A COMPUTATIONAL METHOD FOR EXPLORING MUSICAL CREATIVITY DEVELOPMENT

A. Alexakis
Dept. of Music Studies
University of Athens
tonisq
@music.uoa.gr

A. Khatchatourov
Sony Computer Science
Laboratory
armen.khatchatourov
@gmail.com

A. Triantafyllaki
Dept. of Music Studies
University of Athens
a_triant
@music.uoa.gr

C. Anagnostopoulou
Dept. of Music Studies
University of Athens
Chrisa
@music.uoa.gr

ABSTRACT

The development of musical creativity using non-standard methods and techniques has been given considerable attention in the last years. However, the use of new technologies in teaching improvisation and thus development of creativity has received relatively little attention to date. The aim of this paper is two-fold: firstly to propose a way of formalising the measurement of creativity, and secondly to test whether the use of a particular interactive system built to support musical improvisational dialogues between the user and the computer (MIROR IMPRO), can develop creativity. First, based on previous research, we define a set of variables aiming at evaluating creativity, and we create a computational model to automatically calculate these variables in order to assess the development of creative abilities. Second, we assess the advancement of creativity in 8-10 year-old children, who spent six weeks interacting with MIROR-IMPRO. We used two groups of children in assessing this advancement: a group of children with no musical background (n=20) and a group of young pianists (n=10). We carried out a free improvisation test before the start and after the end of six sessions with the system. The results suggest a potential progress related to a number of these variables, which could be indicative of creativity advancement. The issue of measuring creativity is discussed in the light of these findings.

1. INTRODUCTION

Creativity is a fundamental human ability, and at the same time a particularly challenging concept to define. Various attempts exist to date, and its meaning tends to shift across the various disciplines. Yet however vague and slippery its definition may be, its core features are shared across domains, which makes it possible to model, and in general to become the subject of scientific investigation.

One of the first attempts to formally describe creativity is found in [26], where *Creativity Thinking* is modeled as a four-step process: *preparation* – information, specific knowledge and ideas about the case/problem under question are gathered, *incubation* – work proceeds unconsciously, *illumination* – suddenly the solution emerges, and *verification* – the solution is verified and elaborated. Another step-wise model suggested in [9] where a five-step approach is proposed in problem solving and creative thinking. The idea of problem solving is also closely related with the eminent contribution of J.P. Guilford in the field. Guilford in [11] introduced the idea of *convergent* and *divergent thinking* and associated the latter with creative thinking.

The above approaches to creativity focus mainly on the processes involved in creative thinking. Another aspect of creativity, closely related with attempts to measure or assess creativity, is focused mainly, but not solely, on the product. Creativity as 'product' is defined by Amabile in [5] as one whereby *"...appropriate observers independently agree it is creative. Appropriate observers are those familiar with the domain in which the product was created or the response articulated"*, hence introducing the idea of how a creative product is received and assessed by (as well as situated in) its environment.

But how can creativity be assessed? Guilford in [10] created a test to measure creativity, by assessing divergent thinking. The subjects were given 180 ordinary life objects (e.g. a pencil, a spoon, a cap) that they were asked to score across four dimensions: originality, fluency, flexibility, and elaboration. Extending Guilford's ideas, Torrance developed the Torrance Tests of Creative Thinking (TTCT) [25], while Amabile proposed the Consensual Assessment Technique (CAT) for ranking the creativity of art objects [5]. CAT is based on the idea that expert judges within a field will have a valid opinion regarding the creativity values of an object of art. Gathering and examining such expert opinions may provide a good estimation of the creative worth of an object. A well described application of CAT can be found in [12].

In the field of music creativity, Webster's work [29] continues to be prominent among scholars. Webster built on Guilford's ideas and created a tool to evaluate the creative aptitude of children (ages 6-8), the *Measurement of Creative Thinking in Music (MCTM)* [27]. The MCTM evolved into MCTM-II in [30]. Children's creative thinking is evaluated through a ten-task session, of about 20-25 minutes. The qualities that are scored are musical expressiveness (ME), musical flexibility (MF), musical originality (MO) and musical syntax (MS) [29]. In the specific field of ethnomusicology, Lomax developed the *"cantometrics"* [14]. They are comprised of a set of 37 items measuring group organization, level of cohesiveness, rhythmic features, melodic features, dynamic features, ornamentation and vocal qualities. Later, McPherson in [15] developed measures to assess a musician's ability to perform music creatively. These new measures are pertained to evaluate music learner's performance from memory, by ear and through improvising.

