International Journal of Innovation Engineering and Science Research
(ISSN: 2581-4591)

www.ijiesr.com

# Sonification approach using human body movements and window functions

[1]Daniela Hagiescu
*Advanced Slisys, Bucharest, Romania*

[2]Lidia Dobrescu
*Faculty of Electronics, Telecommunication and Information Technology, University POLITEHNICA of Bucharest, Bucharest, Romania*

## ABSTRACT

In this paper, a new sonification methods is proposed to generate piano music from webcam images and computed human joint coordinates. The music is generated using various window functions implying the pitch and velocity of the generated chords. The influence of the window function and their parameter is investigated.

***Keywords****—Sonification, window function, convolutional pose machines.*

## I. Introduction

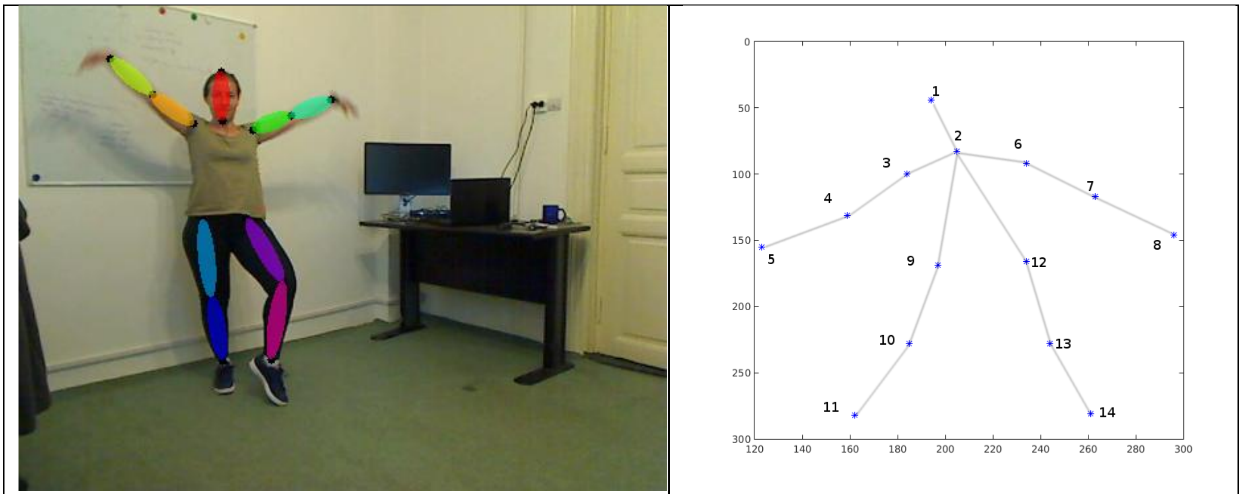
There are numerous sonification techniques that uses data information from various sources (e.g. earthquake data, financial data, etc). The mapping of this data in sounds is usually arbitrary [1]. There are many methods to generate sounds from body movements that uses data from sensors or video recordings [1]-[5]. The approach with sensors attached to the body is sometime unpleasant and restraint the body movements. Using webcams or other image recording devices lead to a cheaper sonification solution and certainly less obtrusive. Examples of such solutions are EyesWeb [6], Motion Composer [7] or Point Motion Control [8] devices.

In [2] the "motiongrams" are used for sonification from video recordings. An image frame is considered as the spectrogram with frequency information on the Y axis and time on the X axis as the basis for synthesizing a sound file [5]. A sonification method based on optical flow computation was proposed in [15]. A musical note is played depending on the optical flow magnitude [5]. The pitch corresponds to the location and flow direction of the peak and the velocity (or intensity) of the note corresponds to the magnitude [2]. This method is very complex due to optical flow computation. In [5] a simpler method that uses the body joint coordinates obtained by the convolutional pose machines (CPM) approach [9] was proposed. The piano sounds were generated using a layout of pitches from low-level harmonic notions [2]. In this paper a novel approach based on applying different window functions to the computed pitch and velocities values is investigated. This is an original approach since they are not applied on the piano sounds themselves. Therefore, differently aesthetically pleasing piano sounds are generated from the same body movements.

The structure of the paper is the following: Section II concisely describes the approach using CPM, Section III the proposed approaches for sonification are presented and the conclusions close the article.

