Understanding Stimulant Parameters in Horror Clips via RNN

Eren Güney (eren.guney@metu.edu.tr)
Department of Architecture, METU
Ankara, 06800Turkey

Abstract

The project tries to understand which parameters are important while watching a horror clip. The aim will be reinterpreting and improving digital space according to sound, brightness, peak level data of the video and emotional data taken from individuals. People will watch generated videos from a root video which are furnished with different parameters to have controlled experiment. At the end, the data sets will be fed into recurrent neural network machine learning algorithm so that one will have a clearer idea of what sounds, brightness and peak combinations in horror videos people are scared of. Finally, new horror digital spaces can be designed according to the results of the machine learning algorithm. If newly organized video scare the people more than the first one, according to quantifiable emotional data of users, it will be counted as successful.

Keywords: digital space, horror movie, recurrent neural network, space design

Method and Composing Data Set

Method

Method will be classified as two stages. First is data collection and interpretation in recurrent neural network. Namely, after collecting metadata of the videos and collecting fear level of participants as an output, by feeding them into recurrent neural network structure, some patterns an important parameters tried to be reached.

Composing Data Set

At the beginning, to compose the data set, individuals watched different generated videos from a root video named as *"The RooM : VR 360° horror".*[1] All the 9 videos were generated to compose different features to have controlled experiment.

	Video-1	Video-2	Video-7
Audio	1	0.25	1
Background Audio	1	0.25	1
Overall-Brightness	1	0.5	1
Peak Voice	1	1	0.25
Peak Gap	1	1	0.6

Figure 1: Showing Parameters of Three Generated Videos out of Nine

Through this video, user's emotion was tracked by the help of Windows Emotion API and verbal questionnaire had done at the end of the video, in case Windows Emotion API fails. By recording individual's faces through computer's webcam and according to user's facial expressions it will extract the emotional data, namely "degree of fear". Degree of fear is used as output.

Other data types that compose training examples were sound, brightness, peak sound level data, the time (seconds) passing between two peaked sound level and fear level before a given point data of the video. Fourteen sample moments were taken from the video, especially the moments which have peaked sound and low brightness. The sample seconds from the 304 seconds long video are:
time_seconds =[0, 7, 14, 44, 75, 88, 125, 147, 158, 203, 248, 265, 288, 299]

After interpreting the data on Python by using the tools as *wave, sys, numpy*[2] audio of the video is transformed into

[1] C. (2016, June 23). Retrieved December 17, 2017, from https://www.youtube.com/watch?v=9vgBDiDpLmU

[2] Reading *.wav files in Python (n.d.). Retrieved December 17, 2017, from

graphical and quantifiable data. So that, the points where sound is peaked up can be seen and is ready to compare with brightness data.[3] Brightness data is also taken according to time samples.

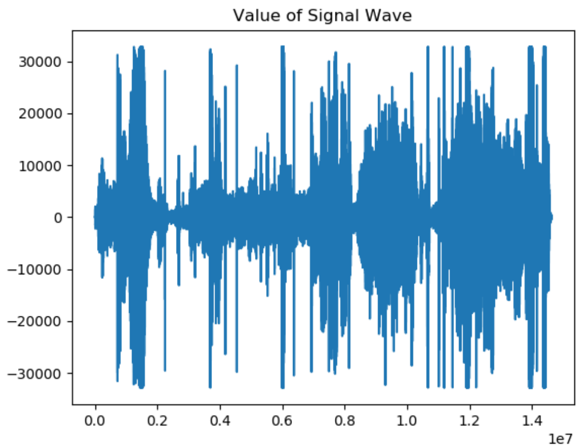


Figure 2: Sound Analysis of the Video-1

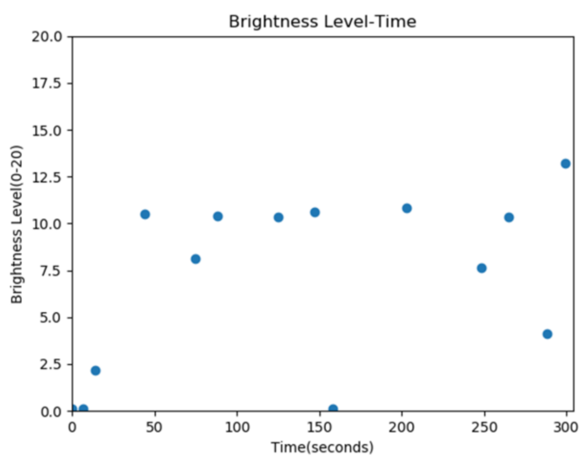


Figure 3. Brightness Analysis of the Video-1, according to time samples (seconds)

