To appear in “The future of research”, Sony CSL 20th anniversary

The Future of Content is in Ourselves

François Pachet

Sony CSL-Paris

1 Introduction

The word “Content” is the ugliest word one can imagine to describe the most valuable creation of mankind. Content is everything that is, or flows, inside containers. A strange way indeed to talk about the products of our digital culture: music, films, photographs, books, games; in short everything which is produced for other reasons than necessity. The success of this word is probably related to the paramount and probably excessive importance of distributors in the present state of our society. A view which culminated with the erratic visions of Jean-Marie Messier, during the creation of the Vivendi conglomerate, who explicitly proposed to view objects of creation as free fluids.

However misnamed, content becomes more and more important. As democracy develops and spreads in the world, likewise do wealth, individualism and, as a consequence of this general increase in well-being, interest in, if not dependency on, content, in the form of movies, music, games and media in the large. Internet and mobile communication can only increase this omnipresence of content in our everyday lives. So although a large part of the world still fights against poverty, dictatorship or hunger, the future of content is a key question for our developed societies.

This question has so far been addressed essentially from the viewpoint of distribution and access: Recent years have seen a strong focus on the development of technologies and culture to share and distribute content. This effort was incredibly successful, as a large part of our society can indeed now access freely, although often illegally, a large part of our cultural patrimony.

Although this situation creates huge problems for the traditional content industries - see e.g. the never-ending collapse of the music industry - it should not be forgotten that it is the direct consequence of an age-old dream of the western society, the dream of *accessibility for all* (a motto of the European Commission for instance, among many others institutions) that can be traced back to *les Lumières*, an era when knowledge and culture were, for the first time in history, explicitly considered as a production deemed for universality. Today’s peer-to-peer systems are but an instantiation of these century-old visions in which culture, in the noble sense of the term, should definitively be separated from mercantile considerations.

Now that everything has or will become accessible, the question of what is next to come is a legitimate one. The mass of available content is now such that a “Babel library” effect can be observed: the existence of an item of interest does not suffice to make it actually available. In many cases, recreating it from scratch is easier than finding it.

Although the current focus of information technologies lies in networking aspects of communities, I argue that the future of content lies not only in information exchange and our relations with others. So-called *personalization* technologies have addressed the issue of content recommendation and sharing, bringing the idea that content could be tailored to users, for instance through automatic recommendations systems, the most famous of which is collaborative filtering, introduced in the 90s. With these technologies, the distribution issue is basically solved, at least technically.

I claim in this chapter that the future of content lies not only in the efficiency of its distribution, but rather in the nature of content creation, an issue which has so far been left mostly untouched by technology research. More precisely, I claim that the future of content creation lies primarily in the ability of individuals to realize their creative potential themselves, rather than picking up existing content out of seemingly infinite repositories. This ability is itself closely dependent on the existence of tools that reveal us, rather than tools that connect people together in ever expanding networks of acquaintances.

Of course, not everyone is a creator: transforming ideas, possibly ill-defined, naive or uninformed, into actual objects of interest requires craft, discipline and learning, some life-long. However, I claim that content technologies can substantially boost individual realization, and help us develop inner dialogs through which personal content can emerge, that would otherwise be left unexpressed. More than sharing and communication, the major issue we have to face is that of expressive power: how to create those objects of desire when we do not know how they are constructed?

I propose to look at this question by examining some of the research projects we conducted at CSL for the past 10 years. These projects have addressed these questions with a particular focus on so-called *reflexive interactions*. These interactions are designed in such a way that users can create objects of interest (mainly musical ones in my case) without being specialists, and through mirror-like, man-machine interactions designed in particular ways.

2 Reflexive Interactions

The notion of reflexive interaction stemmed from a series of experiments in music interaction, involving an interactive music learning system (the Continuator, see Section 3.1). The idea behind reflexive interactions is to produce interactions in which an object has to be constructed (e.g. a melody, a taxonomy, etc.) not directly through a traditional construction scheme, but indirectly, as a side-effect of an interaction taking place between a user and an image of himself/herself, typically produced by a machine-learning system. Technically, this image is necessarily going to be imperfect, for many reasons, including the intrinsic limitations of machine-learning systems, but it is precisely this imperfection which is going to produce the desired side-effect.

The idea that an imperfect mirror is more interesting than a perfect one is perfectly illustrated by the famous *mirror scene* of the Marx Brother’s movies “Duck Soup” (1933), in which Harpo pretends to be the mirror image of Groucho (see Figure 1) but inserts, in an increasing manner, various “imperfections” in the replication (such as using a hat with a different color). These imperfections push Groucho to explore this increasingly unmimetic image of himself for about 10 minutes until he convinces himself that the image is not him, when a decidedly unbelievable imperfection arises, namely the appearance of a third image of himself, played by Chico.

Figure 1. Groucho (and Harpo) Marx in the mirror scene of “Duck Soup” (Leo McCarey, 1933).

2.1 The Tickling Metaphor

Aristotle, in Parts of Animals, book III (Aristotle, 350), stressed the human specificity of tickling: “For when men are tickled they are quickly set a-laughing, because the motion quickly reaches this part, and heating it though but slightly nevertheless manifestly so disturbs the mental action as to occasion movements that are independent of the will. That man alone is affected by tickling is due firstly to the delicacy of his skin, and secondly to his being the only animal that laughs”.

Figure 2. A tickling robot arm of the kind used for the tickling experiment by Sarah-Jayne Blakemore (Shadowrobot).