Simonton in [21] performed computerized content analysis to assess the melodic originality of 15,618 themes of 479 classical composers, from Josquin des Pres to Shostakovich. Simonton defined a number of variables each of which pertain to different qualities of the case under investigation. Similarly in [22] he investigated 1919 compositions of 172 classical composers, spanning almost 500 years. A panel of experts manually scored several of the above variables, prior to the computer analysis.

Regardless of how well they approach the notion of creativity, the above measures require more or less the engagement of (often numerous) human experts in scoring. They also employ statistical averages in order to eliminate human errors and individual particularities.

At the same time, the broad introduction of computer technology in music educational processes created the possibility to computationally automate the whole process. Hence it becomes more and more pertinent to come up with proposals that require no human intervention, even if the range of the investigated qualities is decreased.

The introduction of new music technologies in the educational process involves also the introduction of new interaction paradigms between the user and the machine. An example of new interaction paradigms are Interactive Reflexive Music Systems (IRMS) [18], and in particular the MIROR IMPRO system [20], which was developed within the MIROR project [1] as the evolution of *The Continuator* [2][3][4][17][19]. The core concept in such a system is that basic musical elements can be taught and musical cognitive processes can be developed not only by the traditional teacher/learner dipole but also by the direct interaction of the learner with the system, without the involvement of a human instructor.

The application generates different kinds of output melodies based on the user's musical input, stimulating the reflexive interaction between the user and the application. This generation is based on a specific Markovian mechanism designed by Sony CSL Paris, allowing a meaningful musical output. Namely, the output is composed of what the user could have played herself, i.e. a constrained recombination of musical elements previously played by the user. In this way, each response of the system is composed of musical material close to the user's style, but at the same time proposes to the user to explore, as the next, step, new ways to express musical ideas.This study explores the use of MIROR IMPRO in developing young children's improvisational skills - recognised as a central component of musical creativity [29]. Therefore, it would seem important to develop a methodology of evaluating the creativity that arises as a result of engaging with such a system. This may later be integrated into the system in order to give real time information to the user and to record such information for a traditional trainer/learner session that may subsequently follow. The aim of the paper is to propose a way of measuring creativity in children playing the keyboard; and to use this model in order to assess creativity in children with and without musical background, comparing their pre and post tests (before and after an intervention of 6 improvisation sessions using the MIROR IMPRO system).

The paper is structured in the following way: in the **Methods** section, the technical description of the work is laid out, including the data collection process, the knowledge representation schemata, computational details and the description of the variables used to assess creativity. In the **Results** section, the results of the work are presented and subsequently discussed in the last section (**Discussion**).

2. METHODS

2.1 The Goal

In this section, a model and a computational method to measure creativity is introduced. Specifically, we describe the musical corpus we used, the knowledge representation schema, algorithmic details and particularities, and finally the creativity measuring model, realized as a set of measures/variables.

2.2 Data collection

Within the framework of the psychological experiments related to the MIROR project, a number of children's musical improvisations on a MIDI keyboard were performed. The keyboard was connected to MIROR IMPRO system. Each improvisation session is comprised of a dialogue of music phrases that are alternately human and machine generated. Each of these phrases is recorded onto a different MIDI channel and thus it becomes straight-forward to extract all human phrases.

The data we used comes from two experiments with MIROR IMPRO and young children - one with non-musicians and one where children had been studying the piano from between 1-4 years.

The reasoning behind this sampling is the following. In our initial work with non-musicians we found that the keyboard as an object (rather than the interaction with the system itself) seemed to draw the attention of the children. We then introduced a second sample of children who were already familiar with the keyboard, as a way to eliminate the effect the keyboard may have on the interaction and hence the musical output from this interaction. In this paper we present the analysis from both groups of children.