After graphical representation, quantifiable data sets of sound for video-1 is [12, 32, 126, 108, 134, 74, 131, 26, 120, 65, 131, 90, and 131,140]; and the brightness is [0.11, 0.11, 2.18, 10.49, 8.12, 10.4, 10.32, 10.63, 0.11, 10.82, 7.62, 10.34, 4.11, 13.24]

Moreover, to be more precise some new parameters are added to current structure. These are sound peak number before a given time point indicated as:

[0, 0, 0, 0, 0, 1, 1, 1, 2, 2, 3, 3, 4, 5]; and the time(seconds) passed between a sound peak and the previous peak shown as:

[[0, 0, 0, 0, 0, 13, 50, 22, 33, 45, 45, 17, 40, 11]

Finally, output set is created by the help of Windows Emotion API and a verbal questionnaire regarding “how was participant scared by giving points from 0-10”. The quantifiable fear level output for video-1 is:

[10,20,40,40,80,30,80,70,100,70,100,70,100,90]

Algorithm

One hidden layered recurrent artificial neural network algorithm seems very applicable for this project because it has time bounded structure. The parameters are as indicated before. Moreover, what makes this NN structure as RNN structure is that, the hypothesis of the fact that the output in a time sample is also impacts the output of the next time sample. Therefore, output (timesample-1) is used as a direct input for the structure.

The expected hypothesis result according experience in experiment process, sound (Derbyshire, 2010) is more influential on output (level of fear) than brightness level. However, without dark environment high sound level or without high sound level, dark environment is insufficient to raise the level of fear. So there is balance between two main parameters and this balance can be visualized by weights present in neural network algorithm. By mapping the values of parameters between (0-1) as an input, this RNN diagram can be drawn:

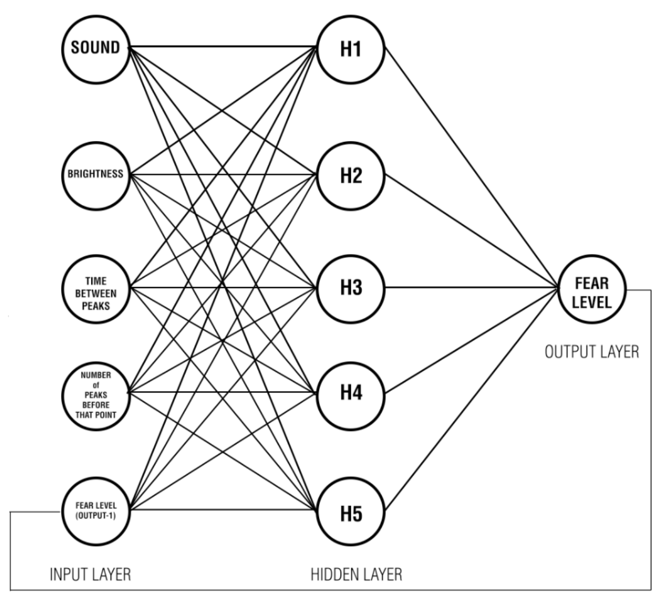


Figure 4: Structure of RNN

As a result this process is done for all the videos, in order to extract weight parameters to understand which parameters are important under different conditions.

[3] What are some methods to analyze image brightness using Python? (n.d.). Retrieved December 17, 2017, from https://stackoverflow.com/questions/3490727/what-are-some-methods-to-analyze-image-brightness-using-python

Results

With learning rate= 0.01, after 100.000 iterations weight matrices were obtained.