However, it is not clear whether Aristotle had already noticed the impossibility of self-tickling, that is of triggering laughter with self-tickling (so-called *gargalesis*, i.e. hard, laughter-inducing tickling, as opposed to *knismesis*, i.e. light, feather-like tickling). Recently, Sarah-Jayne Blakemore from the London Neuroscience institute did a breakthrough experiment in which a tickling robot arm, remotely controlled by a button, would tickle various subjects (Blakemore et al. 2000). She noticed that the self-tickling impossibility extended as far as button pushing: laughter was induced only when the button was pushed by another subject. This experiment, coupled with brain imagery would suggest that the cerebellum is able to somehow inhibit the laughter circuitry, and therefore to act as a "detector of non-self".

Furthermore, she noticed that if a delay (a fraction of a second) was introduced between the moment the button is pushed and the moment the arm is activated, then the cerebellum would not be able to perform its computation to inhibit the laughter, and self-tickling became, then, possible, somehow by fooling the cerebellum.

Of course, this experiment has a lot of consequence for neurosciences, in particular to better understand inhibition mechanisms and the role of the cerebellum in sensory-motor actions. In our context however, we interpret it differently. This experiment showed that contrarily to the intuition it is possible to self-tickle, but this requires carefully designed machinery, involving reflection, delay, and specific conditions yet to be elicited.

The question raised here draws from this experiment, and the long observed relation between laughter and creativity: if self-tickling is indeed possible through artificial machinery, can we build similar machineries for other human activities, in particular involving creativity?

2.2 Definition and Examples

Reflexive interaction is a particular class of man-machine interactions whose goal is precisely to create stimulating user experiences. Their focus is not to solve a given, well-defined problem, such as querying a database, but rather to help users express hidden, ill-formulated ideas. This expression is performed indirectly, as a side-effect of an interaction based on the systematic exploitation of powerful machine learning algorithms.

The idea that machines can act as mirrors is not new. It is the central metaphor of the vision of our computerized society developed in (Turkle, 84). However, in our context, we take it more literally, as we design systems that effectively build virtual images of users in several disciplines. These images are built with the help of real-time machine-learning components, which build models of the users that are continuously updated.

The notion of interactive reflexion is related to the notion of feedback, as it involves a potentially infinite interaction loop between a user and an image of him built artificially by a computer. Similarly, reflexive interactions exploit only information coming from the user, and do not rely on preexisting information or databases, i.e. they operate in a closed world. However, as opposed to feedback systems, reflexive interactions do not consist in feeding back the output of a system to its input (Figure 3). It consists in influencing the actions of the user by providing him with a carefully designed image of himself. Technically the main difference with a feedback system is the presence of a time-based machine-learning component between the user and this image: a reflexive interaction system performs a continuous learning of the user behavior which produces a continuously updated mirror image (Figure 3).

This definition is intentionally broad to encompass different scenarios, ranging from interactive music systems to taxonomic and search systems as well as content creation systems, as illustrated in the next sections.

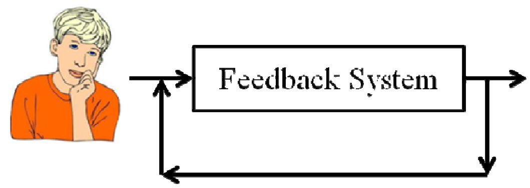

Figure 3. In a traditional feedback system (e.g. the Larsen effect, or the Karplus-Strong synthesis algorithm), the output is directly fed back to the input.

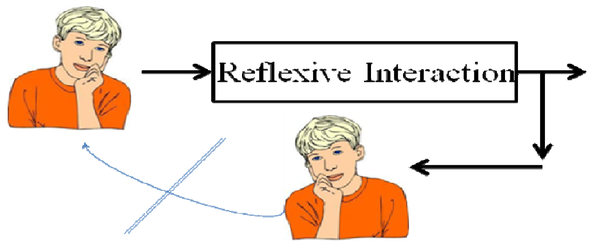

Figure 4. In a reflexive interaction, the output of the system is a continuously updated mirror image of the user.

A good example of artificial reflexive interaction is given by the “persuasive mirror” experiment realized by Accenture in collaboration with Stanford hospital. In this project, a user would see his image as shot by a camera located in front of him, on a screen, with some transformations. These transformations, performed using digital image processing, would model the natural ageing process, itself parameterized by the dietetic behavior of the user. In this context, the long-term impact of a fat diet is immediately visible, and hopefully frightening enough to push people, children in particular, to change their eating habits (Andrés del Valle and Opalach, 2005).

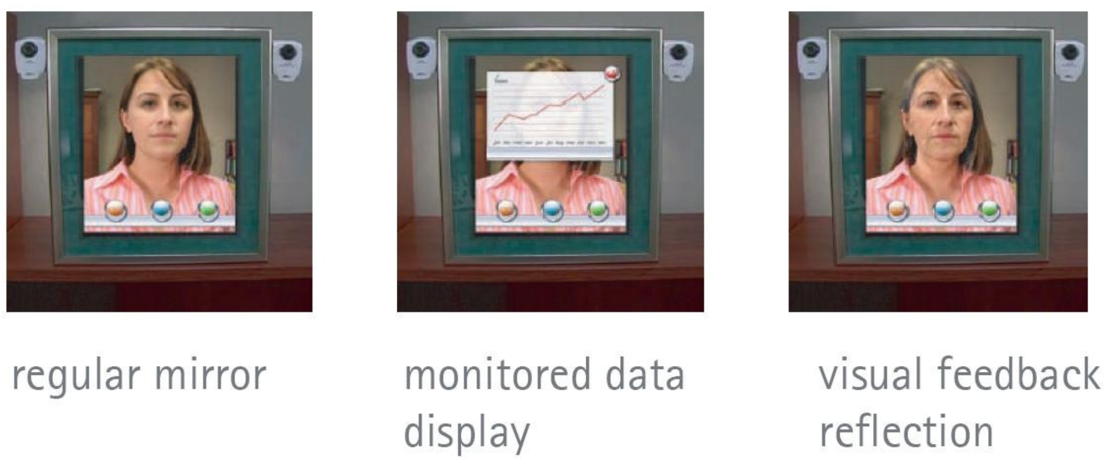

Figure 5. The Persuasive mirror (Andrés del Valle, 2005).