The study with the young pianists took place in a small music school and involved 10 children (six girls and four boys) playing alone with the MIROR IMPRO system for six weeks (that is six sessions of 15 to 20 minutes). The study with the non-musicians took place in a primary school and involved 20 children (sixteen boys and four girls) playing with MIROR IMPRO across six weeks, in similar conditions. In both studies we proceeded to conduct a pre- (before the six weeks) and post-test (after the six weeks) with the children. This consisted of asking each child individually to improvise a short tune (1-2 minutes long) on the keyboard.

We compare the pre-test sessions to the post-test sessions of both the young pianists' and the non-musicians' sessions, in order to find out if their creativity developed by their post-test session. In this way, we might begin to attribute such development to their in-between sessions where they interacted with the MIROR IMPRO system, in order to explore further the use of IRMS in the development of children's musical improvisations and creativity.

The young pianists pre-corpus consists of 5218 note events having duration of 2,359,916 msecs. The post-corpus consists of 2427 note events having duration of 662,627 msecs. The non-musicians pre-corpus consists of 8990 note events having duration of 2,022,753 msecs. The post-corpus consists of 6477 note having duration of 1,030,853 msecs.

2.3 Knowledge Representation

The concept of a symbolic musical corpus raises the issue of music knowledge representation. Having in mind the data manipulation task, the viewpoint representation formalism was chosen to be used [8], as it offers great flexibility in surfacing the attributes of the musical objects. It also offers a direct and straight forward representation on corresponding data structures. The concept of viewpoint is lately gaining popularity among researchers, due to its capability to capture in a well-defined representation set of symbols, a big variety of the musical features of musical data.

The musical object on which a viewpoint is defined can here be a single note or a sequence of notes, viz. a segment. Here the notion of a segment is used to describe the whole melody played by the child.

On the note level, several viewpoints were calculated: pitch (as MIDI number), pitch class, onset, duration, ioi (interon-set time interval), trail (time interval between a NOTE OFF event and the consecutive NOTE ON), fnitioid (time interval from first note in track), seqint (melodic interval – pitch distance from previous event), contour (rising: 1, static: 0, falling: -1) and several others.

Segmental viewpoints [7] are also constructed. For each segment a set of segmental viewpoints is calculated, such as the number of notes in the segment, the duration etc.

Segmental Viewpoint	**Description**
sd[seq]	Standard deviation of sequence seq
uniq_patt[seq]	Number of unique patterns in sequence seq
diff_patt[seq]	Number of different patterns in seq
tot_patt[seq]	Number of total patterns in seq
Avg_sise[seq]	Average size in number of note events of seq
Avg_dur[seq]	Average duration
Tot_size[seq]	Total size in number of note events of seq
Tot_dur[seq]	Total duration
Inteval (small,medium,large) [seq]	Percentages of interval divisions
Note (small,medium,large) [seq]	Percentages of pitch divisions
Rhythm (small,medium,large) [seq]	Percentages of rhythm divisions
velocity (small,medium,large) [seq]	Percentages of dynamic divisions
Texture[seq]	Measures how “thick” is the music
Cluster[seq]	Number of chords in seq

Table 1. Segmental viewpoints used.

2.4 Computational Processing

The computation proceeds by reading one by one all MIDI files in a directory (a directory with MIDI files is considered a corpus) and building from the corresponding MIDI events a sequence of viewpoints. Consecutively, repeated patterns within each viewpoint sequence are extracted.

Thus, the identification of patterns can be seen as a problem within the stringology domain. As such, in order to identify common patterns suffix arrays [16] are employed. Suffix arrays provide an easy to implement and fast way to locate each and every common substring within a string. In [24], suffix arrays technique proves its capability and its efficiency on a much larger corpus.

For constructing the suffix array, the well-known QuickSort comparison sort algorithm is used in this work. The suffix array can be scanned and common patterns can be reported, along with their frequency, their length and their locations within the corpus.

2.5 Creativity Variables

In order to assess creativity we used a set of variables that we calculated for each subject, for the improvisation tests that took place before and after the training. The idea of assessing creativity through a set of metrics (realised as variables) is drawn directly from the creativity literature, as most of the scholars are proposing to measure creativity based on a set of measures, scored by one or more experts. Our aim is to come up with a set of metrics that are scored automatically, eliminating thus the need of experts. As evidenced in the creativity literature, we assume that advancement in musical variation and diversity is an indicator of musical creativity.

