14] or WordNet-Affect [16]. First, Whissell's dictionary has been created with words from literary and poetic texts. The other two contemplate mostly affective words, which are not enough to assess the whole meaning of a poem. Second, the number of words contained in Whissell's dictionary is greater than in ANEW and Wordnet-Affect. As a result, the majority of words used in poems are mostly recognized by Whissell's dictionary and not by the other two.

The WDAL itself operates as a licensed stand-alone application that assesses the affective information of a poem in terms of activation, evaluation and imaginary dimensions. It also provides a detailed classification of each word in the poem according to the following states: pleasant, nice, fun, passive, sad, unpleasant, nasty, active, high imagery and low imagery. Currently, we only consider the pleasant, nice, fun, sad, unpleasant and nasty states.

Nevertheless, if we associate emotions only to words, then it might happen that the emotional connotation of the poem is distorted. A way to avoid this is to tag the poem using a more general structure than the word, but smaller than a stanza (i.e. a unit within a larger poem). This would be the lines of the poem, which are the ones that convey the ideas the poet wanted to express. As a result, from the global analysis of all the words in the poem, we depict if it is emotionally positive or negative. In a second analysis, we break down the poem in lines. Then, the prevailing emotion in the words of a line is the one that is applied to it. For instance, if a line contained more unpleasant words and the poem was assessed as negative, then that line would be tagged as "unpleasant".

After this process is completed, the poems with their corresponding emotional tags are stored in the repository.

3.3 Poems Selector

This module is the one that accesses the poem repository, and based on the words recognized from the user selects the poem to be performed. The selection of the poems is done in a top-down approach. At first, the selector considers all the available poems in the repository, extracts their words and sends them to the Voice Recognition and Word Cloud module. When the first word is recognized, the selector searches for it in the current set of poems, keeping only the poems that contain that word. Moreover, the selected word is stored, so it does not appear again in the word cloud. This process is repeated twice more, guaranteeing that the poems in the final set contain the three recognized words.

Finally, if the final set contains more than one poem, one of them is picked randomly and sent to the Virtual Character module.

3.4 Virtual Character

This module has been divided into two submodules very closely related: the facial animation module and the speech synthesis module.

3.4.1 Dynamic Facial Animation

One of the main characteristics of The Muses of Poetry is the dynamic generation of facial animations. In this way, there is no need of having pre-rendered animations for each poem. It represents a great advantage because any new poem can be added into The Muses without effort on the animation side. The development framework that provides this functionality is named Frapper², created at the Institute of Animation, Filmakademie Baden-Wuerttemberg.

In order to produce dynamic animations, first a set of pose animations are created in the third-party software Maya, and then exported to Frapper in the form of files containing the references to the object and its animation properties. Currently, these animations correspond to universal emotions like "anger", "sadness", "disgust" and "joy"; to idle states like "waiting" or "thinking"; or to "confirmation" of a recognized word. Being this the first prototype, we preferred to keep a reduced set of emotions, which can be enhanced in future iterations of The Muses. Figure 3 shows two of the expressions manifested by the current female character.



Figure 3. (Left) Expression of "sadness". (Right) Expression while "waiting"

To generate a poem animation, Frapper takes into consideration the emotional tags in the text of the poem. For each tag there is a number of associated animation files, which are randomly chosen by Frapper, providing variability to the poem animation.

Another important feature of Frapper is automatic lip-syncing. It is accomplished by using the visemes information provided by the speech synthesizer. Thus, having previously defined the poses for each of the selected visemes (i.e., A, O, E, CH, I, F, M), these are triggered when they are recognized in the text to be recited.

Additional rendering elements provided by the framework are tears simulation, eye redness, and wrinkles generation.

3.4.2 Speech Synthesis

To create the voice of the character we used a third-party voice synthesizer, provided by SVOX³.

It produces not only a more natural voice, but also the visemes information required by Frapper to generate automatic lip-sync. Similarly, it allows us to control the pitch and speed of the voice, which would produce variations that enhance the emotionality of the voice. In the current version of The Muses of Poetry these features are not yet exploited, consequently, the emotional content of the poem is manifested mostly through facial expressions.

4 Philosophical aspects of a Virtual Poetry Reader

As seen in the previous sections, The Muses of Poetry is a combination of multiple disciplines and technologies used to achieve an engaging poetic experience. Nevertheless, when showed to students and colleagues, a number of issues were pointed out regarding the intelligence and creativity of the character and the installation. Therefore, the question we asked ourselves is: what do we need to make our character more intelligent and creative?