One interesting, and differentiating, aspect of interactive reflexion is the “always successful demo” effect, due to the manipulation of user’s intimate characteristics. With non reflexive man machine interactions, users are constantly checking the accuracy or performance of the system they interact with. A typical example of a non-reflexive interaction is an automatic audio classifier. This classifier, given an audio file provided by the user would classify this file as e.g. “Speech” or “Music”. Any misclassification of the system will typically be interpreted by the user, who knows the correct answer, as an error. Conversely, let us consider a “reflexive” equivalent of an audio classifier: a “vocal lock” system that attempts to identify users based on their voice. This system continuously updates its model from the feedback given by users (for instance “correct” or “false” identification). As it is well known in the literature, voice recognition systems are never perfect, and suffer from still poorly understood problems, notably the non uniform distribution of voice features in populations (Doddington et al., 1998). However, because the system’s action is based on the user voice, in this case, it is very likely that the output of the system is never considered faulty in case of errors. Rather, users will naturaly tend to interpret the system’s deficiencies as coming from the characteristics of their voices (so-called “wolves”, “sheep” or “lambs”). More generally we have observed that reflexive interactive systems produce demos which always “work”, because of the involvement of the user. This is a superficial characteristic of reflexive interaction, but an intriguing, and defining one.

3 Spiraling Thoughts and Experiments

To illustrate the idea that tickling robots can be designed with interactive reflexion as a key paradigm, I describe in this section three projects which can be interpreted as an attempt to build tickling robots in various domains of musical creativity. These systems are designed as reflexive interaction systems, so as to reproduce, at least metaphorically, the reflexive situations of the tickling robot experiment. The projects address the following domains: musical style exploration (the *Continuator* project), musical genre categorization (the *MusicBrowser* project), and music composition (the *DesignGame* project).

3.1 Continuator: From Frustration to Flow

The Continuator project popped out of the mind of a frustrated musician. As a jazz improviser, I have been bothered for a long time by my technical limitations. Just like many guitarists of my generation, I felt inspired by such great talents as John McLaughlin or Al Di Meola. I also wanted to be able to play these fast, harmonically consistent, seemingly infinite notes streams that blew away the listeners. And like many others, I worked hard to master the instrument, to be able to play these scales, to understand harmony, and to be able to spit out musical phrases while the harmony develops, always bolder and always faster... But my ideas would always develop faster than my hands.

The idea to use Markov processes to analyze and generate music is not new, and dates from the very beginning of computer science and information theory (Brooks et al, 1957). Many refinements to this idea were brought to these early models, culminating with the composition systems of David Cope (1996). But Continuator was the first interactive system to be able to learn and respond in real time, from arbitrary input phrases. These first Continuator-generated phrases, although linear, already produced a remarkable effect because they would capture recurring patterns which were not necessarily made explicitly by the user, in a continuous, potentially infinite stream in which the user would somehow recognize himself, sometimes enthusiastically, sometimes reluctantly.

I dreamed of a machine that would help me concretize my musical ideas, faster, better, further. In a way you could say that I was looking for an extension of my musical brain that could produce the phrases I had in mind, while letting me control them and define their very musical substance. The Continuator at its beginning (in

1999) was able to play these fast, endless musical phrases, using a novel combination of a machine-learning algorithm applied to musical streams and a real-time phrase generator. These experiments confirmed that it was indeed possible to generate endless, harmonically challenging phrases in real time. Part of the initial frustration had been overcome.

But I wanted more. The next step was to have a system that would also produce polyphonic material, with other rhythms than linear 8th notes. This required some adaptation of the algorithm and led to a new version of the system (Pachet, 2002). During this phase I worked intensely with the musician György Kurtag Jr., who continuously experimented with the system as it was being developed in 2000 and 2001. Also the various sessions with pianist Bernard Lubat during this period helped me to build a robust system, suited for intensive concert sessions (e.g. at Ircam in 2002, see Figure 6) as described in (Pachet, 2002b).

Figure 6. Bernard Lubat playing with the Continuator during a concert at Ircam, October 2002.

Many other threads developed, in particular the combination of several Continuator systems together where different inputs could be mingled into one system. One of the most striking results is shown in “Double Messieurs”, a movie by Olivier Desagnat involving Kurtag Sr. and his son (see Figure 7).

Figure 7. G. Kurtag, father and son, in the movie « Double Messieurs » By O. Desagnat, 2002.

However, the musically most interesting sessions were probably the ones with pianist Albert Van Veenendaal, who very quickly learned how to play with the system to extract the most significant responses from it. The encounter with Albert is yet another example of a winding road. We met initially to conduct a “musical Turing test” for a radio broadcast on VPRO, a Dutch public network.