The following variables were used:

V1 – Standard Deviation. Standard deviation is a metric on how much away from the average falls most of the values. A low standard deviation means that data tend to be close to the average. We calculate this for the sequence of three viewpoints – MIDI numbers, intervals and rhythmic values. It indicates the diversity of the musical vocabulary.

V2 – Number of patterns with frequency 1. We identify all sequences of the 3 viewpoints (notes, intervals, rhythmic values) that appear only once in the corpus. We borrowed this idea from the lexical analysis in [23], as it seems to indicate novelty and musical variety. Suffix arrays make straight forward the identification of those patterns, since we count the number of rows in the array that has no common with their next.

V3 – Average Size, Duration. The idea of this indicator is taken from Webster's MCTM [27][28]. We calculate two variants of this variable. First, we calculate the segmental viewpoints size (in number of notes) and duration (in msecs) for each subject. Then we calculate the average of all segments per subject. Second, we calculate the total size and total duration for each subject.

V4 – Ratio of different per total patterns. This variable is drawn by analogy from lexical content analysis in psychotherapy [13] and is used also in [23]. There are evidence that the greatest the ratio of different words per total words the greatest the lexical diversity [13]. So we assume that the greatest the above ratio the greatest the musical variability and hence the musical creativity. We identify all sequences of the 3 viewpoints (notes, intervals, rhythmic values)

V5 – Interval Variation. This is an indicator on musical intervals diversity. We calculated the segmental viewpoint `interval(small, medium, large)`. Then we calculate for each subject's music (viz. each MIDI file) the percentages of small, medium and large intervals. We assume that small intervals are less than 4 steps and large ones more than 8 steps – a step is a semitone.

We assume that the more evenly distributed the percentages are that more variation we have. This applies also to V6, V7 & V8

V6 – Pitch Variation. We calculated the segmental viewpoint `note(low, medium, high)`. Then we calculate for each subject's music the percentages of low, medium and high pitches. We assume that low pitches are below F3 (MIDI number 53) and high ones over C#5 (MIDI number 73).

V7 – Rhythm Variation. We calculated the segmental viewpoint `rhythm(slow, medium, fast)`. Then we calculate for each subject's piece of music the corresponding percentages. We assume that medium rhythmic values are with the notes that has more or less the quarter note duration; that is 500 msecs for our MIDI files. Hence we take +/- 10% of that for identifying the slow and fast rhythms.

V8 – Dynamics Variation. We calculated the segmental viewpoint `velocity(soft, normal, hard)`. For identifying the dynamics of notes we take into consideration the velocity recorded along with the notes within the MIDI file. The velocity takes values in [0, 127] range. We calculate for each subject's music the percentages, similar to the above variables. We assume the piano range lays below velocity value of 40 and the forte one above 60.

V9 – Texture Richness. For all notes in each subject's corpus we sum up their duration. Then we divide the duration of each piece of music with the total duration of all notes. The more notes we have (and the more lengthy they are) the less the value of V9 will be. It indicates how much populated with notes the music is.

V10 – Clusterness. For each segment we calculate the number of simultaneities. It is an indicator of the number of chords and consequently the richness of harmony produced. A simultaneity occurs when a `"note on"` MIDI event is transmitted while others `"note on"` events are still alive.

3. RESULTS

Table 2 reports the mean values on pre and post conditions for the two groups, non-musicians and musicians. The general trends indicate advancement in creativity when we compare mean values on pre and post sessions.