After an in-depth research, we found that the answer to this question poses a number of issues. Even though they are planned for future work, they make us think of all the implications when creating an Affective Virtual Poetry Reader. These thoughts are discussed in the following subsections.

4.1 Perception

"What people think when interacting with The Muses of Poetry?"

This is the query that came to our mind when we showed it to an audience for informal testing purposes. It is worth noting that so far we have not performed any perceptual test that would give us qualitative and quantitative results, given that this is still a work in progress. However, we tried to show occasionally the installation to an audience formed mainly by animation and interactive media students, computer scientists and poets, in order to get their feedback.

The answers have been as varied as the audience that responded, showing the importance of perception when dealing with an installation of this kind. One of the more recurrent topics was the voice, which will be explained in a further subsection.

The visualization of the character was another important issue, with opinions regarding its abstraction, its expressiveness, or the way it should be displayed. The majority of the persons, including poets, reacted positively to the facial expressions, agreeing that it conveyed emotionality. Nonetheless, they were also distracted by the effect of what they called "robotic" voice. There was one case where a person manifested the opposite reaction. He found the voice completely appropriate but the expressions subtle.

From these comments, we realized that the expectations of regular users when faced with a realistic human character make them want to see and hear a real person reciting and expressing the affect in a poem. That leads to the problem that any artefact, however small, diminishes the whole experience.

² http://frapper.animationsinstitut.de/

³ http://www.nuance.de/products/SVOX/index.htm

Given than the modification of the synthesized voice to make it sound human-like (i.e. with the nuances and expressiveness of natural voice) was not possible at this stage; and as a matter of fact, it remains a challenge in the Speech Synthesis field, we decided to change the facial appearance of the character. By moving away from a realistic character into a more abstract one, we wanted to see if the "robotic" voice could be accepted as part of the character, and not as a failure in the character.

To address this topic, we asked a group of students from the Filmakademie Baden-Wuerttemberg to come up with more abstract concepts of virtual characters. One of the designs is shown in Figure 4, a virtual head made of particles that swirl around, producing the effect of a constantly moving mask.



Figure 4. Concept design for an abstract character

The speech in this abstract character was made by simply opening and closing its mouth, according to the words of the recited poem. To have an idea of the effect that it would produce, we presented it to a small audience with experience in animation. Again the reactions where very dissimilar. On the one hand, some of them were fascinated by the swirling effect. On the other hand, some expected to see facial expressions during the interaction.

After this second experiment, we concluded that when working with abstraction and imagination, it is very difficult to use one type of character to target all audiences. Imagination and perception play a crucial role, hence the need to address these two points. One possibility in this direction would be to have a character that could change according to the user's perception of poetry, making it a first step into creating intelligent "adaptable" characters. In the near future, we will allow the participants to choose from a number of characters, in a way to personalize the experience.

4.2 Semantic Analysis

As explained in Section 3.2, currently we are using an affective dictionary to perform the semantic analysis of the poems and extract their affective content. Previous works have dealt with text analysis of poetry using different techniques, like statistical analysis to assess the poetry style of different poets [11], logic rules and pattern matching to extract metaphors from the poems [8], or classifiers like Bayesian and Multilayer Perceptron to extract poetic features to differentiate prose from poetry [18].

The main challenge in The Muses of Poetry is that all the poems are free verse. It means that they do not use a consistent meter pattern or rhyme, tending to look and sound like prose. Therefore, most of the techniques used in previous works result inefficient.

The next step would be to find a way to automatically extract syntactically well-formed sentences, and group them in a way so they convey one complete meaning. Then, we would need to teach the system how to recognize elements like the affective cues in it, and tag the text accordingly. Breaking the poem into its lines would also give the "space" needed to make the character pause. This would result in a more intelligent affective poetry analyser.

4.3 Poetic Voice

According to Rachel Blau DuPlessis (cited in [13]), poetry "is the kind of writing that is articulated in sequenced, gapped lines and whose meanings are created by occurring in bounded units... operating in relation to pause or silence". From this definition it can be inferred that poetry is sound. Hence the importance of identifying the lines in a free verse poem to know its periodicity and its gaps.

