Non-Realtime Sonification of Motiongrams

Alexander Refsum Jensenius
University of Oslo, Department of Musicology, fourMs lab
a.r.jensenius@imv.uio.no

ABSTRACT

The paper presents a non-realtime implementation of the sonomotiongram method, a method for the sonification of motiongrams. Motiongrams are spatiotemporal displays of motion from video recordings, based on frame-differencing and reduction of the original video recording. The sonomotiongram implementation presented in this paper is based on turning these visual displays of motion into sound using FFT filtering of noise sources. The paper presents the application ImageSonifyer, accompanied by video examples showing the possibilities of the sonomotiongram method for both analytic and creative applications.

1. INTRODUCTION

Motiongrams were originally developed for analysing the motion of dancers and musicians, with the aim of visualising spatial motion features over time [1]. Due to the visual similarity of motiongrams to spectrograms, motiongrams have also been used as the basis for sonification, through a method I call *sonomotiongram* [2]. The first implementation of the sonomotiongram method was focused on creating *realtime* sonifications of the motiongrams, and the sonification was based on an interpolated oscillator bank. Realtime here means that it is possible to listen to the sonification while watching the original video, hence listening to the sound of motion as it unfolds.

A realtime implementation is useful for realtime applications, such as in sonic feedback or in creative applications. It is less useful, however, for applications in which long video recordings need to be analysed. For such material it would be better to use the high temporal capacity of our auditory system to listen through long video recordings at a much higher speed than the video could be watched.

This paper presents a non-realtime implementation of the sonomotiongram method, based on FFT filtering of a noise source. This implementation allows for (much-)faster-than-realtime sonification of the input motiongrams. The paper starts with an overview of the motiongram and sonomotiongram methods, before the non-realtime implementation of the sonomotiongram method is shown. Finally, some examples of how the method can be used for analytical and creative applications are presented and discussed.

2. BACKGROUND

2.1 Motiongrams

A motiongram is a visual display of (human) motion, created by frame differencing and averaging a video file, as illustrated in Figure 1. This makes it possible to see the temporal unfolding of motion features on the X axis, and the vertical location of the motion on the Y axis. Motiongrams therefore give a holistic representation of the spatiotemporal unfolding of motion from a video recording, albeit only in one spatial dimension. This is because information about the spatial distribution of motion in the plane that is averaged over is represented by only one pixel for each row (see [1] for details). Thus a horizontal motiongram visualises vertical motion, while a vertical motiongram visualises horizontal motion.

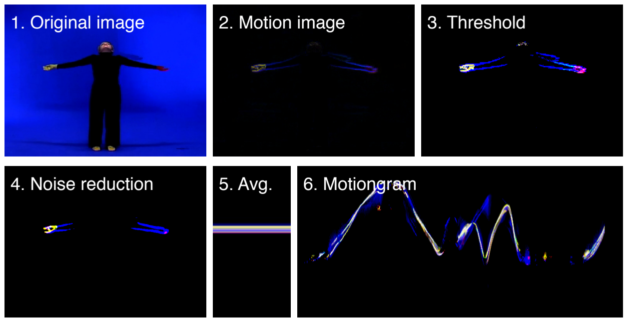

Figure 1. The steps involved in creating a motiongram: (1) original video image, (2) frame differencing, (3) thresholding, (4) noise reduction, (5) averaging over each row, (6) drawing the average matrices over time.

2.2 Sonomotiongrams

The sonomotiongram method is based on what could be called an "inverse FFT" process. The idea here is to treat a motiongram as if it were a spectrogram, with frequency information on the Y axis and time on the X axis, as illustrated in Figure 2. In the first implementation of the method, this was accomplished using an interpolated oscillator bank [2]. The implementation presented in this paper is based on doing FFT filtering of a noise source based on the matrix values of the motiongram. Both implementations result in a direct sonification of the image, in which lower sound frequencies are based on pixel values in the lower part of the image, and vice versa.

Even though they may appear to be visually similar, a motiongram is, in fact, very different from a spectrogram.