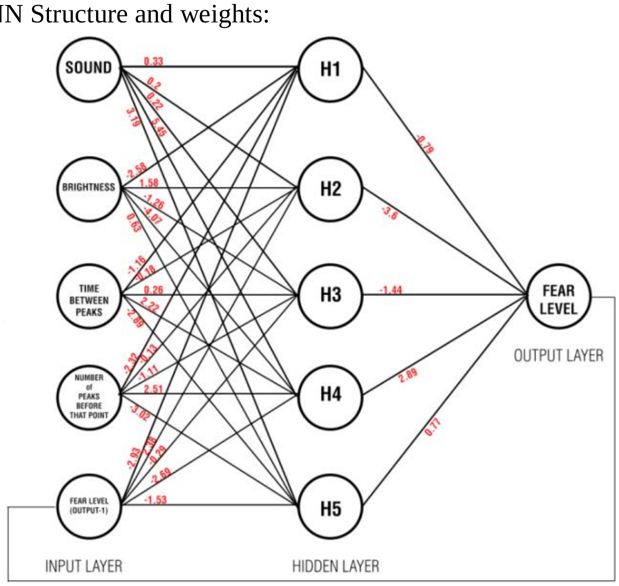
Wanted Output	Hypothesis Output After Training
[[0.1]	[[0.1]
[0.2]	[0.2]
[0.4]	[0.4]
[0.4]	[0.4]
[0.8]	[0.75]
[0.3]	[0.37]
[0.8]	[0.85]
[0.7]	[0.65]
[1.]	[0.96]
[0.7]	[0.69]
[1.]	[0.96]
[0.7]	[0.70]
[1.]	[0.98]
[0.9]]	[0.86]]

Figure 4: Comparison of Outputs

According to results, weight 1 is:

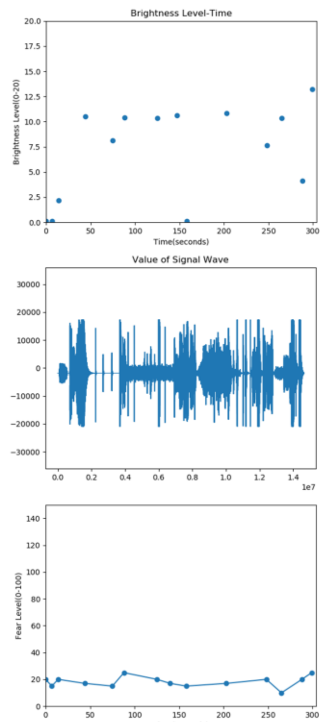
0,33	0.2	0.22	5.45	3.19
-2,57	1,58	-1,26	-4,07	0.63
-1,16	-0,18	0,26	2,22	-2,89
-2.32	-0.13	-1.11	2.51	-3.02
-2.93	2.38	-0.29	-2.69	-1.53

RNN Structure and weights:



Interpretations on Results

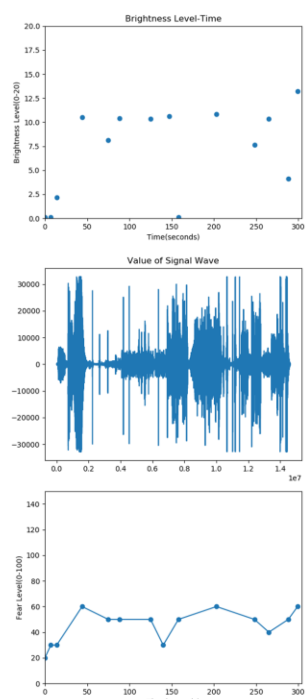
According to weights, for video-1, sound is more effective stimulant than the others. While sound level is at the top other data is not that important.



Video-2:

[[0.66 -0.08 0.71 0.41 0.32]
[-0.44 0.92 -0.82 -0.53 -0.06]
[0.54 -0.61 0.71 0.96 -0.68]
[-0.09 -1.14 -0.44 -0.92 0.82]
[0.17 1.05 0.85 0.73 -0.57]]

When the sound is too low, all the weights are similar to each other, there is no observable distinction on which parameter is more stimulant