The principle consisted in having two jazz critics listen to Albert playing on a Midi-controlled grand piano (a Disklavier) linked to the Continuator, and try to guess whether it was him or the system playing, at any moment in time. The comments of these critics were recorded and broadcasted on the Dutch radio VPRO in June 2004. With the Disklavier producing the same sound whether played manually or controlled by the computer, the critics could only base their judgment on their analysis of the music. The test showed that the difference was not detectable (the critics would be correct about 50% of the time), so that Continuator would pass the test (Veenendaal, 2004). Of course, the playing style was free improvisation, and not structured composition as in the experiments by Cope (1996). But this was precisely the aim of the test, which was probably the first Turing test of free improvisation. This positive result shows that the system can somehow fool listeners (as well as the user himself…), at least for a short duration.

Figure 8. A Jazz Turing test for the Continuator organized by the Dutch Vpro radio station.

Even though two jury members deemed the test itself successful (see Figure 8), we noticed that the system was more convincing when the pianist would be frantic, playing quick material rather than slow phrases. We agreed to meet again later for another session where the focus would not be Turing any longer, but rather “How to play slow music with the Continuator”. The results of these sessions are probably the most interesting pieces of music played with Continuator so far (see Figure 9).

Figure 9. Albert Van Veenendaal improvising with the Continuator.

It turned out that the system was more intriguing than expected as I realized something very interesting: apart from the technical aspects of the learning algorithm and the real time generator, it became clear that the *subjective aspects* of this new kind of musical interaction were extremely unusual. Quite often, people experienced “Aha”-phenomena while interacting with the Continuator and some even seemed to get addicted. So the next question was obvious: How can we understand why and how the system created such reactions?

The question was not so much a technical than a psychological one. It turned out that very few psychologists studied excitement or even “fun”. The closest studies I could find were writings on Flow theory by Csikszentmihalyi (Csikszentmihalyi, 1990). In his view, two mental states are primordial: boredom and anxiety. Between them lies a region of Flow, where challenges match skills, and in which people experience “optimal” states, are able to concentrate, to forget time, and create new goals in a totally autonomous way, the so-called autotelic state (Steels, 2004). The next questions were therefore: “Is the Continuator a Flow machine?” and “How can we look at Continuator from this perspective?” Basically, the answer was to look at children.

Some preliminary experiments were conducted in Paris with 3-year old children (maternelle Bossuet Notre Dame, Paris 10[e], see Figure 10). The children’s reactions were enthusiastic: they became suddenly interested in the keyboard, had fun with the answers produced by the system and most of all were able to focus their attention for extended periods of time, sometimes up to 40 minutes.

Figure 10. A 3-year old child playing with Continuator.

These experiments were, however, not systematic, and still looked more like scientific “hobbyism” than anything else: sessions were not always properly recorded and the protocol was not rigorous. In short, I was not sure what I was looking for, but there were sufficient intuitions that this was a very interesting area to pursue.

A decisive breakthrough occurred during my encounter with Anna-Rita Addessi from the University of Bologna, which led to psychological experiments with a well-defined protocol and a systematic study of the impact of the Continuator on early childhood musical development. We quickly set up a one-week session in a secondary school in Bologna (La Mela), where protocols were established, sessions were organized and videos were shot involving the Continuator. The next years were devoted to the analysis of these videos (Pachet & Addessi, 2004).

We needed a guideline to assess the impact of Continuator (with versus without). Flow theory turned out to be particularly helpful to this aim, as it is the only psychological attempt to describe these particular mental states where people engage themselves entirely in their activity, regardless of anything around them (Figure 11). More specifically, Flow gave us a list of precise criteria we could measure, such as increased attention spans, development of novel musical behaviors, autonomous discovery of turn-taking protocols, and many other fascinating phenomena. These measures could be compared in various situations, such as with or without the Continuator, but also in single child machine-interaction or with two children.

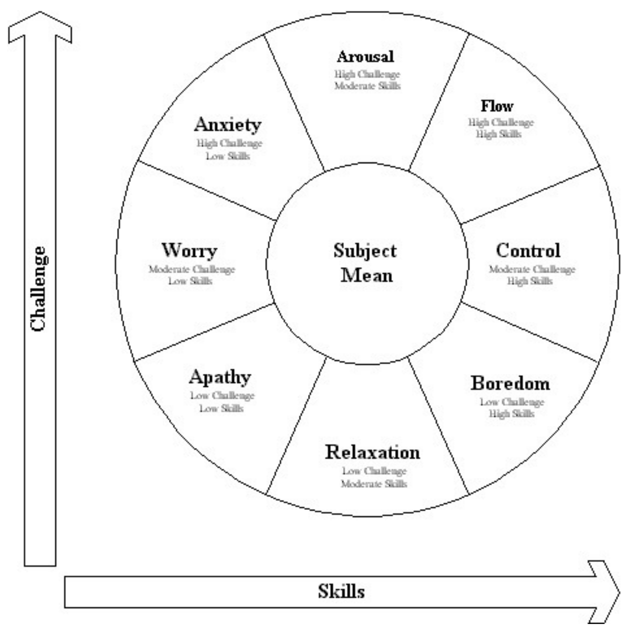

Figure 11. The state diagram of Csikszentmihalyi's describes several emotional states such as boredom and anxiety, in terms of the relationships between challenge and skills.