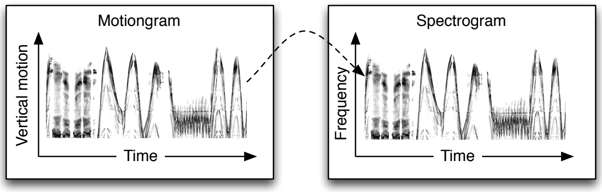

Figure 2. A sketch of the sonomotiongram method, showing how the motiongram matrix is "mapped" to spectral audio data.

A spectrogram of an audio recording displays the energy level of the frequency bands resulting from doing a Fourier transform on the audio. A motiongram, on the other hand, is a reduced display of a series of motion images. There is no analysis being done when creating a motiongram, it is only based on a reduction algorithm. The simplicity of the approach may be seen as a problem, but it is also what has made motiongrams useful in several different application areas [3].

Despite the fact that motiongrams and spectrograms represent different features, they share one property: the temporal unfolding of shapes of either motion or sound. Furthermore, the Y axis in a motiongram represents vertical motion, which is often associated with pitch/frequency [4], meaning that there is also a conceptual link between the Y axes in a motiongram and a spectrogram.

2.3 Image sonification

While the idea of using a motiongram as the basis for sound synthesis is novel, the general idea of sonifying an image has been around for decades. An early example of such an idea is the Pattern Playback machine built by a group of speech researchers in the late 1940s [5]. This system made it possible to "draw" shapes that could afterwards be played back as sound. Iannis Xenakis developed the UPIC system in 1977, which made it possible to create complex timbres by drawing with a digital pen on a computer screen [6]. Nowadays, the idea of making sound from drawings is available in the Metasynth software, along with the possibility of sonifying any type of images and photos [7].

There are also examples of how audio analysis software, like AudioSculpt [8] and SPEAR [9], allow for screen-based manipulation of spectrograms and resynthesis of the manipulated image into sound. This makes it possible for researchers and composers to edit the timbral content and development of sounds in the visual domain.

Closer to the non-realtime sonification approach presented in this paper are examples of how image sonification strategies are used in art installations and realtime applications. One example here is the installation SoundView allowing the user to move a pointer device over an image while listening to the sound [10]. Here the pointer can be thought of as a "tape-head" that scans through the image following an auditory information seeking principle [11]. Other related projects include the 2D spatiotemporal mapping strategies presented in the case of EEG sonification [12] and video sonification based on Hilbert curves [13].

3. IMAGESONIFYER

The original implementation of the sonomotiongram method was presented in [2], and was developed in Max/MSP/ Jitter as modules for the open framework Jamoma [14]. The non-realtime version presented here has been created as a standalone Max patch and application called ImageSonifyer [15]. A screenshot of the user interface of the application is shown in Figure 3.

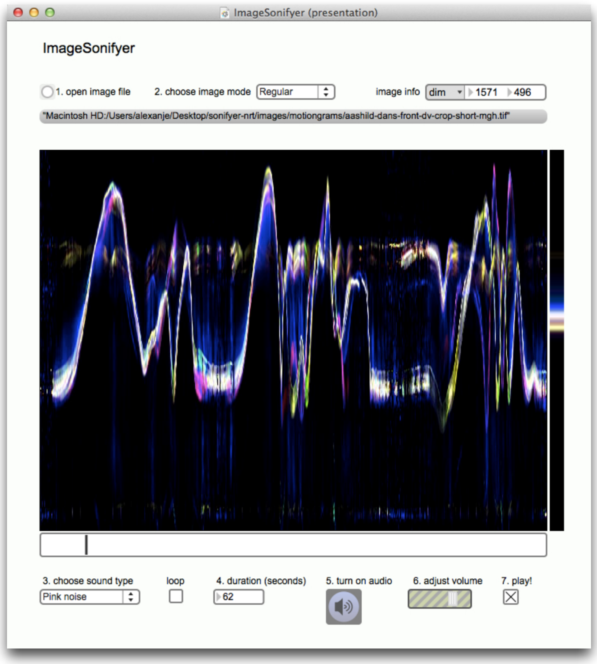

Figure 3. Screenshot of the ImageSonifyer application.