## II. The CPM based sonification method

As shown in [5] the Convolutional Pose Machines (CPM) method is used to find the skeleton joints coordinates from webcam images. The convolutional pose machine is a human pose detector. The human pose detector algorithm finds 14 human skeleton joints for head, upper neck, shoulders, elbows, wrists, hips, knees, and ankles. An example of webcam image and skeleton joints is shown in Figure 1 [5].




a) Webcam image and CPM computed joints coordinates [5]

b) Skeleton joints [5]

Figure 1. Webcam image and skeleton joints

Two convolutional neural networks are trained for detecting a person in the image frame and to detect skeleton joints. Each stage produces belief maps being composed of a sequence of convolution and pooling layers. More details about the CPM method can be found in [9]. As shown in [5] there is a correlation between the obtained joint vectors due to the nature of the huma body and the possible movements. The CPM method is very good in handling non-standard poses in case of different relative camera views [9]. The few mismatches that might happen can be removed by using the Dynamic Time Warping (DTW) distance between consecutive normalized joint coordinates vectors [5]. The DTW algorithm [10] computes the optimal alignment between two time-series and proved a valuable tool in handwritten evaluation (e.g. [11]). A threshold of 0.05 was used to detect outliers [5].

*2.1The computation of the pitch and velocity*

The pitch mesh pairs method from [2] is used. The pitch mesh is obtained from vertically stacked pitch chains in parallel octaves [2]. The musical tone uses the dominance bit, the pitch, and the register values. The last two values are normalized between 0 and 128. The chromaticism of the generated music is altered by modifying the thresholds for the normalized coordinates. The method used in this paper for pitch and velocity computation is described in more details in [2] and [5].

If the music contains multiple notes on the same pitch, some values are filtered out and the filtered parameters are played using the Microsoft MIDI Mapper as the midi output device [5]. As shown in [5], the complexity of this approach is smaller in comparison with competing methods. There are numerous ways to obtain different piano sounds by varying the threshold, pitch range, register range and beat duration. However, for some human body movements consecutive identical notes are generated and there is a need to diversify them while still taking into account the movement for these consecutive image frames.

## III. THE PROPOSED METHOD BASED ON WINDOW FUNCTIONS

The normalized joint coordinated are used to create a matrix. A pause for a fraction of a second on each column is considered. A note is triggered based on the normalized horizontal and vertical coordinates values and their location on the pitch mesh pair. The nearest pitch on the selected pitch mesh determines the pitch of the triggered note, and the value of the pixel determines its velocity [5]. The window functions are applied on the column of the matrix. In this paper two window functions are considered: the Hamming window and the Triangular window.

*3.1 The Hamming window*

The Hamming window it is a raised cosine window of length n having the filter coefficients w(i) = 0.54 - 0.46*cos(2*pi*i/(n-1)) [12]. Figure 2 shows the histograms of the pitch and velocities vectors after passing them through a Hamming window of length 7. It can be noticed that this simple approach modifies the pitch and velocities histogram values and, therefore, create different sounds.