During these analysis, we noticed many interesting behaviors, occurring after the initial phases of surprise and excitement. Notably, several children started to invent new playing modes, sometimes really innovative (with the sleeves, the mouth, the elbows, etc.). A particularly interesting moment was when we could literally see a child discovering and understanding the notion of "musical phrase". This was clearly indicated by a typical launching gesture, ending the phrases of the child, and very similar in shape to the spontaneous gesture performed by professional musicians (see and compare Figure 6 and Figure 12).

Figure 12. Children launching the Continuator after having finished their musical « phrases ».

We had collected enough information for the next five years to come (Addessi & Pachet, 2005). Later, we managed to get other researchers in the area of music psychology and education to become interested in our approach: *interactive reflexive music systems*. Work still continues in this direction with many other subjects of study being identified and investigated, notably the long-term impact of these systems on musical development, the sensitivity of children to musical "personality", the relation between musical behavior and physical movements, and the ability to invent new interaction modes (e.g. Ferrari et al. 2004).

The various experiments produced many important improvements to the initial system, but also stressed its limitations. In particular, the automatic generation of interesting rhythmic information remains open: the Continuator, in its standard mode, does not have a precise idea of tempo for instance. Harmony is also problematic, as the system is harmonically deaf. Some extensions of the Continuator were investigated to address these issues. I devised in particular an interactive mode in which the output of the system is "corrected" using external harmonic information, thereby introducing the notion of reflexive harmonization (Pachet, 2006).

An even harder problem is structure: musical phrases generated by the Continuator do not have a clear beginning or ending. This type of information, like rhythm or harmony, is indeed extrinsic to the notes and therefore does not fit well with the Markov view of time sequences. Finally, the basic interaction mode of the Continuator was turn-taking or question-answer. We had to invent and program many other interaction modes, corresponding to various musical situations (in particular *reflexive harmonization* and so-called *harmonic attraction*, see Pachet, 2006). But these designs were made manually, one-by-one, which raised what turned out to be a particularly fundamental problem: the need for dynamically creating interaction modes, in reaction to unpredictable situations. Indeed, a truly flexible interactive system should be able to adapt, or even possibly create on-the-fly interaction protocols, as humans often do in real situations. A good music teacher, for instance, can switch freely from an explicative to a listening or accompanying mode with his student. In this line, a new collaboration was initiated with Sergio Krakowski, a Brazilian tambourine player. Sergio builds interactive music systems which are controlled by a Pandeiro, a simple percussive instrument with about 6 different sound classes (Roy et al., 2007). His goal is to bootstrap basic interaction protocols that can be used in real-time to invent interaction modes, without having to pre-program them. These modes targeted include commands such as "play a short-duration chord on this type of sounds" to more complex modes such as "repeat the last pattern I played until I decide you should stop".

Figure 13. Sergio Krakowski experimenting with a reflexive interaction system that creates modes on-the-fly.

But what about this initial desire to overcome the frustration I experienced as a guitar-player? It is obvious too that this chain of experiments, scientific investigations and encounters of all kinds are all ingredients of this quest for understanding the creative act of musical production. At each step of the process, fascinating results were obtained, either scientific (e.g. concerning the study of child development), technological (concerning for instance machine learning algorithms for musical style), or musical (concerts with the Continuator are a particular, and hitherto unheard of form of musical expression). However, the very act of improvisation, even through reflexive interaction, produces a specific type of content. Improvization is a practice designed for real-time performance, concerts, as opposed to music composition, which aims at producing universal objects, to be contemplated, listened to again and again. Of course, the limit between imrpovization and composition is fuzzy: some Jazz improvisation do achieve the status of intemporal objects (the famous chorus of Charlie Parker are literally written, more than improvized), and some composers (e.g. Mozart) are able to compose music on-the-fly, without a need to backtrack, as if they were improvising. But as stimulating as they can be, sessions with the Continuator do not produce reusable content, i.e. fully-fledged pieces of music. One strong limitation was the lack of a linguistic component that would allow the user to structure the music stream produced by the system.

So the next question in this spiral of thoughts was naturally to shift the attention to linguistic features, in particular taxonomic thinking.

3.2 Reflexonomies: Mirrors Plus Taxonomies

The Continuator is clearly an instance of a reflexive interaction system, but it is not the only one. The Music Browser project started in 2002 to investigate how to design music categorization systems tailored to the tastes of users. Indeed, the explosion of available music titles in digital form created a pressure for automatic categorization tools. Several approaches were developed. On the one hand, purely manual approaches consist in letting experts categorize music and making these categorizations available on-line. The *All Music Guide* effort targets the systematic description of all music in the world to this aim. On the other hand, automatic approaches try to extract this information from the signal itself. These approaches are interesting, but robustness and precision appear to be intrinsically limited and therefore forbid their use in commercial contexts (Aucouturier et Pachet, 2003). Another approach is the exploitation of social information, such as collaborative filtering, or, more recently, *social tagging*. In these approaches knowledge comes from the community. Each user tags or annotates a piece of music (or any other type of cultural information). This information is then aggregated and made explicit for a given community. The most popular tags can be detected automatically, resulting in the emergence of a robust description language, or lexicon, also called a *folksonomy*.

However, in all cases a fundamental problem remains: tagging automatically is error-prone, and tagging manually is tedious.

The idea to apply reflexive interaction in this context is therefore natural: tagging is a way to define a personal language, and it turned out that reflexive interaction is particularly well suited to this task.

More precisely, we introduced a new notion, intermediary between taxonomies (created by experts) and folksonomies (emerging from the behavior of a community): *reflexonomies*. A reflexonomy is basically a taxonomy created by a single user through a reflexive interaction. The general schema introduced in Figure 4 is instantiated as follows: the inputs of the system are classification actions, such as creating a new tag or associating a title to a tag. The learning mechanism consists in building a classification model from the tagged examples, and updating it continuously after each interaction.

