Groove Machine

Authors: Kasper Marklund, Anders Friberg, Sofia Dahl, KTH, Carlo Drioli, GEM, Erik Lindström, UUP
Last update: November 28, 2002

1. General information

Site: Kulturhuset-The Cultural Centre of Stockholm, September 27 2002.
Context: A part of a public event called Virtual Scratches/Moving Grooves
Involved partners: KTH, GEM, UUP, DIST
Involved artists: Dancer Ambra Succi, Bounce dance company, DJ 1210 Jazz, virtual scratching, normal scratching and groove composer.

2. Aim

- To investigate how expressive cues in dance can be used to control the music in a real time live artistic setting.
- To test simple yet powerful means of collecting motion data and interact with others in real-time, live performance.
- To investigate if three intended emotions (happy, sad , angry) performed by the dancer could be predicted by the video analysis.
- To explore how a dancer could "play" flexible voice samples in which the tempo and the pitch of the voice could be independently controlled in real time.
- To assess audience feedback (see separate report)

3. Concept

The general idea was to make an artistic performance in a popular style, namely hip-hop, combining both dancing and scratching – two important parts of the style. The main difference from more traditional performances was that the dancer was able to control the music acting both as a DJ and musician, instead of just following the music passively.

In Part A, the dancer controlled the mood of the music by changing dancing style. A patch was developed in EyesWeb that from the video input (of the dancer) extracted three different cues: General quantity of motion (QM), maximum velocity of gestures in the horizontal plane (VEL), and, the time between gestures in the horizontal plane (IOI – interonset interval). The cues were inspired from previous experiments in prediction of the emotional coloring of music performance, thus QM and VEL corresponded to sound level and IOI to the inverse of tempo. Another part of the patch predicted the intended emotion by making a qualitative judgement from these three cues. The prediction output was controlling a software sound mixer that blended together three different grooves composed as to express the three emotions. These grooves were rather long loops of audio with drums, bass and some additional accompanying instruments playing typical hip-hop patterns. They were all in the same tempo and key so that they could be mixed together. The dancer could then gradually change the continuously playing music by changing dancing style.

Part B featured the DJ that played the virtual scratcher and normal turntables – a collaboration with the EU-project SoB.

In part C, the DJ played drums manually on the turntable using scratching techniques while the dancer played on a virtual voice sampler. The patches for sinusoidal analysis/synthesis developed by GEM were used to trigger two voice samples, one on of each side of the camera view. The hands were tracked using the man-parameters extraction in EyesWeb. The voice sample was triggered when the hand reached out to either the right or the left. The vertical position of the hand determined the pitch of the voice and the width of the vertical bounding rectangle of the whole body controlled the playing speed.

Experiences from other productions indicate that it is important to include some pedagogical hints so that the audience conceptualises that the dancer is really controlling the music. For this reason, the dancer's silhouette and associated bounding rectangle were projected for the audience on a small screen to the right on the stage, see Figure 1.