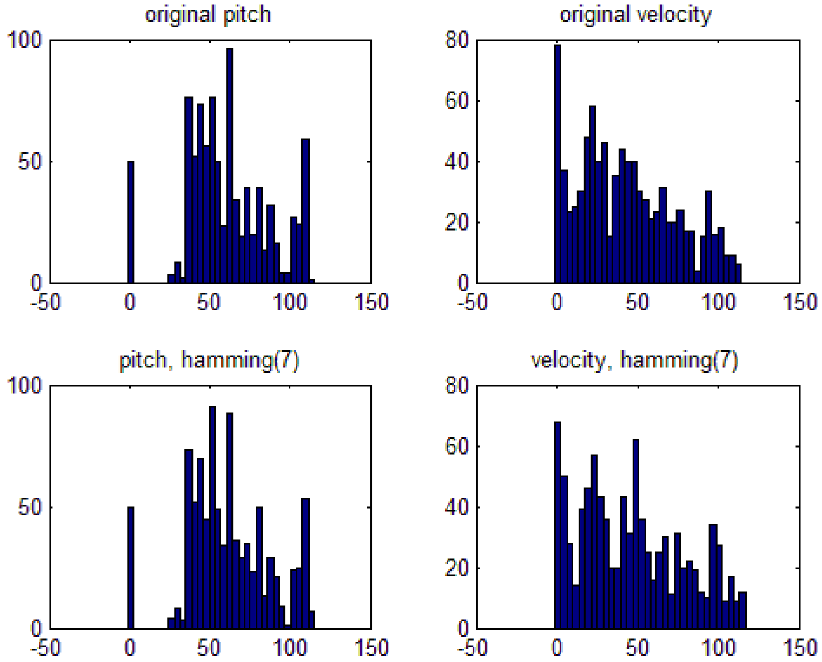



Figure 2. Left side: The histograms of the original pitch and the Hamming windowed pitch (n = 7). Right side: The histograms of the original velocity and the Hamming windowed velocity (n = 7).

Also, it can be noticed from Figure 3 that the Hamming window change the velocity values more than the pitch values. Figure 4 show a comparison of the effect of the window length on both pitch and velocities values. It can be noticed that there is a higher variance on the velocity's values than on the pitch values. Also, the window size influences the generated piano sounds. Therefore, multiple sounds can be generated by using variable length windows while still connecting them with the human body movements for several consecutive frames.

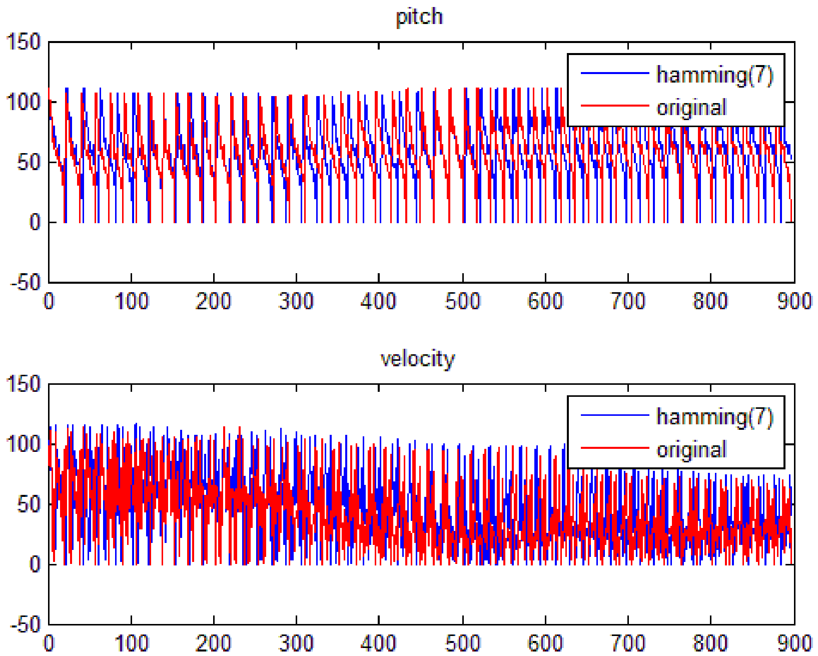



Figure 3. Top: The original pitch value and the Hamming windowed pitch values (n = 7). Bottom: The original velocity values and the Hamming windowed velocity values (n = 7).

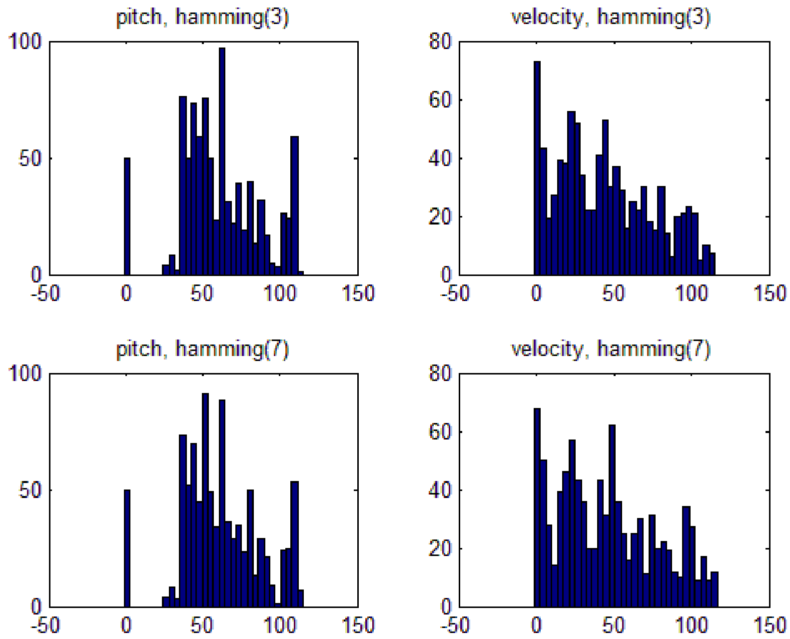



Figure 4 The histograms of the pitch and velocities. Top: Hamming windowed (n = 3). Bottom: Hamming windowed (n = 7).

