VREED Dataset Exploration

Georgina R. Fenton
Student ID: 2404965
Applied Artifical Intelligence and User Experience, MSc
CMP513: User Experience Research
Unit 1.

▼ Rubrick:

Grade/mark / Criteria (weight)	A 4.5-4.0 excellent	B 3.5-3.0 very good	C 2.5-2.0 good	D 1.5-1.0 satisfactory
Research question (10%)	Research question identified is clearly logically derivable from the data set; research question fully justified.	Research question identified is logically derivable from the data set; research question is justified.	Research question identified is, in general terms, logically derivable from the data set; research question is partially justified.	Research question identified only partially logically derivable from the data set; research question is poorly justified.
Data provided (15%)	Excellent, detailed overview of data, what they mean and what they allow to investigate.	Very good overview of data, what they mean and what they allow to investigate.	Good overview of data, what they mean and what they allow to investigate. More detail could be provided.	Satisfactory overview of data, what they mean and what they allow to investigate, however, description lacks detail.
Analytical approach (15%)	Selection of analytical methods is pertinent to the data set and correct in all aspects.	Selection of analytical methods is pertinent to the data set and mostly correct.	Selection of analytical methods is partially pertinent to the data set and correct in many parts.	Selection of analytical methods is not overly pertinent to the data set and may be partially wrong.
Tests selected (20%)	Excellent and detailed justification of tests selected. Tests all appropriate.	Very good and detailed justification of appropriate tests selected.	Good justification of tests selected; tests mostly appropriate.	Satisfactory justification of tests selected; tests not all appropriate.
Test results (20%)	Excellent presentation of all data with appropriate visualisations of findings; correct calculations of results.	Very good presentation of data with appropriate visualisations of findings; correct calculations of results overall.	Good presentation of data with mostly appropriate visualisations of findings; mostly correct calculations of results with minor inaccuracies.	Satisfactory presentation of data with some visualisations of findings; some calculations of results are noticeably inaccurate.
Interpretation (20%)	Interpretation is insightful with respect to the data and links back to the research question.	Interpretation is consistent with the data and links back to the research question.	Interpretation is partially consistent with the data and partially links back to the research question.	Interpretation is mostly inconsistent with the data and only shows minor links with the research question.

Introduction:

Research Question

How accurately do participants' self-reported emotional states correspond to the data from Galvanic Skin Response (GSR) collection?

Independent Variable: Emotional Category of a 360 VR clip

Dependent Variable: GSR Data

The VREED paper supplies expansive self report and physiological data regarding participants' emotional states. As work often relies on self reported emotions, it is of interest to explore how closely these two methods of emotion data collection mirror each other. The paper's dataset presents a fitting opportunity to compare directly the attitudnal, subjective response of a participant to their behavioural, physiological response. This is because it supplies a large amount of physiological data through GSR and ECG time series data (measuring skin conductivity and heart activity, respectively), and alongside this it supplies the participant's immediate reporting of their emotional state through self-assessment manikins (SAM), visual analogue scales (VAS), and presence questionnaires (PQ).

For this research question, GSR is to be explored. GSR Is used as a proxy for sympathetic arousal, which would be expected to increase during intense, or highly activated, emotional states. Therefore, it is expected to see higher GSR readings correspond to higher self-reported arousal on the SAM or VAS scales. Exploring the validity of GSR as a reflection of sympathic arousal is powerfully important, and an unexpected result may lead to questionning the validity of studies that build upon this assumed proxy.

Data Overview:

The data provided is expansive, and for the research question exploring the relationship between GSR data and emotional state, it can be simplified as follows.

Two datasets were used. The first was the physilogical data. The GSR/ECG data comes in .dat format, with one file per participant. Within this, each file holds a data dictionary of {Labels:Data}. The Labels are split into 1 baseline, and then the 4 emotional quadrants based on the Circumplex of Emotion (CoE) (1). They are:

- 0 High Arousal, High Valence
- 1 Low Arousal, High Valence
- 2 Low Arousal, Low Valence

- 3 High Arousal, Low Valence
- 4 Baseline

The 'Data' part of the {Labels:Data} dictionary has 3 dimensions: video index, channel, and data length.