3.1 Features

The ImageSonifyer application is based on opening pre-made image files of motiongrams, and use these images as the starting point for the sonification. Motiongrams can be created in Max using the above-mentioned Jamoma modules, and can also be created using the standalone application VideoAnalysis [16]. It is also possible to use other types of images as input to ImageSonifyer, which will lead to more "traditional" image sonifications. The following features are available:

Load image file Any image file supported by QuickTime can be loaded and displayed. There is no limit on the pixel size of images that can be loaded, but the image will always be displayed at a fixed 4:3 ratio to get a full view of the image and to avoid scrolling. The sonification will still be based on the original image data and not on the reduced image presented on the screen.

Image mode A raw motiongram usually has a white foreground (pixel value 255) on a black background (pixel value 0). For visual reasons, however, it may be convenient to invert the motiongrams so that they

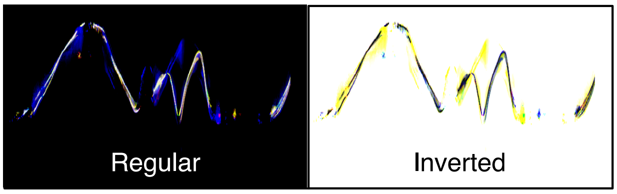

Figure 4. Illustration of the difference between a regular and inverted motiongram.

end up with a black foreground on a white background, as illustrated in Figure 4. There is an option to (re)invert the motiongram in ImageSonifyer in case a user loads an inverted motiongram and wants to sonify it as a regular motiongram.

Sound source The current implementation only allows for choosing between pink and white noise as the source material for the synthesis. In future research it will be interesting to explore other sounds as source material for the synthesis.

Panning Since motiongrams have a temporal direction from left to right, the horizontal location in the image is used to control the panning from left to right in ImageSonifyer. Hence, the sound starts in the left channel, and then gradually pans over to the right side when moving through the image.

Playback Sound can be played back by hitting the space bar button on the keyboard. Looping of the playback can be turned on and off with a toggle.

Duration The ability to freely select the duration of the sonification allows the user to experiment with both fast and slow sonifications of the same material. The duration value will default to the duration of the original video recording, assuming that the video was recorded at 25 fps. This means that an image file with a width of 1500 pixels will be played back over 1 minute. The user is free to set other durations (in seconds) to alter the playback speed.

Scrubbing As an alternative to a linear and clocked playback of the image file, it is also possible to scrub through the image using the mouse. Then the horizontal position of the mouse will control the location of the sonification, and the vertical position will control the sound level.

An example of ImageSonifyer in use can be seen in Video 1. [1]

3.2 Implementation

The sonification part of the ImageSonifyer application is inspired by the Metasynthy patch presented in [17], and the Max pseudo patch in Figure 5 shows an overview of the implementation. The first step is to load an image file into a `jit.qt.movie` object, and read individual matrix columns from this image using a `jit.submatrix` object. These numbers are passed to an MSP buffer using `jit.buffer~`, and used as the basis for an inverse FFT process using the `pfft~` object. There is also a simple crossfade function used to pan the sound from left to right following the position in the image.

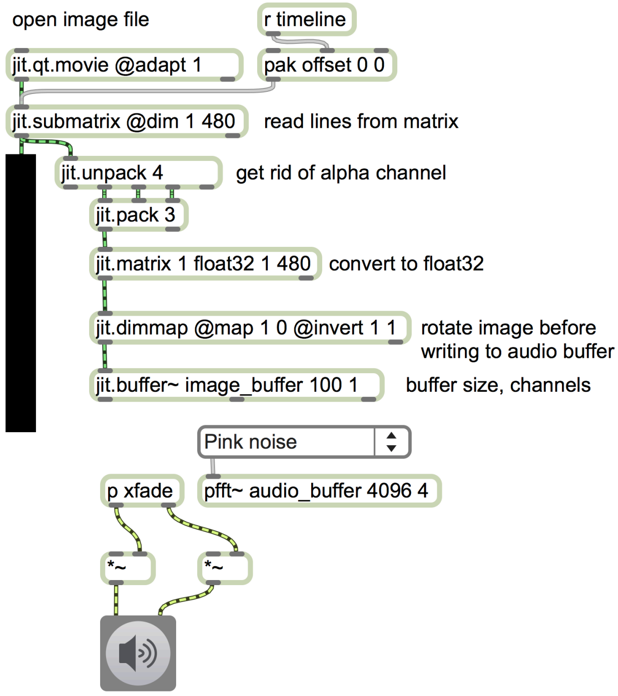

Figure 5. Pseudo Max patch showing the sonification process from input image file.