*3.2 The Triangular window*

The triangular window can be considered as the convolution of two length (n-1)/2 rectangular windows, where n is an odd number and having the filter coefficients w(i) = 1-2*|i|/(n-1). Figure 5 shows the histograms of the pitch and velocities vectors after passing them through a triangular window of length 7.

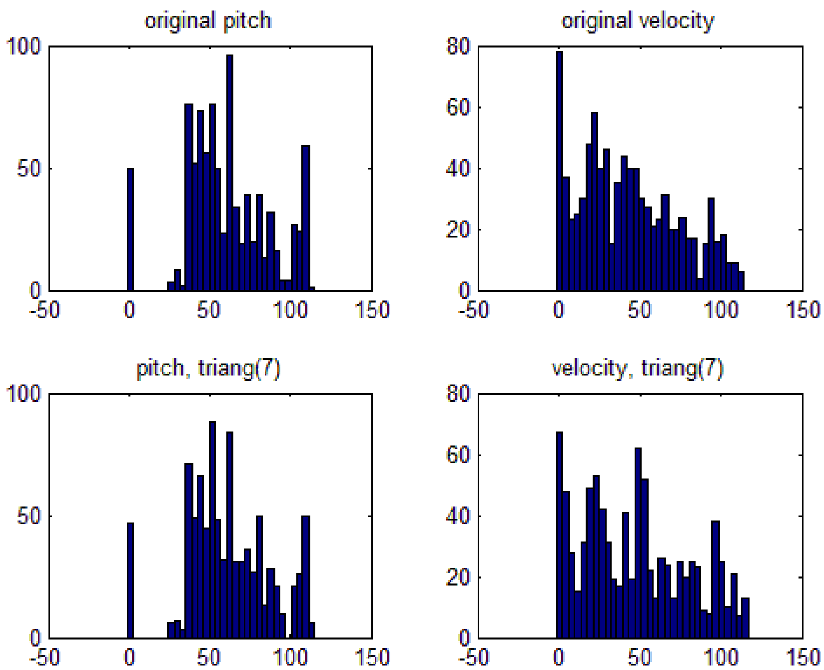



Figure 5 Left side: The histograms of the original pitch and the Triangular windowed pitch (n = 7). Right side: The histograms of the original velocity and the Triangular windowed velocity (n = 7).

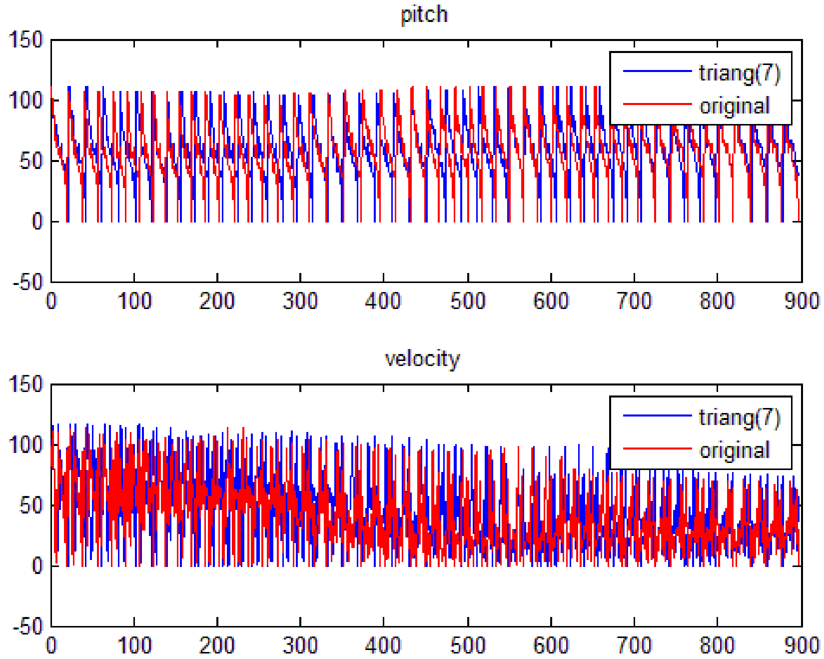



Figure 6 Top: The original pitch value and the Triangular windowed pitch values (n = 7). Bottom: The original velocity values and the Triangular windowed velocity values (n = 7).

