

Computer Based Statistical Composing

Stefan Schuster

Diplomarbeit Jahrgang 2011/2012

Masterstudiengang Digitale Medientechnologien, FH @U%JA2\$""%

dms.schuster@gmail.com

Introduction

In the past twelve months various studies were conducted to develop a universal terminology to describe and analyze film music and its effect on the reception of a film scene. Furthermore an attempt was made to interpret this kind of music statistically and to find recurring patterns related to certain film genres. Finally the gathered data was interpreted and processed for the use in an algorithm to compose film music automatically.

Listening Test

At the beginning of the process to develop an algorithm that is capable to compose film music adequately and automatically a database of 250 film clips of various genres was conducted by students of the Fachhochschule St. Pölten. This database was used as the ground truth for subsequent studies and was analyzed in various ways. A terminology to categorize emotions and film scenes at the same time, based on psychological, sociological and psychoacoustic parameters was generated to be able to capture to whole effect of film music on the human mind. Based on Darwin's book *The Expression of the Emotions in Man and Animals* in which he defined six universal face expressions: anger, grief, fear, joy, surprise and disgust, progressive symbols were constructed to be verbally closer to the subject film and its music. Therefore the used symbols are: fear, action, freedom, joy, war, desolation, tragic and romance. These categorizations were used to tag every clip in the database and consecutively to group them in smaller clusters. The next step was to develop a listening test that is capable to find recurring patterns within the symbols and deliver statistically interpretable data of the use of instruments, triggered emotions, importance of melody or rhythm and unique features of a single symbol.

Prior to the actual listening test, the audio track of each clip was extracted and clips that didn't comply the following rules were excluded:

- Only orchestral instruments should be used in the clip
- No dialogs
- A complete motive should be hearable within 10 seconds

A group of experts analyzed afterwards the music to find for each clip an instrument section e.g. brass or strings that is sonically dominant and if possible an instrument e.g. violins that is within its group dominant.

After these first classifications the database shrunk down to 120 clips and the first leads to describe the symbols statistically were found so the next step was to conduct the listening test. The test itself consisted of 6 clips and each clip was analyzed in three steps:

- A circumplex model to gain information about the psychological effect of the music. The axes of the model were: active/passive and positive/negative. The results of the distances and the angles to the origin were rounded.

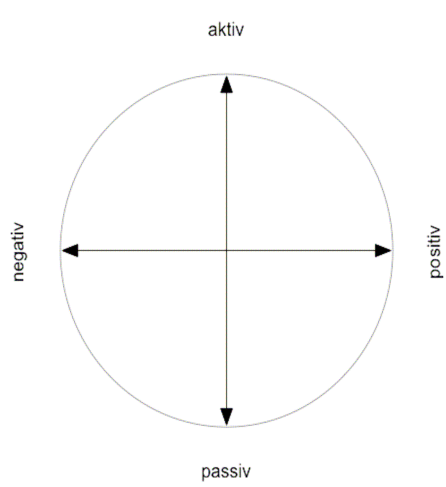


Abb. 1 Original circumplex model

- The next step was the question which symbol described the clip the best. Only one answer was allowed so that a clear definition was possible:
 - romance
 - fear
 - war
 - desolation
 - action
 - tragic
 - freedom
 - joy

- Finally a question why the symbol was chosen was asked. Multiple answers were possible:
 - rhythm
 - melody
 - instrumentation
 - other

Based on the circumstances the 120 clips were split up into 20 individual tests whereat each test consisted of different symbols. Eight rounds of testing were conducted so that every clip had at least 4 evaluations and only 1 minute was given to rate each clip, in order to get spontaneous answers. Four answers may seem not enough to find significant results but the priority is not the clip itself, it's the belonging to a certain symbol and the analysis of the symbol. The results from the tests viewed together with the analysis of the music instrument wise which happened before showed that there are recurring patterns within certain symbols and that some symbols can be identified with a high percentage rate even without seeing the images:

- *Action* showed a high percentage of recognition. The test persons felt that the music is activating but showed no tendencies towards negative, or positive, so the assumption is that the images are important for the interpretation how the viewer absorbs music in action scenes. The music has a harmonic structure with brass instruments, mostly trumpets in the foreground. Instrumentation was mostly chosen as reason what defines the symbol.
- *Fear* was detected similar to action with a high percentage of 88%. These scenes have an atonal structure with strings in the foreground and the lead melody is mostly played by violins or celli. These scenes are perceived as negative and activating.
- *War* scenes are dominated by their melodies played by violins. Key features in the scenes are percussion instruments and snare patterns such as marching snares.
- *Freedom* shows a tonal character with *melody* as the dominant reason in combination with the instrumentation. A primary instrument or group couldn't be identified in the analysis of the clips. The music is perceived as positive but shows no tendency towards active or passive. The clips showed a match of 84 percent.
- *Joy* showed the highest matching rate with 92 percent but no clear tendencies towards a reason what defines the symbol. The music is dominated by violins and woodwinds.
- *Romance* was practically not identifiable because of the close relation to Joy and Freedom. It showed no clear reason why the symbol was chosen. The harmonic structure is mostly tonal and played by violins. The music is perceived as positive and passive therefore these scenes have a relaxing effect.