	Non-musicians		Musicians	
	Pre	Post	Pre	Post
V1 pitch SD	10.75	13.16	8.84	9.65
V1 interval SD	10.08	10.75	9.36	9.24
V1 rhythm SD	0.93	0.97	15.11	19.84
V2 unique pitch	23.90	30.00	20.3	17.8
V2 unique interval	39.70	40.3	27.5	24.9
V2 unique rhythm	23.85	24.15	46.4	40.0
V3 Nb notes / segmented	48.70	48.42	42.62	29.42
V3 duration /segmented	12324	7598	25299	9822
V3 Nb notes / total	449.5	323.85	521.8	242.7
V3 duration/ total	101138	51543	235992	66263
V4 different pitch	0.35	0.37	0.29	0.31
V4 different interval	0.32	0.35	0.25	0.35
V4 different rhythm	0.29	0.30	0.31	0.38
V5 variation interval small	57.87	59.00	50.45	49.92
V5 variation interval medium	15.30	18.13	25.05	25.09
V5 variation interval large	26.82	22.79	24.50	24.98
V6 variation pitch low	13.85	20.09	12.25	15.62
V6 variation pitch medium	58.30	50.71	55.35	55.00
V6 variation pitch high	27.84	29.20	32.40	29.37
V7 variation rhythm slow	12.22	11.60	69.99	53.60
V7 variation rhythm medium	4.42	3.52	7.13	10.35
V7 variation rhythm fast	83.36	84.90	22.88	36.05
V8 variation dynamics soft	37.26	15.59	14.76	8.11
V8 variation dynamics normal	27.30	14.93	31.13	26.89
V8 variation dynamics hard	35.44	69.49	54.10	64.99
V9 texture richness	0.89	0.70	1.35	0.66
V10 clusterness	17.43	21.60	19.56	26.39

Table 2. Variables mean values for non-musicians and musicians, on pre and post session.

However, due to a small sample size and limited number of treatment sessions, not all of shifts are statistically significant.

The pre – post treatment comparison was performed with asymptotic Wilcoxon signed rank test with Pratt zero handling (with *coin* package in *R* software [31]). The two groups were assessed in a separate manner, so that no direct statistical comparison between groups was made.

The tables below report only statistically significant differences between pre- and post-conditions, for the variables not reported below no significant difference was found. For variables V1, V2, V4, V5medium, V6 we predicted greater values in post session. i.e. greater values indicating the progress of creativity. For variables V5small and V5large we predicted smaller values in post session (see the explanation in the **Discussion** section). Accordingly, a one-tailed test was used for these variables. For variables V3, V7, V8, V9, V10 no directional hypothesis was made. Accordingly, a two-tailed test was used.

3.1 Non-musicians

	MEAN	STD DEV	MEDIAN
Pre	10.75	3.34	10.87
Post	13.16	2.88	13.72
Z = -2.65, p-value = 0.004 (one-tailed)			

Table 3. V1 – Standard Deviation on pre- and post-corpus.

As seen in Table 3, the average pitch SD was higher in the post-session than in the pre-session, indicating that greater variety in the notes used.

	MEAN	STD DEV	MEDIAN
Pre	101137.65	36301.93	96031.50
Post	51542.65	19238.46	49255.00
Z=3.40, p-value=0.001 (two-tailed)			

Table 4. V3 – Duration, total.

As it can be seen from Table 4, the average total duration was almost two times shorter in the post-session than in the pre-session.

	MEAN	STD DEV	MEDIAN
Pre	15.30	6.51	16.20
Post	18.13	6.00	18.45
Z = -1.75, p-value = 0.039 (one-tailed)			

Table 5. V5 – Percentages of medium intervals

As it can be seen from Table 5, the average medium intervals were more present in the post-session than in the pre-session.

	MEAN	STD DEV	MEDIAN
Pre	37.26	25.40	29.98
Post	15.59	12.32	11.93
Z = 2.65, p-value = 0.008 (two-tailed)			

Table 6. V8 – Dynamics Variation, soft.

As it can be seen from Table 6, on the average, "soft" dynamic was more than two times less present in the post-session than in the pre-session.

	MEAN	STD DEV	MEDIAN
Pre	27.31	9.11	28.06
Post	14.93	9.58	14.07
Z = 3.06, p-value = 0.002 (two-tailed)			

Table 7. V8 – Dynamics Variation, normal.

As it can be seen from Table 7, on the average, "normal" dynamic was more two times less present in the post-session than in the pre-session.

	MEAN	STD DEV	MEDIAN
Pre	35.44	24.67	34.40
Post	69.49	19.54	70.40
Z = -2.99, p-value = 0.003 (two-tailed)			

Table 8. V8 – Dynamics Variation, hard

As it can be seen from Table 8, on the average, "hard" dynamic was more than two times more present in the post-session than in the pre-session.