Technically, this model is based on a timbral analysis of music titles. The acoustic *features* used to analyze the music files are basically MFCC coefficients, computed on successive frames, and aggregated using *Gaussian Mixture Models.* Like Markov models, GMMs capture essential characteristics of data distributions, although in the continuous domain. We have shown elsewhere that this particular approach was well suited, and in some sense, optimal, to model polyphonic music (Aucouturier et al, 2005). Figure 14 shows how this approach works when applied to the Beatles catalogue, after having been trained on a variety of titles from other artists: most of the Beatles are unsurprisingly classified as *Pop/Brit*. Titles classified in less predictable classes can in fact be explained using musical arguments (e.g. the soundtrack of *Yellow Submarine* is classified as *Classical*, and indeed is a mostly orchestral tune).

The problem with this automatic classifier is that, regardless of the performance of the classifier and feature extraction scheme used, the classes derived from this automatic analysis have to be "understood" by the user. The class system (ontology) we have used was designed by Sony Music experts, and there is no reason arbitrary users can indeed understand, e.g. the subtle difference between *Folk / Pop* and, say, *Pop / Folk* (Figure 14). The only way for a user to understand these classes is by browsing, listening, and … spending time to learn this particular lexicon.

Applying interactive reflexion here is precisely a means to invert this master/slave relationship. In the MusicBrowser, a classification panel is presented to the user, and he can freely drag and drop titles to boxes representing classes (or tags). He can also introduce new tags by typing its name, for a given title. After each classification action the system analyses all the titles, updates a model (GMM) of the corresponding classes, and then uses this model to classify automatically the other titles of the collection. The result is directly presented to the user, who can then, in a manner similar to the Continuator dialogues, decide or not to accept these changes by resetting new tags to some titles, and iterate until the resulting classification looks satisfying.

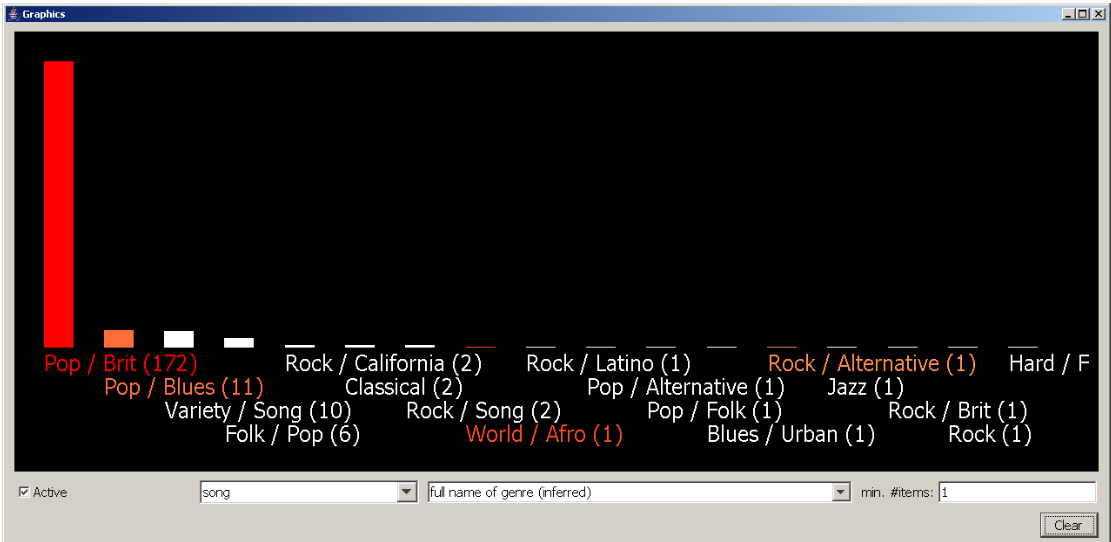

Figure 14. The automatic classification of all Beatles songs using a GMM / MFCC approach. Most of them are classified as Pop/Brit. Some of them are classified in more exotic classes, such as Folk/Pop or Classical.

The resulting reflexonomies can be seen as grounded ontologies. They can then be reused, for instance to classify other, possibly larger, music collections. The difference with automatic systems such as the one described in Figure 14 is that the user will fully understand, and trust, the result of the classification since he participated, with the machine, in the elaboration of the ontology.