- *Romance* was practically not identifiable because of the close relation to Joy and Freedom. It showed no clear reason why the symbol was chosen. The harmonic structure is mostly tonal and played by violins. The music is perceived as positive and passive therefore these scenes have a relaxing effect.
- *Desolation* was identified correctly in 56 percent of the cases and was often confused with tragic scenes. The listeners felt that the *instrumentation* is what matters most and that the music is perceived as negative. The lead instruments are violins or celli.
- *Tragic* showed no clear tendencies in the circumplex analysis and was identified with only 10 percent. There can be no clear structure in harmonics or instrumentation found. This symbol posed the biggest problem in the search for recurring patterns and had to be reanalyzed after the listening test with the conclusion that in most of the cases a piano was used as the lead instrument and therefore didn't match the previous stated rule of classical orchestral instrument usage.

Clustering

Next step in the process was a more detailed analysis of the sonic presence and usage of instruments and a machine based grouping for each symbol. A detailed explanation of how this process was executed can be found on <http://gemma.fhstp.ac.at/publikationen/>.

For each symbol five individual clusters were generated with the most common constellations of instruments. Furthermore an analysis of average sonic presence for each symbol was conducted so that it was possible to place every instrument volume wise from the foreground to the background.

Statistically based compositions

The previous studies conclude in the next step: The composition of film music based on the data collected before. With this procedure the attempt was made to evaluate the data, definitions and observations. Based on the parameters it was tried to eliminate most of the creativity of composing music so that the music is based only on statistical data. Certain parameters like the choice of playing style of the instruments, if not defined by the analysis were left to the composer. For a better understanding of the procedure of conducting a statistical composition an example for the symbol action is shown below:

- Example cluster with sonic presence:

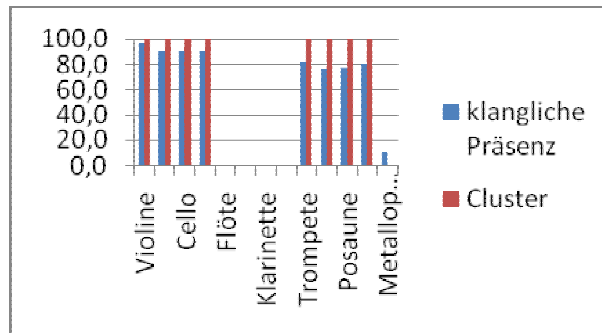


Abb. 2 Action

- Creative parameters:
 - tonal melodies
 - primary group: brass
 - primary instrument: trumpets
 - secondary group: strings
 - often percussion instruments
 - melody/instrumentation is most important

Eight compositions, one for each symbol were conducted and evaluated by another listening test equal to the first one. The results of this tests showed that the chosen terminology, parameters and instruments matched in most of the cases the listening expectations of the audiences and resulted in most of the cases in matching rates above 50 percent. The number may not seem high but it has to be considered that the music was conducted without images. Regarding the reasons what defines the music and the symbol matching rates above 70 percent were achieved, therefore it can be said that a correct interpretation of the main characteristics of the music was made.

The last step to machine based compositions are considerations regarding usability questions and possible implementations of features to improve the output of the algorithm. The main thought is to make the program usable for everybody and that no musical knowledge is needed to use or understand the program. Therefore one should try to avoid any terms that aren't used in everyday language and can lead to confusion in terms of interpretation.

The basic composition happens because of the choice of symbol and then further interfaces were developed to improve the output:

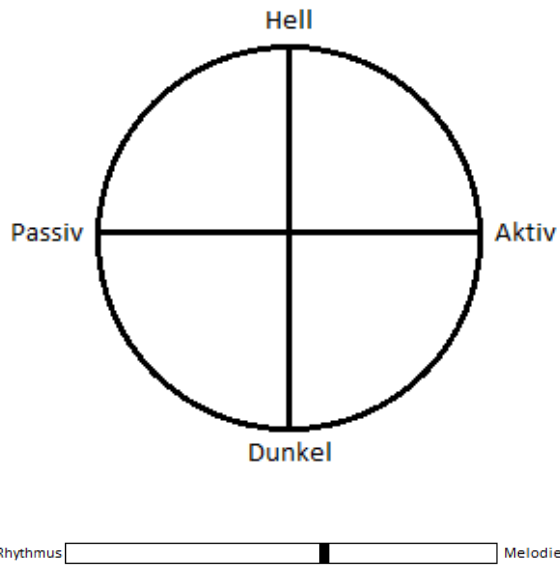


Abb. 3 Modified circumplex model; preference slider

With these two very simple interfaces it would be possible to adjust the instrumentation, playing style, rhythm and tone of the composition. The circumplex model has the axes bright/dark (for the tone) and active/passive (rhythmic accents and/or usage of percussion). The slider shown above affects the instrumentation and playing style and can shift the balance between rhythm and melody.

The future will show if a program can be used to compose film music for temp tracks and home videos or even cinema projects. The results presented in this paper lead to the conclusion that, if there are ways to aid the algorithm in the process, an output can be achieved that should match the expectations of a broad audience.