	MEAN	STD DEV	MEDIAN
Pre	0.89	0.26	0.86
Post	0.70	0.07	0.72
Z = 3.92, p-value = 0.001 (two-tailed)			

Table 9. V9 – Texture Richness

As it can be seen from Table 9, on the average, the musical excerpt played by the child is more "populated" in the post-session than in the pre-session (smaller values of this variable reflect more "populated" excerpt).

3.2 Musicians

	MEAN	STD DEV	MEDIAN
Pre	235991.60	111207.17	257527.50
Post	66262.70	31756.15	57980.50
Z= 2.60, p-value = 0.009 (two-tailed)			

Table 10. V3 – Duration, total

As it can be seen from Table 10 average total duration was more than three times shorter in the post-session than in the pre-session

	MEAN	STD DEV	MEDIAN
Pre	0.25	0.06	0.26
Post	0.35	0.07	0.38
Z = -2.29, p-value = 0.021 (two-tailed)			

Table 11. V4 – Ratio of different per total, intervals.

As it can be seen from Table 11, the average ratio of different intervals was higher in the post-session than in the pre-session.

	MEAN	STD DEV	MEDIAN
Pre	22.88	6.51	16.20
Post	36.05	22.17	31.60
Z = -2.09, p-value = 0.037 (two-tailed)			

Table 12. V7 – Rhythm variation, fast.

As it can be seen from Table 12, the average percentage of fast rhythm was almost twice higher in the post-session than in the pre-session.

	MEAN	STD DEV	MEDIAN
Pre	1.35	0.66	1.21
Post	0.66	0.04	0.68
Z = 2.80, p-value = 0.005 (two-tailed)			

Table 13. V9 – Texture Richness.

As it can be seen from Table 13, on average, the musical excerpt played by the child is almost twice more “populated” in the post-session than in the pre-session (smaller values of this variable reflect more “populated” excerpt).

4. DISCUSSION

Both musicians and non-musicians improvised on the keyboard. In general, it was observed that musicians, who were keyboard players, improvised by creating musical sequences based on their previously known pieces. Non-musicians, who were not familiar with the keyboard, played mostly in the form of gestures, such as upward and downward melodic movement, oscillation between two notes, continuous repetition of a pattern etc. (for more information see [6]).

The students' teachers were supportive of our sample's participation in the study, although their role in the process was not studied nor was the impact of children's participation measured in some way, when they returned to their 'normal' musical activities. A follow-up study may be able to explore this aspect, particularly teachers' perceptions of students' musical skills after having participated in such activities.

Webster in [29] suggests that certain divergent, imaginative skills among others, are also critical to creative thinking, such as musical extensiveness (the amount of time invested in creative imaging), flexibility (the range of musical expression in terms of dynamics, tempo, and pitch) and originality (the unusualness of expression). Our variables explored mostly variance in flexibility, between the pre and the post test.

4.1 Non-musicians

The pre tests and post tests for the players without any musical background show some differences, which could potentially be attributed to the use of the MIROR IMPRO system. More specifically, the standard deviation of the pitches used increases in the post test. This shows that the children start to be more adventurous and explorative in their choice of pitches, using a bigger range of the piano.

While the pitch standard deviation increases, the medium intervals also increase, compared to small and large intervals. This fact could indicate that children stop playing at random, in all the registers (i.e. they don't make huge intervals any more between high and low register), and they avoid repetitions of the same note (i.e. they don't use very small intervals any more). Instead they use intervals that are more or less typically used in music, of medium size.

Another interesting difference between pre and post test is that children play louder, which could indicate a stronger confidence in their playing, and at the same time use more notes in the same amount of time, to create a thicker texture. However, it is interesting that in the post test they also play for significantly less time. This could be seen in two ways: the first suggests that they play in a more focused way, given the above significant results, for less time, while the second proposes that they might be getting tired by the time they reach the post test, and decide to play less.

4.2 Musicians

Before discussing the results of the pianists, there is one fact that needs to be explained in order to better evaluate the results. Children with a background in piano playing, during the pre test, played mainly their known pieces from the piano lesson, and improvised less. Therefore, their pre test has a lot of features that we would normally find in known music. By the time the children reach the post test, all of the children leave the security of the known pieces and prefer to play more freely their own tunes. We believe that this can be attributed to the use of the MIROR IMPRO system, as there was scant interaction with the researcher throughout the study. The post test improvisation session is also significantly shorter. As they played more freely, it could be explained as more focused improvisational playing.