In the current implementation of the sonomotiongram method there is no use of the colour information in the image. Some experimentation has been done in using the colour information to alter the timbral quality, but this has not ended up working particularly well. After all, the main aim of this method has been to sonify motion features, and quite often the colour information in a motiongram is not particularly relevant. It will, however, be interesting to explore the use of colours at a later stage.

4. SONIFICATION EXAMPLES

The following sections will present some examples of how different types of video material can be sonified using the ImageSonifyer.

4.1 Sonification of standing still

What type of motion can be observed when a person attempts to stand physically still? This has been the topic of some recent experiments in our lab, using a high quality motion capture system able to detect motion at the scale of millimetres [18]. An alternative measurement approach is to place a video camera on the head of a person standing still, since even small motion in the head will lead to a large amount of changing pixels in the recorded image. Figure 6 shows the motiongram resulting from a 10-minute recording of a person standing still with a sports camera (GoPro

[1] Video examples are available from www.arj.no/smc2013/

Hero 2) attached to the head. The recorded image is not interesting in itself, but the motiongram and the sonification are able to represent the rhythmic pattern and the temporal development of the micromotion (Video 2).

4.2 Sonification of a high-speed guitar recording

How does a sonification of moving guitar strings sound like? Figure 7 shows a motiongram of a high-speed recording of a single strumming of the strings on an acoustic guitar. The recordings were made with a high-speed video camera (Phantom V711), at a speed of 7 500 frames per second and with an image resolution of 1280 x 800 pixels. Due to the memory limitations of the camera (21 GB), the maximum recording duration was 1.1 second, which results in a video file with more than 8 000 frames.

Video 3 shows a playback of the video file at 100 frames per second, and Video 4a and 4b shows sonifications of the recording with durations of 10 and 1 seconds, respectively. These sonifications were created slightly different than the other sonifications presented in this paper. Motiongrams are usually created by averaging over the rows in the video matrix. A regular motiongram, however, would not work so well for this particular recording, since the strings are not entirely horizontal in the recording. Thus the averaging happening when creating the motiongram would make the "height" of each string larger than they really were, and will lead to an imprecise rendition of the actual motion. The solution has therefore been to create a motiongram using a slit-scan approach, selecting a single pixel column in the middle of the sound hole on the guitar, and using this pixel column for drawing the motiongram before doing the sonification.

4.3 Sonification of long videos

While the sonomotiongram method was mainly developed for studying music-related body motion, it may be relevant for other applications as well. One example is that of the sonification of long video recordings. Such sonifications may be used to listen to rhythmic patterns and structural changes in the recordings, which may not otherwise be easily recognisable by watching the video in its entirety, or looking at different types of compact visualisations.

Figure 8 shows a motiongram and Video 5 shows a 60-second sonification of a 7.5 hour documentary film of the scenic train ride from the city of Bergen on the west coast of Norway to the capital Oslo. Fortunately, the Norwegian broadcasting company NRK has decided to release a full HD recording of the documentary, with a creative commons license allowing the reuse of the material [19].

When creating a sonification of the Bergensbanen recording, I decided to start out with what I call a *videogram* instead of a motiongram. The difference between the two is that a videogram is based on averaging the input video image instead of the motion image, that is, skipping step (2) in Figure 1. The end result is a videogram in which the colours reflect the colours of the original image, which is more meaningful for a recording in which movement of the camera is as prominent as movement within the image.

4.4 Sonification of abstract images

It is, of course, also possible to sonify other types of images than motiongrams or videograms, thereby using the ImageSonifyer application more like MetaSynth. Although this was never the intended use of the application, it can be creatively interesting to explore how different types of abstract images sound like, such as shown in Video 6a, b, c, and d. Such images are also interesting to explore through the scrubbing functionality of ImageSonifyer.

5. DISCUSSION

Although still in an exploratory state, the sonomotiongram method, and its implementation presented in this paper, has been versatile for both analytic and creative applications. On the analytic side, the possibility to create fast sonifications of long videos is useful, since it allows for listening to long videos in a much shorter time than it would have taken to watch through the recordings. I also find that listening to the sonifications reveal other features than what can be seen from the motiongrams, particularly when it comes to rhythmic and periodic elements in the material.