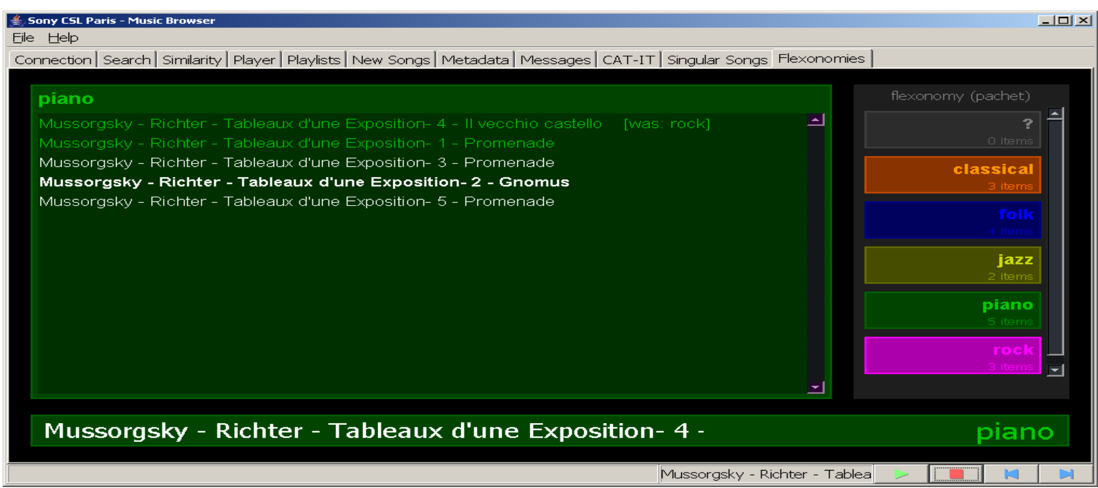

Figure 15. A reflexonomy created through a reflexive interaction. Here, the user can see the impact of four classification actions (introduction of 4 tags: *classical*, *folk*, *jazz* and *piano*, and one example per tag) on a collection of 15 titles.

The construction of reflexonomies produces interactions of a similar nature as the Continuator ones. Some experiments were conducted with students at the University of Bologna (Figure 16) to evaluate this aspect systematically. An interesting aspect of these interactions is the shift they operate in the goals of the user: initially the classification is somewhat artificial (why, after all, should one classify titles?). After a while, thanks to the nature of the interaction, the activity becomes autotelic, and users classify not so much to categorize music, but to better understand their own way of classifying. Although systematic Flow studies were not conducted as with the Continuator, we consider the MusicBrowser as yet another instance of a Flow machine, in which the difficulty of a self-imposed task increases as more information is given to the system.

Figure 16. A session with MusicBrowser reflexonomies in Bologna.

3.3 Overcoming the Frustration of Technical Languages: Interactive Combintarial Design

As we saw with the Continuator and the MusicBrowser, a part of the frustration in expressing creativity lies in the difficulty in understanding the technical languages that govern the structure of the objects of study. Being able to improvise is difficult, not only technically, but also because it requires knowledge about harmony, melody, rhythm, that can only be acquired through long hours of practice. Similarly, classifying music requires the understanding of genres and various musical categories that typically require a long apprenticeship. Although these two activities are of a very different nature, they lend themselves naturally to reflexive interaction games. Taxonomic thinking does introduce a linguistic component in the loop, but creating taxonomies is not creating content: a last ingredient is still needed to produce actual content objects. This is where *combinatorial design* was introduced, as a way to bridge the gap between taxonomic thinking and object design.

3.3.1 Combinatorial Design

The last project I describe here concerns the particular problem of designing digital objects (i.e. that can be manipulated by machines) using *ad hoc* languages. In this activity, called *combinatorial design*, the aim is to provide users with a reflexive interface of a novel type for creating various kinds of objects such as colors, melodies, sounds, logos or simple texts. The creation of these objects typically requires the knowledge of corresponding technical languages: the language of melodies (involving features such as repetitions, patterns, arpeggios, scales, etc.), the language of colors (involving technical representation spaces such as RGB, HSV and others), the language of sounds (requiring the deep knowledge of particular sound synthesis algorithms), etc. The main idea here is to alleviate the burden of learning these technical languages by putting together two components: 1) a tagging system in which users can describe existing objects with their own set of tags. Similarly to the MusicBrowser, this tagging system is associated with a machine-learning system that continuously learns mappings between user tags and technical features extracted automatically from the objects, and 2) a combinatorial object generator, that allows people to explore variations of these objects using linguistic modifiers, based on their tags. These modifiers consist in selecting an existing object and then applying transformations expressed as "more X" or "less X", where X is a tag introduced by the user himself. The combinatorial object generator is then able to construct a new object which is both "similar" to the initial object, while being slightly more (resp. less) of category X. This more/less property is determined by the probability of the variation to be categorized as X, by the system (trained with examples previously tagged by the users).

A simple and typical example of this interactive system is a tool for creating colors. In this system, colors are initially randomly created by the system. The user can tag colors using his own, possibly subjective words. Examples of tags can be "happy", "bright", "sad", "red" or "blue". As soon as a tag is associated to a color, the system automatically retrains a classifier for this tag, using a set of predefined technical features. In the color example, the features are the R, G and B components of the RGB representation, as well as the H, S, V components of the HSV representation (features here are not necessarily mutually exclusive). Once the system has finished the training phase (which takes only a fraction of a second in this case), the user can select an arbitrary color in the panel (such as a yellowish one), an arbitrary tag in his lexicon (such as "blue"), and ask for a color which is "close" to the yellowish one, but "more blue". The system then creates a new color which is both close to the initial one (the yellowish one), but that it considers as more "blue". By repeating this operation (more blue) the color, in this case is progressively transformed and becomes greener and greener (yellow plus blue yields green). At any point the user can select a different tag, e.g. the tag "bright", to create eventually a brighter, greener color.

At any point during the session new tags can be introduced, and existing tags can be updated (removed, added, etc.), and the classifiers are automatically retrained.

In other words, this system allows users to create grounded lexicons to describe objects (a reflexonomy, as in the MusicBrowser), but more importantly allows them to use this lexicon as a tool to create new objects. As a consequence, the system turns descriptive languages into actuators, and substitutes the task of having to learn technical languages, by the task of creating grounded lexicons.