In the post test, their ratio of different per total intervals used is higher, which means that there is less repetition and more originality in their playing. At the same time, pianists play almost twice as fast as in the pre test, which could indicate more confident playing, especially as this is coupled with less soft and timid playing. Like the non-musicians, they also use more notes per unit of time, to create a thicker texture.

4.3 General discussion

The work described here is introducing a model for measuring creativity and creativity development. This model in essence defines and describes musical creativity via a set of attributes realised as distinct variables. While the utilization of a set of variables for describing creativity is something that most of the scholars in the field are employing (see section 1), the appropriateness of a particular variable can always be under question. For example, is it valid to hypothesise that different distribution in the (small, medium, large) range of intervals (that is variable V5) indicates musical creativity advancement? Of course in general, in the borderline cases this hypothesis holds true; for instance if a `interval(95, 3, 2)` tuple is becoming a `interval(40, 40, 20)`, the player is musically exploring a larger interval range and this seems to be consistent with musical creativity development in the literature. But in most in-between cases the extent to which changes in the variables indicates creativity development is open to discussion. In general the concept of creativity evades a clear definition and the issue of assessing creativity development is a challenging topic which can be dealt with in many ways. Future work will include fine tuning of variables, eventually defining significant limits on experimental basis.

5. CONCLUSION

This study firstly proposed a set of variables to measure creativity in music, based on existing literature on creativity assessment, and secondly investigated the development of creative music improvisations of young children, after playing an Interactive Reflexive Music System called the MIROR IMPRO. It drew on two examples, a group of 20 non-musicians and a group of 10 young pianists, and measured the development of their creativity in free improvisation before and after six sessions of using the system.

The non-musicians' post test free improvisations include higher diversity of musical vocabulary, more medium intervals and richer texture, indicating a sensible progress in improvisational creativity. At the same time, they include more intensity in dynamics, indicating more confident playing behaviour. Interestingly this seems also to be the case with the young pianists, as their post tests include similar features. In their post tests, however, there is more use of different intervals with less repetition and faster playing, even though they move away from the familiarity of their known piano pieces by this final session. It can be argued that the differences between pre and post tests observed in the musicians and non-musicians may be due to more than increased familiarity with the keyboard, that is the differences observed may be due to the use of the MIROR IMPRO system to develop creativity.

Further analysis of the in-between six sessions with MIROR IMPRO may provide more ideas regarding the precise variables that seem to shift across sessions in both groups of melodies. Future work also includes the direct comparison of the two groups, to investigate the differences between the young pianists and the children with no musical background, as well as the introduction of a control group to assess an eventual development of creativity without MIROR-IMPRO.

This would allow also fine tuning of the creativity assessment model and its testing in various new settings in order to improve the definition of the variables used, as well as the introduction of new related variables.

Acknowledgments

The work described in this paper forms part of the MIROR European project [1], co-funded by the European Community under the Information and Communication Technologies (ICT) theme of the Seventh Framework Programme. (FP7/2007-2013). Grant agreement n° 258338

6. REFERENCES

[1] **M**usical **I**nteraction **R**elying **O**n **R**eflexion official website: http://www.mirorproject.eu/.

[2] A. R. Addessi and F. Pachet, "Childen's Interaction with a Musical Machine" in Proceedings of the 3rd Conference "Understanding and Creating Music", Caserta, 2003, pp. 11-15.

[3] A. R. Addessi and F. Pachet, "Experiment with a Musical Machine: musical style replication in 3 to 5 years old children" in B. J. Music Ed., 22, 1, 2005, pp. 21-46.

[4] A. R. Addessi, F. Pachet and R. Caterina, "Children Confronting an Interactive Musical System" in Proceedings of 8th International Conference on Music Perception & Cognition (ICMPC8), Adelaide, 2004.

[5] T. M. Amabile, The social psychology of creativity. Springer-Verlag, 1983.

[6] C. Anagnostopoulou, A. Alexakis and A. Triantafyllaki, "A Computational Method for the Analysis of Musical Improvisations by Young Children and Psychiatric Patients with No Musical Background" in Proceedings of 12th International Conference on Music Perception & Cognition, Thessaloniki, 2012

[7] D. Conklin and C. Anagnostopoulou, "Segmental pattern discovery in music" in INFORMS Journal on computing, 18, 3, 2006.