On the creative side, the scrubbing functionality of ImageSonifyer has proven creatively inspiring to work with. Here the application can be used for image-based improvisations, or using an image as a "score" for a fixed composition. This has already been tested in a concert, and will be explored further in future performances.

One of the positive sides of the sonomotiongram method is its flexibility, being able to sonify image files based on all sorts of video material: long and short recordings, different types of image resolution and qualities, different image framing (close-ups of hands to full body motion). The implementation in Max has been stable and reliable, and it runs comfortably on a normal laptop.

That said, there are also several issues that will have to be explored further in future research:

Dimensions Since they are based on averaging over each row in the video matrix, motiongrams are limited to displaying the distribution of motion in only one spatial dimension [1]. It will be interesting to explore different ways of sonifying multiple spatial dimensions, including two dimensions for regular video recordings and three dimensions for recordings from depth-cameras. This could, for example, be done through spatial or harmonic relationships.

Time and temporal resolution A challenge when working with video recordings as the source material for sonification is how to handle the temporal aspect. This is particularly apparent when working with realtime sonification of video recordings, due to the poor temporal resolution of video as compared to audio. Faster-than-realtime sonifications may overcome this problem, and allows for utilising the potential of our auditory system. Still it is important to find a balance between the temporal aspects of motion features in the video recording and the audio features of the sonification.

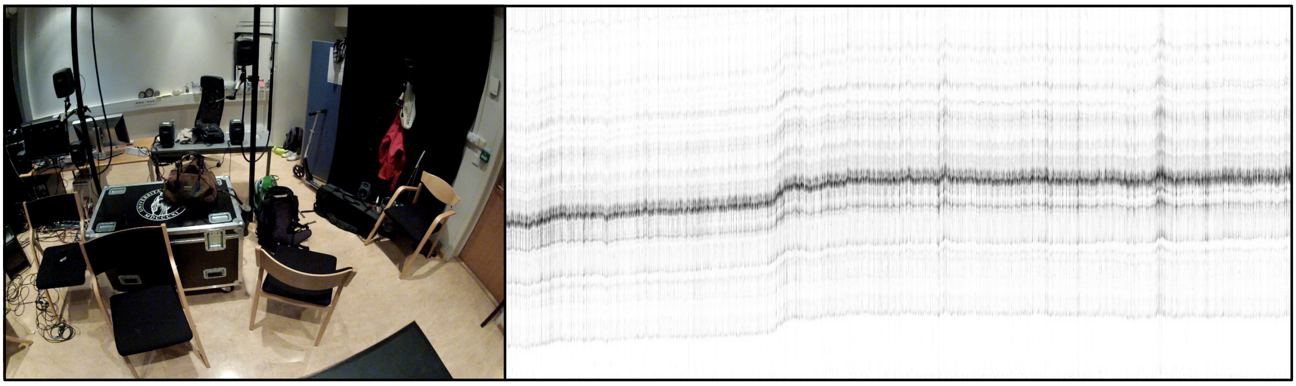

Figure 6. One frame from the video (left) and motiongram (right) of a person standing still for 10 minutes with a camera on the head. The motiongram illustrates the micromovements in the head of the person standing still. The sonification can be heard in Video 2.