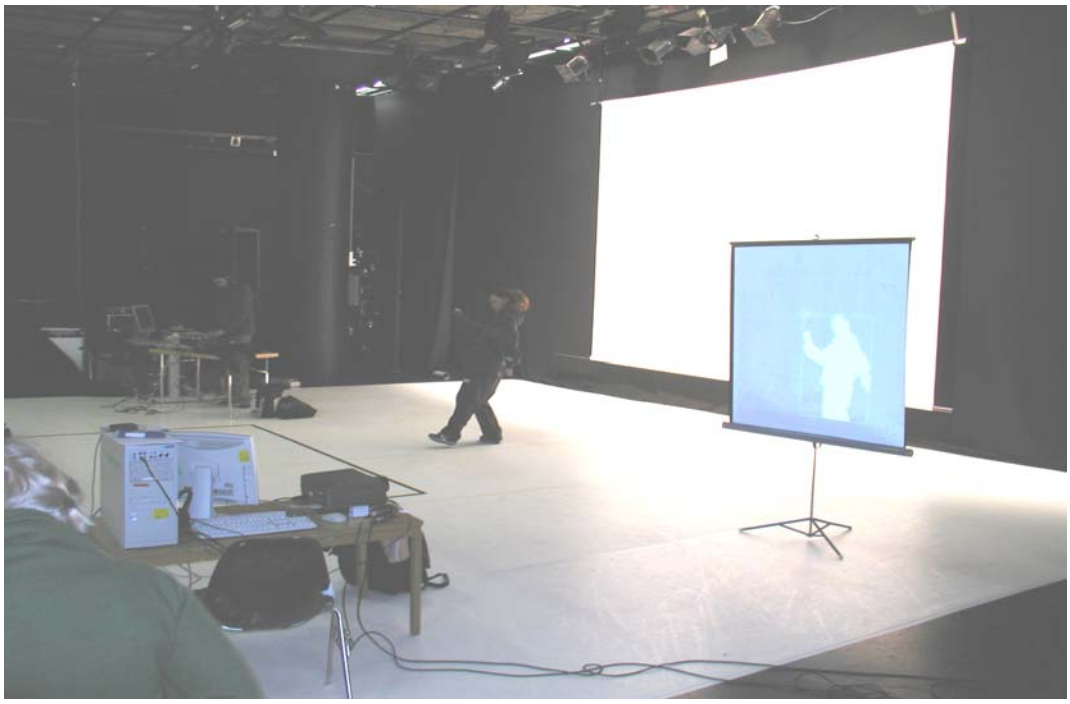


Figure 1. A picture of the stage. The dancer is in the middle, the camera in the front of the dancer, and the silhouette and the bounding rectangle is projected on the small screen to the right. The DJ is standing in the rear of the picture to the left of the dancer.

4. Relation with MEGA

Most of the features of the performance were exploiting results derived directly from the MEGA work packages. The analysis of the dancer used two parts resulting from work within WP3: a motion cue extraction and an expressive mapper. The music output featured mainly the voice processing algorithms made in WP6.

The material from the event can also be used in the subsequent work in the WPs. The video input of the camera was recorded as well as the output of the cue extraction during the performance.

5. Technical description

5.1. Hardware and software set-up

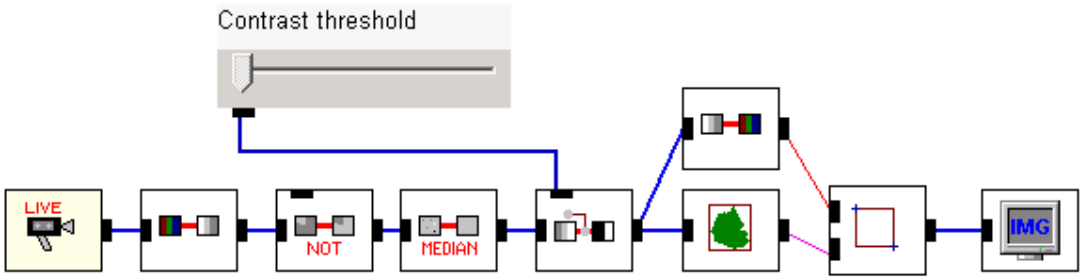
Equipment: Sony DV video camera, Pentium IV, 2GHz PC computer with a Matrox frame grabber for video input and a SoundBlaster live card for audio output, video projector, EyesWeb 2.5 for all software processing.

The video camera was positioned low in face of the dancer, so as not to disturb the audience's line of sight. EyesWeb output of recorded motion cues was projected on a small screen beside the dancer. EyesWeb audio output was fed to the scene's PA.