The question is then *how to make our character transmit the sound of poetry?* The answer again is not trivial, specially when there is not a specific way to read poetry to take as guidance. Each poet has its own style, which makes it more difficult for the system to learn a particular "reciting" pattern. If we take as example the style of poet William S. Burroughs or Robert Frost, we could hear that they had a characteristic way of accentuating the words, almost as if they "sing" the poem. However, there are other poets who read in a way that resembles the current style in The Muses of Poetry. One of the reasons might also be that the metrical feet of the poems in the installation is *Iamb*, which it is said is the nature of the English Language.

Another related issue is the emotionality in a computer generated voice. When it comes to emotions in speech, so far the majority of speech synthesizers in the market and in academia offer voices that are not yet expressive enough.

Buckhardt and Stegmann [2] identified the features of the speech signal as "spectral (sound of the voice), prosodic (melody of the speech), phonetic (kind of spoken phones, reductions and elaborations), ideolectal (choice of words) and semantic features". However, achieving speech with the right features that simulate human speech would require a great investment of time and resources. A proof of this is that despite the number of methods proposed for achieving expressive speech synthesis (e.g., rule-based, data-driven, among others), it still remains as a challenging research issue. Thus it is not our intention to create our own speech generator, but to be able to finetune parameters in existent ones that allow us to have a more natural voice with emotions.

A proposed solution was the use of human-recorded voices, but then it would go against the automation achieved in our system. The Muses of Poetry has been conceived as an installation where the character poses the expressions and generates the speech that is more adequate to certain poem.

Basic rules that could be applied to make our characters read poetry with more emotionality and vigour would be:

- Poems come in lines, but pausing at the end of every line will create a choppy effect and interrupt the flow of the poem's sense. Readers should pause only where there is punctuation, just as you would when reading prose, only more slowly [3].
- Contrary to the rule above, sometimes the pause in-between lines are actually pauses, but so short that is like the reader's breath inhale. A possibility would be to identify the correct spacing in-between lines pauses and simulate the inhaling sound.

- Another kind of pause would be before connectors: "and", "but", "or"; or after "so". Nonetheless, this is not a fixed rule and does not happen always.
- Some poets recite the end of the line much slower to give emphasis to the meaning of the poem. This can be done with sad or happy poems, but not with unpleasant or funny poems, where probably a raise in the volume would produce the desired effect.

A concluding remark regarding speech synthesis has to do again with the issue of perception. When people see and interact with a realistic character, the expectations are high and any flaw in the representation is taken as an "uncanny" effect. But, what would happen if the character is so abstract that there is no way to link it to a known voice? Would the poetic effect be the same? Would the audience perceive it as a real intelligent entity that analyses the poem and manifests the emotions it conveys?

These are questions that remain open and we intend to investigate and test in the further development of The Muses of Poetry.

5 Discussion and Future Work

The Muses of Poetry is the name of the project that aims to create an interactive installation where emotional poetry "reciters", or muses, transmit not only the words of a poem but also its intrinsic emotional meaning.

The innovative combination of human-computer interaction (HCI), media design, computer graphics, voice synthesis and semantic analysis resulted in the first prototype of The Muses of Poetry. This has helped us to assess which should be the path to follow in order to convey a very engaging poetic experience.

One of the first things we noticed was the way people perceived poetry. For some of them, poetry belongs to the realm of abstraction and imagination. Thus, their expectations when faced with a realistic character reading poetry made them find a number of flaws and elements that were not suitable. The voice of the character being too robotic, the animations being too subtle, or the lack of an environment that separates the audience from the real world, were part of the feedback of the users who tried the installation.

Another element to take into consideration is the need for a more "intelligent" character that would make each interaction unique. This could be achieved, for instance, by adding a mood element to the semantic analysis. Thus poems would be analysed differently depending on the mood of the character. Moreover, this mood could change depending on the interaction with the user, or the amount of sad, happy, or unpleasant poems the character has read so far.

Intelligence would also mean having a more emotional voice that goes according to the affective state of the poem. Currently, our prototype does not consider emotionality in the voice due to technical reasons, but it is intended for future work. The proposal is to link visual and vocal expression in such a manner that, independently of the character, gives the impression of a living entity reciting poetry.

As it can be seen, there is plenty of room for improvement. Many questions are left opened but are intended to be answered in future iterations. So far real-time rendering, automatic lip-sync, natural interaction, and affective semantic analysis have been invaluable tools that allowed us to come to this point. Now it is time to endure the task of giving intelligence to the system to achieve the ultimate goal of transporting the user to the world created by the poet and manifested in The Muses of Poetry.