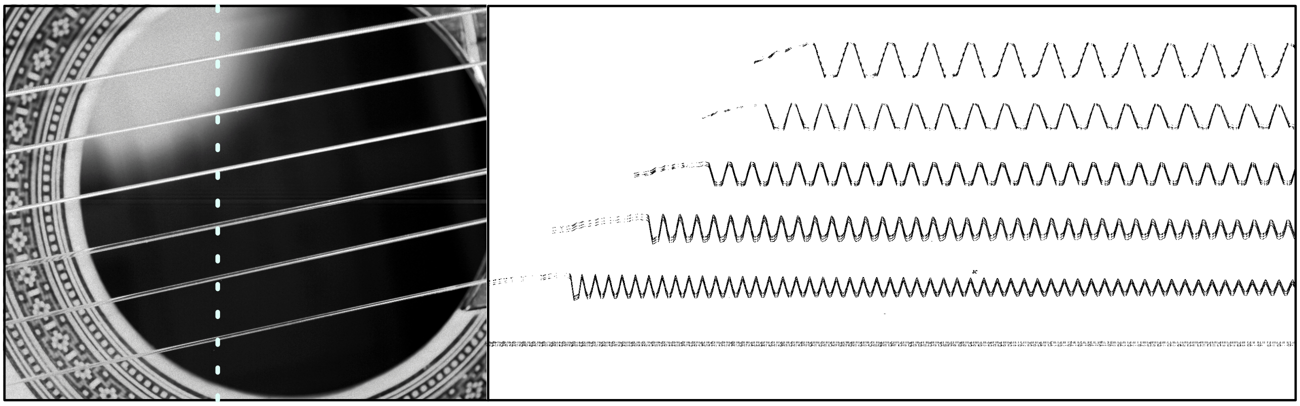

Figure 7. Motiongram of a high-speed guitar recording: one frame from the original video (left), motiongram of the first 1000 frames of the video recording (right). The motiongram has been inverted for visual clarity (black on white). The dotted line in the original video frame indicates the pixel column that was used for creating the motiongram. The original video can be seen in Video 3, and sonifications can be heard in Video 4a and 4b.

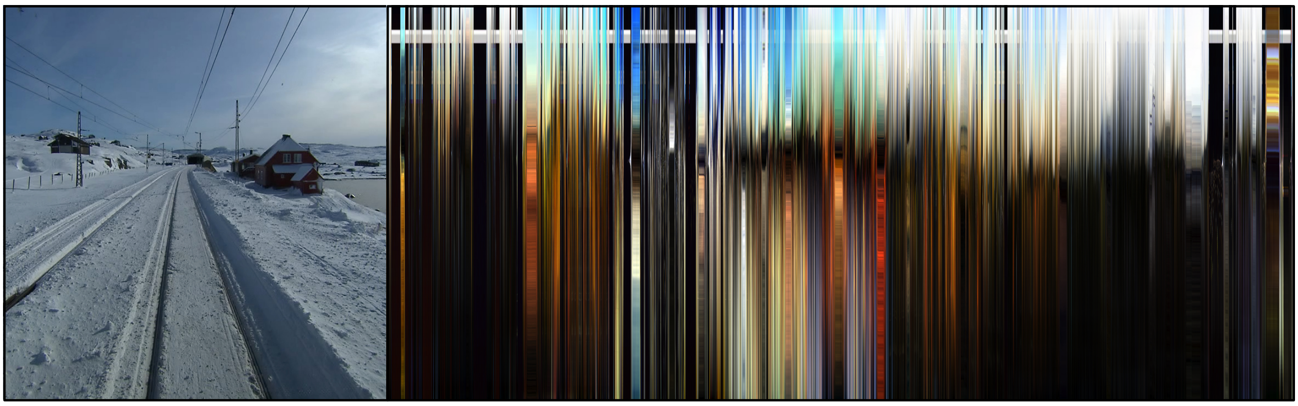

Figure 8. An image (left) and videogram (right) of the 7.5 hour documentary Bergensbanen, a recording of the entire train ride from Bergen to Oslo. The black vertical lines represent when the train were travelling inside tunnels, and the white horizontal stripe near the top is the NRK logo. The sonification can be heard in Video 5.

Analysis and/or performance The original idea of the sonomotiongram method came from an analytic point of view: creating a tool to help in the analysis of various types of music-related motion. While the method certainly works for this type of application, I find that it may be even more interesting from a creative point of view. Here the scrubbing functionality provides the user with a tool for creating what could be called interactive sonifications [20]. It would be interesting to take this one step further by using a realtime motiongram as the basis for such a scrubbing process, which makes it possible to create an interactive loop between motion and sound features.

Colour The current implementation of the sonomotiongram method does not make use of the colour information in the image. Future experiments will look into how the colour information can be used to change, for example, the timbral features of the sound.

Acknowledgments

This research has been funded by the Norwegian Research Council through the project Sensing Music-related Actions. Niko Plath assisted in making the high-speed recording used in Figure 7 at the Department of Musicology at the University of Hamburg. NRK has generously provided a CC license for the Bergensbanen documentary (Figure 8).