Figure 6 shows the original pitch and velocity values and their corresponding triangular windowed values. It can be noticed that similar conclusions to those for the Hamming window can be drawn. There is also a higher variance on the velocity's values than that of the pitch values. By examining Figs. 2, 3, 5, and 6 it is clear that different sounds are generated in most cases by the Triangular window than those generated using the Hamming window. Therefore, the case of generating identical sounds for long periods of time for some repetitive human body movements can be avoided.

Future work will be focused on deriving new filtering methods for the sonification approach and combine various window functions. The use of other possible windows such as Hanning, Kaiser, Blackman [12] will be investigated too.

## IV. Conclusions

An improved piano music generation approach based on human body movements and window functions. The generation of consecutive identical sounds is alleviated through a simple method using window functions operation on columns of joint coordinates matrix. The effectiveness of the method is proved and the influence of the parameters is investigated.

## Acknowledgment

The results presented in this work concerns the research carried out for the "SENTIR" research project, co-financed through the European Regional Development Fund, POC–A.1–A1.2.1–D–2015 grant, research, development, and innovation supporting economic competitiveness and the development of businesses.

## References

[1] T. HERMANN, A. HUNT, J.G. NEUHOFF: The Sonification Handbook, *ISBN 978-3-8325-2819-5, Logos Verlag,* Berlin, 2011.

[2] A. M. TAYLOR, J. ALTOSAAR: Sonification of Fish Movement Using Pitch Mesh Pairs. In Proceedings of the International Conference on New Interfaces for Musical Expression, *Baton Rouge, USA*, June 2015, 28-29, doi: 10.5281/zenodo.1179138.

[3] A. R. JENSENIUS, R. I. GODOY: Sonifying the shape of human body motion using motiongrams. *Empirical Musicology Review*, 7(3), doi:10.18061/emr.v8i2 (2013).

[4] C. ERDEM, K.H. SCHIA, A. R. JENSENIUS: Vrengt: A Shared Body Machine Instrument for Music-Dance Performance. In Proceedings of the international conference on new interfaces for musical expression (nime).doi: 10.5281/zenodo.3672918 (2019)

[5] F. ALBU, M. NICOLAU, F. PIRVAN, D. HAGIESCU: Sonification Method using human body movements. In International Conference on Creative Content Technologies, *Barcelona, Spain*, February 2018, 18-22.

[6] A. CAMURRI, S. HASHIMOTO, M. RICHETTI, R. TROCA, K. SUZUKI, G. VOLPE: Eyesweb: Toward gesture and affect recognition in interactive dance and music systems. *Computer music Journal*, 24 (1), 57-69. Doi: 10.1162/014892600559182 (2000)

[7] MOTIONCOMPOSER software [Online]. http://motioncomposer.de 2021.11.27

[8] POINT MOTION CONTROL software [Online]. http://www.pointmotioncontrol.com/ 2021.11.27

[9] S. WEI, V. RAMAKRISHNA, T. KANADE, and Y. SHEIKH: Convolutional Pose Machines. In IEEE Conference on Computer Vision and Pattern Recognition, *Las Vegas, USA*, June 2016, doi: 10.1109/CVPR.2016.511.

[10] C. S. MYERS, L. R. RABINER: A Comparative Study of Several Dynamic Time-Warping Algorithms for Connected-Word Recognition. *Bell System Technical Journal*, 60 (7), 1389–1409, doi:10.1002/j.1538-7305.1981.tb00272 (1981)

[11] F. ALBU, D. HAGIESCU, M. A. PUICA, L. VLADUTU: Intelligent tutor for first grade children's handwriting application. In IATED 2015, *Madrid*, March 2015, 3708 – 3717.

[12] V. INGLE, J. PROAKIS: Digital signal processing using MATLAB. *ISBN-13: 978-1111427375*, *Brooks/Cole,* (2011)