5.2. Description of the employed patches

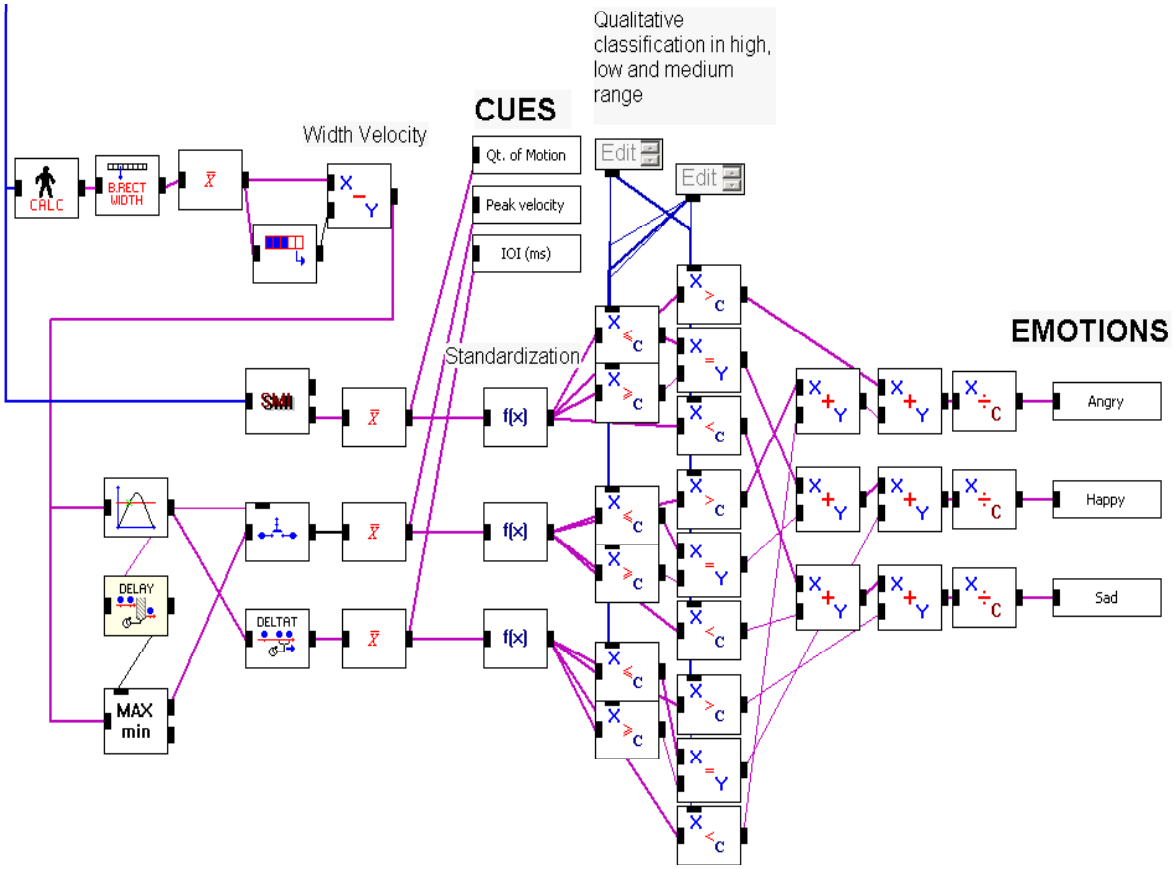
Patch 1 was used in part A of the performance and is divided in three subpatches 1.1-1.3. Patch 2 was used in part C.

5.2.1 Patch #1.1 – basic video input processing



What it does:
Extracts the silhouette and the bounding rectangle of the dancer and provides the monitor output.
How it works:
Video input was found to be optimal when the dancer wore dark clothes seen against a white backdrop. After inverting and filtering the image, a slider connected to an image threshold block was also found most useful for further corrections at the performance. The dancer's silhouette and bounding rectangle was projected on a small screen thus enabling both performers and audience to check for consistency.

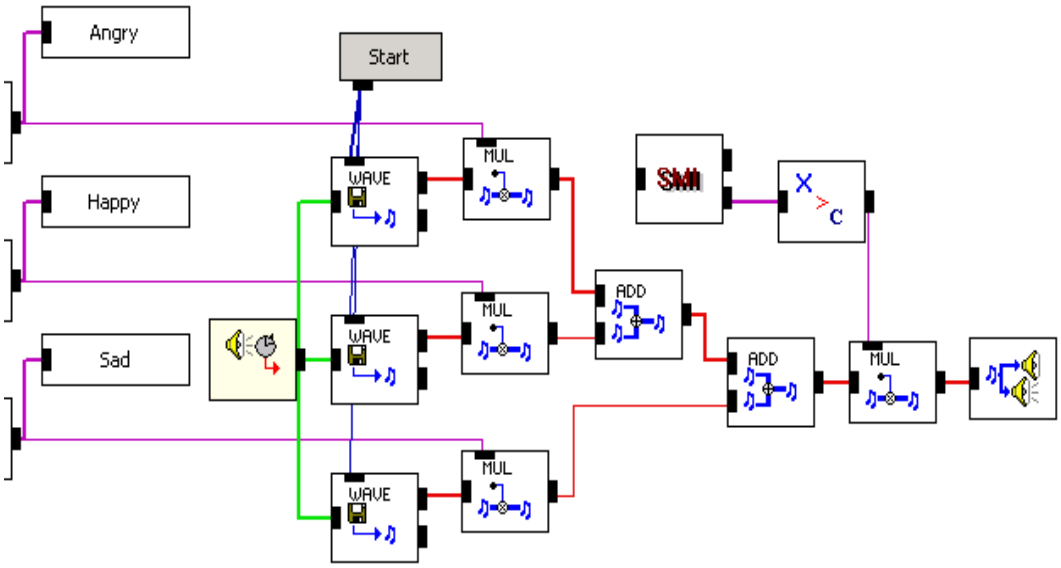
5.2.2 Patch #1.2 –cue tracker and expressive mapper



What it does:
Calculates the three cues Qt. Of Motion (QM), Peak velocity (VEL), and IOI (general motion index) from video silhouette data and estimates the intended emotions using a qualitative predictor.