6. REFERENCES

[1] A. R. Jensenius, "Using motiongrams in the study of musical gestures," in *Proceedings of the International Computer Music Conference*, New Orleans, LA, 2006, pp. 499–502. [Online]. Available: http://urn.nb.no/URN:NBN:no-21798

[2] ——, "Motion-sound interaction using sonification based on motiongrams," in *Proceedings of the International Conference on Advances in Computer-Human Interactions*, Valencia, 2012, pp. 170–175. [Online]. Available: http://urn.nb.no/URN:NBN:no-30588

[3] ——, "Some video abstraction techniques for displaying body movement in analysis and performance," *Leonardo*, vol. 46, no. 1, pp. 53–60, 2013.

[4] K. Nymoen, "Methods and technologies for analysing links between musical sound and body motion," Ph.D. dissertation, University of Oslo, 2013. [Online]. Available: http://hdl.handle.net/10852/34354

[5] F. Cooper, A. Liberman, and J. Borst, "The interconversion of audible and visible patterns as a basis for research in the perception of speech," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 37, no. 5, p. 318, 1951.

[6] G. Marino, M.-H. Serra, and J.-M. Raczinski, "The upic system: Origins and innovations," *Perspectives of New Music*, vol. 31, no. 1, pp. 258–269, 1993.

[7] E. Wenger, "Metasynth [computer program]. http://www.uisoftware.com/metasynth/," 1998.

[8] N. Bogaards, A. Röbel, and X. Rodet, "Sound analysis and processing with Audiosculpt 2," in *Proceedings of the International Computer Music Conference*, Miami, FL, 2004, pp. 462—465.

[9] M. Klingbeil, "Software for spectral analysis, editing, and synthesis," in *Proceedings of the International Computer Music Conference*, Barcelona, 2005, pp. 107–110.

[10] K. van Den Doel, "Soundview: Sensing color images by kinesthetic audio," in *Proceedings of the International Conference on Auditory Display*, Boston, MA, 2003, pp. 303–306.

[11] H. Zhao, C. Plaisant, B. Shneiderman, and R. Duraiswami, "Sonification of geo-referenced data for auditory information seeking: Design principle and pilot study," in *Proceedings of the International Conference on Auditory Display*, Sydney, 2004, pp. 33–36.

[12] T. Hermann, G. Baier, U. Stephani, and H. Ritter, "Vocal sonification of pathologic EEG features," in *Proceedings of the 12th International Conference on Auditory Display*, London, 2006, pp. 158–163.

[13] F. Grond, "Organized data for organized sound. Space filling curves in sonification," in *Proceedings of the 13th International Conference on Auditory Display*, Montreal, 2007, pp. 476–482.

[14] T. Place and T. Lossius, "Jamoma: A modular standard for structuring patches in Max," in *Proceedings of the International Computer Music Conference*, New Orleans, LA, 2006, pp. 143–146. [Online]. Available: http://hdl.handle.net/2027/spo.bbp2372.2006.032

[15] A. R. Jensenius, "ImageSonifyer [computer program]," 2013. [Online]. Available: http://www.fourms.uio.no/software/

[16] A. R. Jensenius and K. Nymoen, "VideoAnalysis [computer program]," 2009. [Online]. Available: http://www.fourms.uio.no/software/

[17] B. Moon, "Pad 28 - metasynthy," 2011. [Online]. Available: http://www.youtube.com/watch?v=rnERzPwRa4g&lr=1

[18] A. R. Jensenius and K. A. V. Bjerkestrand, "Exploring micromovements with motion capture and sonification," in *Arts and Technology, Revised Selected Papers*, ser. LNICST, A. L. Brooks, Ed. Berlin: Springer, 2012, vol. 101, pp. 100–107. [Online]. Available: http://www.springerlink.com/content/j04650123p105646/

[19] Ø. Solstad, "Download Bergensbanen in HD," 2009. [Online]. Available: http://nrkbeta.no/2009/12/18/bergensbanen-eng/

[20] T. Hermann, A. Hunt, and J. G. Neuhoff, *The Sonification Handbook*. Berlin: Logos Verlag, 2011.