The same idea is being applied to more complex objects, such as sounds, chords and melodies. A current experiment consists in using a complex sound synthesis engine (FM synthesis), in which a few parameters allow to generate a wide variety of sounds. Here again, the technical understanding of FM synthesis is a difficult task, mastered only by a few individuals (Chowning and Bristow, 1986). Tags such as "aggressive", "smooth", "brassy", "slappy" can be introduced by a user, and then used to modify existing sounds, to eventually converge to a desired, implicitly defined sound the user "has in his head".

This last project, only briefly setched here, fills the gap between linguistic activities which are often strictly descriptive, and production activities, which are notoriously difficult to reify, and resist machine representation. With combinatorial design, we hope that users will be able to create complex and interesting objects, without necessarily understanding the details of their intimate technical languages: these objects are the future of content.

4 Conclusion

We sustain here the idea that the future of content, a highly societal problem in principle, lies paradoxically in the capacity of individuals to realize their creative potential, more than in connecting them frantically together. We propose interactive reflexion as a paradigm to build tools in which the expression of this potential is achieved through the manipulation by users of their own image. This image can be a stylistic model (the *Continuator*), an ontology (the *MusicBrowser*), or a grounded, active lexicon (*combinatorial design games*). In all cases, objects are created as a side-effect of this interaction, and result most of the time in Flow generating experiences.

Recently, experiments by Steels and Spranger (Steels & Spranger, 2008) have also exploited reflexive interactions as a way for robots to self-teach how to recognize and interpret gestures performed by other robots (see Figure 17). This last scene echoes the mirror scene of the Marx Brothers, in a context where humans have disappeared, but not reflexion: the key ingredient, so I claim, of the future of Content.

Figure 17. A robot engaged in a reflexive interaction, to self-teach gesture interpretation.

5 Acknowledgements

I wish to thank Mario Tokoro and Luc Steels for their continous support and trust. I also thank Jean-François Perrot for insightful discussions and comments on an earlier version of this paper. I also thank the students and collegues who participated in the development of the experiments described here, notably Jean-Julien Aucouturier, Anthony Beurivé and Pierre Roy as well as Anna-Rita Addessi from the University of Bologna.

6 References

Addessi, A.-R. and Pachet, F. (2005) Experiments with a Musical Machine: Musical Style Replication in 3/5 year old Children. British Journal of Music Education, 22(1), March.

Aristotle, Parts of Animals (350 BCE) Book III, translated by William Ogle, eBooks@Adelaide.

Andrés del Valle, A. and Opalach, A. (2005) The Persuasive Mirror: Computerized Persuasion for Healthy Living, Human Computer Interaction International, HCI International, Las Vegas.

Aucouturier, J.J and Pachet, F. (2003) Representing Musical Genre: A State of the Art. Journal of New Music Research, 32(1).

Aucouturier; J.-J., Pachet, F. and Sandler, M. (2005) The Way It Sounds: Timbre Models For Analysis and Retrieval of Polyphonic Music Signals. IEEE Transactions of Multimedia, 7(6):1028-1035 December.

Blakemore, S-J, Wolpert, DM & Frith, CD (2000) Why can’t you tickle yourself? *NeuroReport* 11, 11-16;

Brooks, Hopkins, Neumann & Wright. (1957) "An experiment in musical composition." *IRE Transactions on Electronic Computers,* 6(1).

Chowning, J. Bristow, D. (1986) FM Theory & Applications - By Musicians For Musicians, Yamaha Music Foundation, Tokyo, Japan.

Cope, David. (1996). Experiments in Musical Intelligence. Madison, WI: A-R Editions.

Csikszentmihalyi, Mihaly. (1990) Flow. The Psychology of Optimal Experience. Harper and Row, New York.

Doddington, G. R. Liggett, W. Martin, A. F. Przybocki, M. and D. A. Reynolds, (1998) “Sheep, Goats, Lambs and Wolves: A Statistical Analysis of Speaker Performance in the NIST 1998 Speaker Recognition Evaluation,” in International Conference on Spoken Language Processing, Sydney, Australia.

Ferrari, L. Carlotti, S. Addessi, A.-R. and Pachet, F. (2004) Suonare con il Continuator è unésperienza ottimale ?. Proceedings of International Symposium on Psychology and Music Education.

Pachet, Francois. (2002) The Continuator: Musical Interaction with Style. In ICMA, editor, Proceedings of ICMC, pages 211-218, September 2002. ICMA. best paper award. Extended version in Journal of New Music Research, 31(4).

Pachet, F. (2002b) Playing with Virtual Musicians: the Continuator in practice. IEEE Multimedia, 9(3):77-82.

Pachet, F. (2006) Enhancing Individual Creativity with Interactive Musical Reflective Systems. Psychology Press.

Pachet, F., Addessi, Anna-Rita (2004) When Children Reflect on Their Playing Style: The Continuator. ACM Computers in Entertainment, 1(2).

Roy, P., Pachet, F. and Krakowski, S. (2007) Analytical Features for the Classification of Percussive Sounds: The Case of the Pandeiro. Proceedings of DaFx 07.

Steels, L. (2004) The Autotelic Principle. In Fumiya, I. and Pfeifer, R. and Steels,L. and Kunyoshi, K., editor, Embodied Artificial Intelligence, Lecture Notes in AI (vol. 3139), pages 231-242, Springer Verlag. Berlin.

Steels, L and Spranger, M. (2008) The Robot in the Mirror, Connection Science, 20(3).

Turkle, Sherry (1984) The Second Self, New York: Simon and Schuster.

Veenendaal, A. (2004) Continuator plays the Improvisation Turing Test: http://www.csl.sony.fr/~pachet/Continuator/VPRO/VPRO.htm.