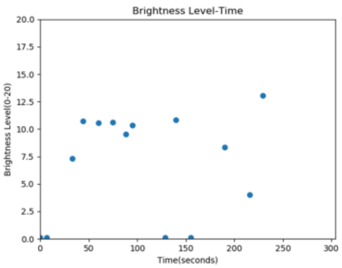


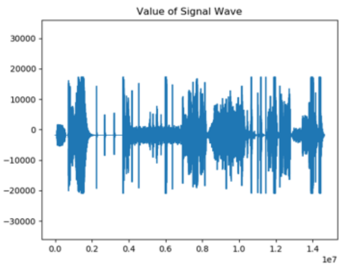
Video-3:

[[0.3 0.75 0.61 0.93 0.35]
[-1.2 1.61 -1.32 0.01 -0.18]
[-0.35 0.09 -0.68 1.33 -0.28]
[-0.76 0.27 -1.06 -0.47 1.06]
[0.11 2.47 1.61 1.08 -0.92]]

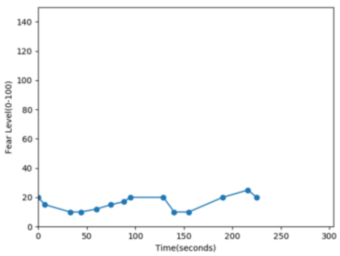
When the sound is low, time between peak levels are shorter, it does not affect any fear of level

Video-5:



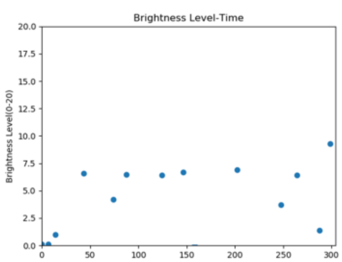


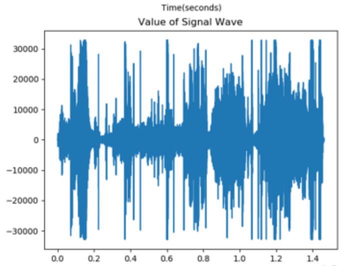
[[0.14 -0.97 0.41 0.55 -0.]
[-0.26 0.63 -0.77 -0.52 0.02]
[0.19 -1.08 0.48 1.02 -0.86]
[-0.47 -1.73 -0.67 -0.84 0.58]
[-0.27 0.64 0.48 0.75 -0.73]]



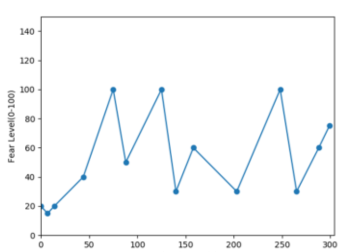
When the sound is low, time between peak levels are shorter, it does not affect any fear level.

Video-4:





[[1.95 -0.88 -0.66 4.14 1.41]
[-1.43 0.89 -5.03 -4.4 -0.19]
[-2.91 -1.07 -0.99 -0.09 -0.26]
[-3.41 -0.6 0.09 -4.41 0.68]
[-4.54 -0.78 3.41 0.78 -0.22]]



When the sound is normal but brightness is lower; it scares people more than average level. From the weights, it can be seen that brightness data is an effective stimulant than the others

CONCLUSION and FUTURE WORK

The project tries to understand which parameters are important while watching a horror clip. However, horror clips are not the ultimate target of this design methodology. In the future, by using this basis, important parameters in designed spaces can be absorbed more scientifically. May be by using more advanced analysis techniques as ECG, Galvanic Skin Response a digital learning space for people with autism can be designed. Moreover fully customizable, constantly changing digital spaces can also be designed by the help of ML algorithms.

References

[1]C. (2016, June 23). Retrieved December 17, 2017, from https://www.youtube.com/watch?v=9vgBDiDpLmU

[2] Reading *.wav files in Python (n.d.). Retrieved December 17, 2017, from

[3] What are some methods to analyse image brightness using Python? (n.d.). Retrieved December 17, 2017, from https://stackoverflow.com/questions/3490727/what-are-some-methods-to-analyze-image-brightness-using-python

[4] Derbyshire, D. (2010, May 26). Ever wondered why the music in horror films scares us? The harsh sounds tap into instinctive fears. *Ever wondered why the music in horror films scares us? The harsh sounds tap into instinctive fears*. Retrieved January 16, 2018, from http://www.dailymail.co.uk/sciencetech/article-1281385/Ever-wondered-music-horror-films-scares-The-harsh-sounds-tap-instinctive-fears.html