How it works:
The peak velocity is defined as the maximum outward velocity of the bounding rectangle width using a max-detector block that is cleared at each new positive zero-crossing. IOI is defined as the time between subsequent zero-crossings, and QM is defined by the SMI block. Each cue is then standardized, crudely assuming it is normally distributed, by dividing it by the standard deviation and subtracting the mean. In this way each of the standardized cues have mean=0 and standard deviation=1. The values for the mean and standard deviation were taken from previous dance performances at the rehearsals. This implied that the system was calibrated according to the personal style of the dancer. The standardized cues were then classified as to whether they were in high, low or medium range. A perfect estimate of an emotion occurred when all three cues were in the range associated with the emotion. Thus an “angry” emotion in the patch above will appear if QM is “high”, VEL is “high” and IOI is “low”.

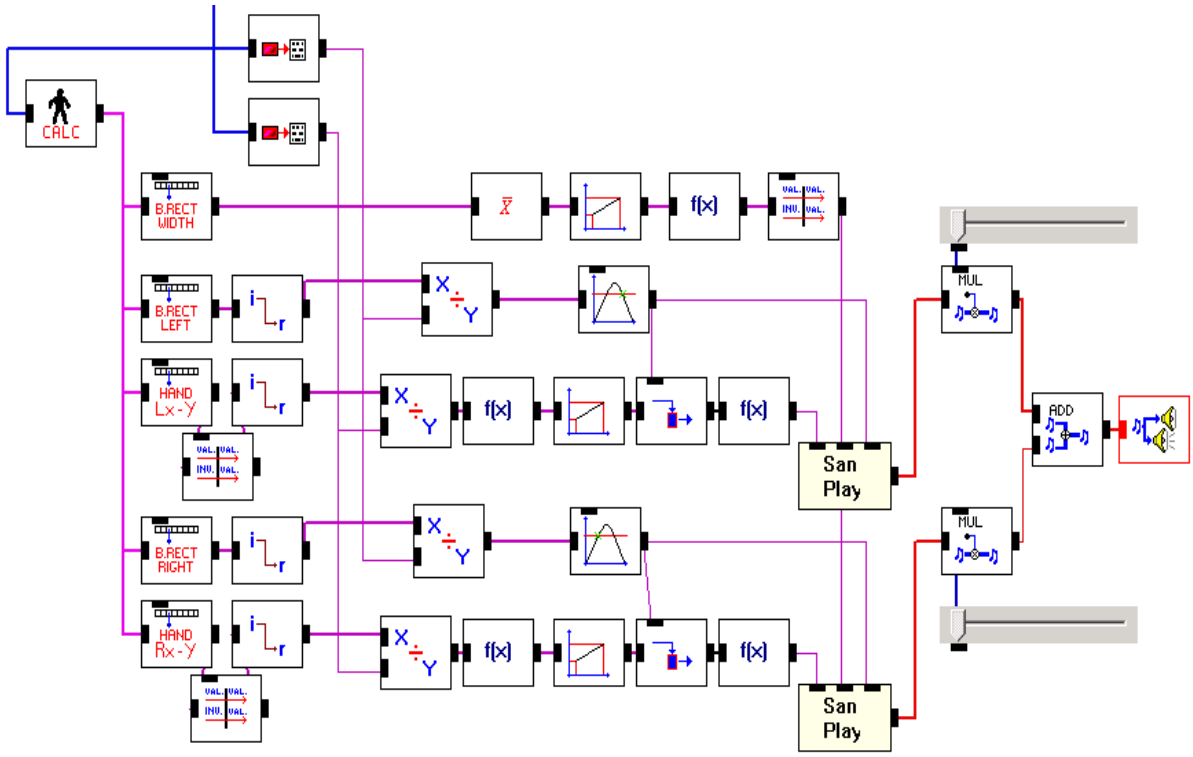
5.2.3 Patch #1.3 –audio grooves player and mixer



What it does:
Mixes the three audio grooves according to the degree of emotion predicted by patch #1.2.