ACKNOWLEDGEMENTS

This work is funded by the *Innovationsfonds Kunst*, from the Ministerium für Wissenschaft, Forschung und Kunst (MWK) Baden-Württemberg. We also thank Tatiana Pushkareva for the animations of the female virtual character, Amit Rojtblat for the concept and animations of the abstract character, Andreas Hykade for his advises on the visual aspect of the abstract character, and Kent MacCarter for being our poet adviser.

REFERENCES

- A.W. Black, H.T. Bunnell, Ying Dou, P. Kumar Muthukumar, F. Metze, D. Perry, T. Polzehl, K. Prahallad, S. Steidl, and C. Vaughn, 'Articulatory features for expressive speech synthesis', in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pp. 4005–4008, (2012).
- [2] F. Burkhardt and J. Stegmann, 'Emotional speech synthesis: Applications, history and possible future', in *German Conference on Speech Signal Processing, ESSV*, (2009).
- B. Collins. How to read a poem out loud. ONLINE, 2004. http://www.loc.gov/poetry/180/p180-howtoread.html. Last checked: 01/25/2013.
- [4] S. Colton, J. Goodwin, and T. Veale, 'Full face poetry generation', in Proceedings of the Third International Conference on Computational Creativity, (2012).
- [5] D. Cope, Comes the Fiery Night, NY: Amazon Books, 2011.
- [6] P. Duhamel and C. Whissell. The dictionary of affect in language [computer software], 1998.
- [7] P. Gervás, 'Wasp: Evaluation of different strategies for the automatic generation of spanish verse', in *Symposium on Creative & Cultural Aspects and Applications of AI & Cognitive Science*, (2000).
- [8] L. S. Hiyakumoto, *PoEM : a parser of emotion metaphors*, Master's thesis, Massachusetts Institute of Technology. Dept. of Architecture. Program in Media Arts and Sciences, 1999.
- [9] J. Kao and D. Jurafsky, 'A computational analysis of style, affect, and imagery in contemporary poetry', in NAACL Workshop on Computational Linguistics for Literature, (2012).
- [10] D. M. Kaplan and D. M. Blei, 'A computational approach to style in american poetry', in *Proceedings of the 2007 Seventh IEEE International Conference on Data Mining*, ICDM '07, pp. 553–558, (2007).
- [11] David M. Kaplan and David M. Blei, 'A computational approach to style in american poetry', in *Proceedings of the 2007 Seventh IEEE International Conference on Data Mining*, ICDM '07, pp. 553–558, (2007).
- [12] K. Kwiatek and M. Woolner, 'Let me understand the poetry. embedding interactive storytelling within panoramic virtual environments', in EVA 2010, pp. 199–205, (2010).
- [13] B. McHale, 'Beginning to think about narrative in poetry', *Narrative*, 17(1), 11–27, (2009).
- [14] F. A. Nielsen, 'A new anew: evaluation of a word list for sentiment analysis in microblogs', in *Proceedings of the ESWC2011 Workshop* on 'Making Sense of Microposts': Big things come in small packages, eds., Matthew Rowe, Milan Stankovic, Aba-Sah Dadzie, and Mariann Hardey, volume 718 of CEUR Workshop Proceedings, pp. 93–98, (2011).
- [15] M. Schröder, 'Expressive speech synthesis: Past, present, and possible futures', *Affective Information Processing*, 111–126, (2009).
- [16] C. Strapparava and A. Valitutti, 'Wordnet-affect: an affective extension of wordnet', in 4th International Conference on Language Resources and Evaluation (LREC 2004), pp. 1083–1086, (2004).
- [17] H. R. Tizhoosh and R. A. Dara, 'On poem recognition', Pattern Analysis and Applications, 9(4), 325–338, (2006).
- [18] H. R. Tizhoosh, F. Sahba, and R. Dara, 'Poetic features for poem recognition: A comparative study', *Journal of Pattern Recognition Research*, 3(1), (2008).
- [19] J. M. Toivanen, H. Toivonen, A. Valitutti, and O. Gross, 'Corpus-based generation of content and form in poetry', in *Proceedings of the Third International Conference on Computational Creativity*, (2012).
- [20] N. Tosa, 'Interactive poem', in ACM SIGGRAPH 98 Conference abstracts and applications, SIGGRAPH '98, pp. 300-, (1998).