For this particular question, this dictionary can be transformed to {Labels: {i, 1, d} with *i* being iterated through, and *d* being collected into a dataframe. The dataframe will hold the participant id, the label associated with the video that elicited this particular GSR response, and the GSR response itself.

The table below shows an example of this.

	Clip	Label	GSR_Feature1	GSR_Feature2
0	0	4	5.059974	0.042331
1	1	1	4.447941	-0.067646
2	2	2	6.304088	-0.274103
3	3	0	4.784788	0.179261
4	4	1	4.493467	0.099771
5	5	0	4.402793	-0.006234
6	6	3	4.394402	0.368861
7	7	1	5.091137	0.013314
8	8	2	4.311366	-0.057073
9	9	2	4.465663	-0.000018
10	10	0	4.734919	0.019231
11	11	3	4.392134	0.017839

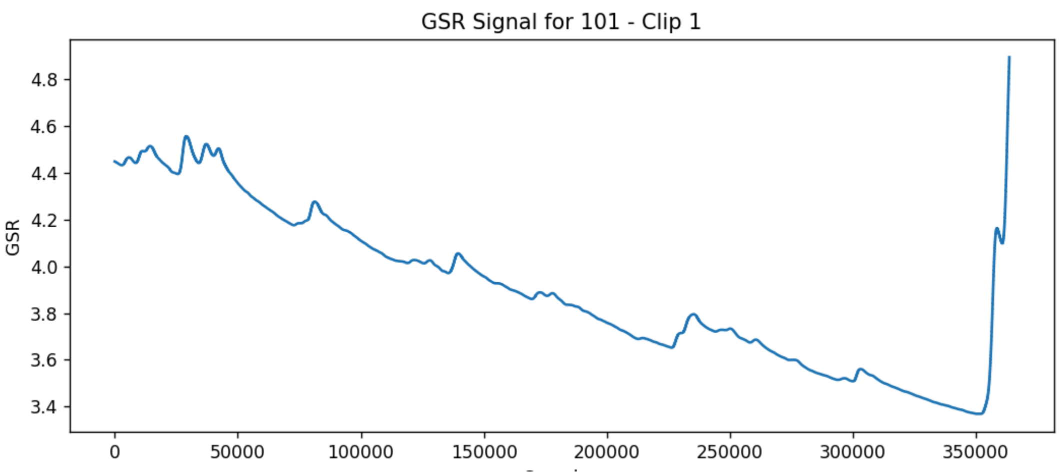

Figure 1: The GSR signal output for participant 101, clip 1.

The second dataset contained self-assessment data following each clip from the participants. The participants used SAM and PQ ratings to score their emotional state after the video, and for this research question the POST-Arousal and POST-Valence scores were used.

The GSR .dat file does not store any unique identifier for the video clips. The first dimension of the Data array in the Data Dictionary is only an index, not a unique identifier, confirmed by comparing the quadrant variable across between the GSR rows and the Self reported rows. This means that the data cannot be merged on individual clips, and instead be merged and analysed as combined quadrants. The ability to assess the physiological GSR data's correlation to the self reported arousal and valence will not be affected by this, however it removes the opportunity to further break any findings into sub questions that would explore differences between participants, or nuances in how different groups aligned the two measures.

The fully processed file held the variables shown below in figure 2.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Participant	Numeric	3	0		None	None	8	Right	Nominal	Input
2	Quad_Cat	Numeric	1	0		None	None	8	Right	Ordinal	Input
3	POST_Vale...	Numeric	18	16		None	None	20	Right	Scale	Input
4	POST_Arou...	Numeric	18	16		None	None	20	Right	Scale	Input
5	GSRMean	Scientific	23	16		None	None	25	Right	Scale	Input
6	GSRSD	Numeric	20	18		None	None	22	Right	Scale	Input
7	GSRChange	Scientific	23	16		None	None	25	Right	Scale	Input

Figure 2

Methodology:

Data Handling

The VREED Dataset has 33 participants (IDs 101-133). Each participant viewed a series of VR 360 video clips while their phsyiological data was recorded, and after each clip were asked to report their emotional state. For this analysis, one participant was excluded due to incomplete GSR data.