How it works:
The three audio loops are all continuously played and looped in synchrony in the three wave blocks. Mixing is performed by simply multiplying the amplitude of each groove by the corresponding "emotion coefficient"- a number ranging from zero to one coming from the patch above. The three channels are finally added together.
Should the quantity of motion fall below a preset threshold corresponding to "no motion" the comparator following the SMI block outputs zero effectively disabling audio output. This means that if the dancer is standing still the music stops.

5.2.4 Patch #2 –Voice sampler



What it does:
Allows the performer to trigger and control pitch and timing of audio samples by reaching out in the space defined by the camera view.
How it works:
Two voice samples are played in the *SanPlay* blocks when a triggering event occurs. The SanPlay blocks perform an IFFT resyntesis from a sinusoidal representation of the sounds, and provide control of tempo and pitch. The original voice quality is preserved by means of a formant-preserving algorithm. The performer can trigger the first or the second SanPlay block by reaching the left or right end of the camera view with his hands. If the trigger event occurs while the voice sample is already playing, the sample starts from the beginning (permitting, for example, a stuttering-like effect for rapidly repeated triggering). The vertical position of the hand at the triggering instant determines the pitch shift factor, and the width of the vertical bounding rectangle of the whole body controls the time stretch factor.

6. Evaluation

After the performance, a questionnaire was distributed to the audience in order to assess attitudes and reactions. The audience's first impression of the *Groove machine* performance was "very positive" (mean 7.8 on a scale *very negative* 0 to *very positive* 10, see Figure 2, left bar). The dancer answered the same questionnaire and her own impression was 10 "very good", and a volunteer in the audiences who tried out the system responded the same. The strength of the audience experience was also high (mean 6.7, middle bar). The dancer's own experience of the performance was scored 9, and the volunteer scored 7. The last question was about to what extent the audience perceived that the dancer had control of the sound events by her bodymotions. Right bar shows that the audience experienced that the dancer was able to control the music to a very large extent, (mean

7.6). However, the dancer, herself, reported a sense of insecurity in controlling the system. For a more complete description of the evaluation, see the separate document.

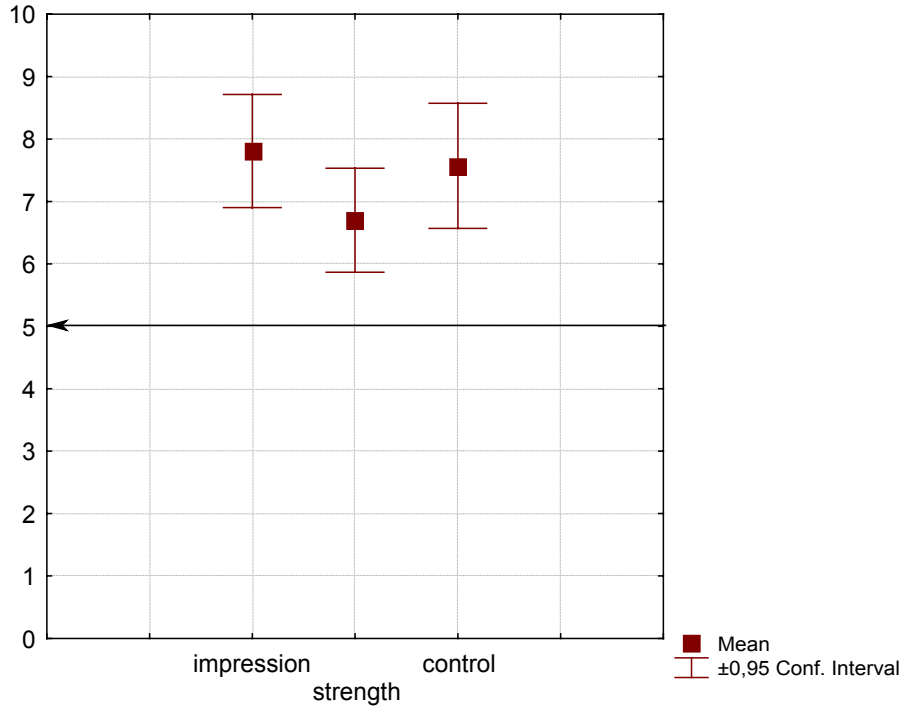


Figure 2. Mean and 95% confidence interval for the audience's perceived impression of the performance (left bar), strength of experience (middle bar), and to what extent the dancer seemed to be able to control the music (right bar).