[8] D. Conklin and I. Witten, "Multiple viewpoint systems for music prediction" in New Music Research, 24, 1, 1995, pp. 51-73.

[9] J. Dewey, How We Think. Heath, 1933.

[10] J. P. Guilford, "Creativity" in American Psychologist, 5, 1950, pp. 444–454.

[11] J. P. Guilford, The nature of human intelligence. McGraw-Hill, 1967.

[12] M. Hickey, "An Application of Amabile's Consensual Assessment Technique for Rating the Creativity of Children's Musical Compositions" in J. of Research in Music Education, 49, 3, pp. 234-244.

[13] O.R. Holsti, "Content Analysis" in G.Lindzey & E.Aronson (Eds.), The Handbook of Social Psychology, 2nd ed., vol. II, Amerind Publishing Co, 1968, pp.596-692.

[14] A. Lomax, Cantometrics: A method in musical anthropology. University of California Extension Media Center, 1976.

[15] G. E. McPherson, "The Assessment of Musical Performance: Development and Validation of Five New Measures" in J. Psychology of Music, 23, 1995, pp. 142-161.

[16] U. Manber and G. Myers, "Suffix arrays: a new method for on-line string searches" in SIAM Journal on Computing, 22, 5, 1993, pp. 935-948.

[17] F. Pachet, “Interacting with a musical learning system: the Continuator” in C. Anagnostopoulou, M. Ferrand, and A. Smaill (Eds.), Music and artificial intelligence: Lecture notes in artificial intelligence 244.5, Springer-Verlag, 2002, pp. 119–132.

[18] F. Pachet, “Enhancing individual creativity with interactive musical reflexive systems” in I. Deliège and G. A. Wiggins (Eds.), Musical Creativity: Multidisciplinary Research in Theory and Practice, Psychology Press, 2006, pp. 359–376..

[19] F. Pachet and A. R. Addessi, “When Children Reflect on Their Playing Style: Experiments with the Continuator and Children” in ACM Computers in Entertainment, 2, 1, 2004

[20] F. Pachet, P. Roy and G. Barbieri, “Finite-Length Markov Processes with Constraints” in Proceedings of the 22nd International Joint Conference on Artificial Intelligence, IJCAI, pp. 635-642, Barcelona, Spain, July 2011.

[21] D. K. Simonton, “Thematic Fame, Melodic Originality, and Musical Zeitgeist: A Biographical and Transhistorical Content Analysis”, in J. of Personality and Social Psychology 38, 6, 1980, pp. 972-983.

[22] D. K. Simonton, “The swan-song phenomenon: Last-works effects for 172 classical composers” in J. of Psychology and Aging, 4, 1, 1989, pp. 42-47.

[23] D. K. Simonton, “Lexical choices and aesthetic success: A computer content analysis of 154 Shakespeare sonnets” in Computers and the Humanities, 24, 1990, pp. 251-264.

[24] I. Knopke and F. Jürgensen, “A System for Identifying Common Melodic Phrases in the Masses of Palestrina” in Journal of New Music Research, 38, 2, 2009, pp. 171-181.

[25] P. E. Torrance, Torrance tests of creative thinking. Scholastic Testing Services, 1966.

[26] G. Wallas, The Art of Thought. Harcourt Brace, 1926.

[27] P. R. Webster, “An assessment of musical imagination in young children” in P. Tallarico (Ed.) Contributions to symposium: The Bowling Green State University symposium on music teaching and learning, Bowling Green, 1983, pp. 100–123.

[28] P. R. Webster, “Refinement of a measure of creative thinking in music” in C. K. Madsen & C. A. Prickett (Eds.), Applications of research in music behaviour, Tuscaloosa, AL, 1987, pp. 257–271.

[29] P. R. Webster, “Creativity as Creative Thinking” in Music Educators J. 76, 9, 1990, pp. 22–28.

[30] P. R. Webster, Measure of creative thinking in music-II (MCTM-II). Administrative guidelines. Unpublished manuscript, Northwestern University, 1994.

[31] http://cran.r-project.org/web/packages/coin