Initially, each phsyiological file (*.dat* file) was loaded using Python's pickle library. The GSR signal (stored as channel 0) was extracted as an array and stored in a dataframe alongwith the participant ID and clip 'label' (the quadrant from the Circumplex model of affect (1)). This produced a large raw data file (*GSR_all.csv**), which was the summarised into GSR Mean, GSR SD, and GSR range (calculated by maximum value minus minimu value), per participant per quadrant. It would have been more suited to summarise this data per participant per video clip, but as aforementioned, the dataset did not provide the correct IDs to perform that merge. The python script used to perform these merges is attached in the appendix under *aggregateGSRandPOSTbyQuadrant.py*. This merge also drops the 'baseline' data due to it also not being included in the POST data file.

The data was merged with the 'Post Exposure Ratings' data file to add the 'POST_Aroual' and 'POST_valence' variables to the rows. This was resulted in the final dataframe that would be processed in SPSS - found in the appendix under '*GSR_POST_Final.csv*'.

**GSR_all.csv is not included in the appendix due to it exceeding upload limits. Please get in contact to request access to this file.*

Foundational Analysis

Descriptive statistics, including means, minimums, maximims, and SDs of the GSR measures and POST emotional state data were calculated overall, and for each emotional quadrant in order to give a broad understanding of the data. Awareness of these overview values sets an appropriate foundation before moving on to the more complicated upcoming tests, improving the ability to understand results at a glance.

Then, normality tests were run to assess the distribution of the data. This is crucial as it will determine which variant of the further tests (in correlations, ANOVAs) will be appropriate. Due to the size of the data (33 participants, 1 exclusion), the Shapiro-Wiltz was chosen as it is known to perform better when sample sizes are smaller than N=50 (2).

Analysis

Pearson's correlations were ran between all the GSR data and both the POST Arousal, and POST Valence variables. The correlations of most relevance to the research question was the GSR Range (maximum minus minimum value) compared to the POST Arousal, and then compared to the POST Valence. This test was appropriate due to these variables all being of a continuous data type, and that correspondance is the key component of the research question. Pairwise correspondance between GSR Change and POST Arousal would reveal whether the *physiological* arousal changed with participants' self reported intensity of emotion. The POST Valence and GSR Change correspondance would reveal whether GSR responses were influenced by emotional positivty or negativity.

A one-way ANOVA was also run in order to test whether GSR reponses differed significantly between the CoE emotional quadrants. An ANOVA is suitable for this analysis because the IV, the quadrants (Quad_Cat variable), was categorial of four groups and the dependent variable, GSR Change, was continuous. A significant result would reveal that at least one quadrant results in a significantly different level of skin conductivity. Before performing a post-hoc of the ANOVA results, Leven's test was assessed and Turkey HSD used to identify which specific categories differed or not, as equal variance could now be assumed.

Results:

The following section shows the results from the statistical tests run on SPSS. Please note, the files containing these results is stored in the appendix under *Results.spv*.

Foundational Analyses Tests

Descriptive Statistics:

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
POST_Valence	129	1.0000000000	8.3333333333	4.8604651163	1.9846186637
POST_Arousal	129	1.0000000000	9.0000000000	3.9095607235	2.0377308739
GSRMean	131	-1.1996461E-5	2.7995924E+1	9.8384375E+0	6.0012894E+0
GSRSD	131	.03889315525	2.4030543271	.59922508274	.39817527269
GSRChange	131	-6.379296E+0	7.3558918E+0	8.2094177E-1	2.3619173E+0
Valid N (listwise)	129				

Figure 3.

Figure 3 shows the descriptive statistics results performed over the physiological data and self assessment data

The descriptive statistics for each quadrant are included in the '*Results.spv*' file attached in the appendix.

Normality Tests:

***Note**: The Quad_Cat numbers correspond to:*

- *0 High Arousal, High Valence*
- *1 Low Arousal, High Valence*
- *2 Low Arousal, Low Valence*
- *3 High Arousal, Low Valence*
- *4 Baseline*

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Quad_Cat	Statistic	df	Sig.	Statistic	df	Sig.
POST_Valence	0	.183	32	.008	.933	32	.048
	1	.169	32	.021	.950	32	.147
	2	.160	33	.031	.937	33	.056
	3	.158	32	.041	.938	32	.067
POST_Arousal	0	.149	32	.068	.899	32	.006
	1	.156	32	.047	.929	32	.037
	2	.135	33	.134	.947	33	.112
	3	.134	32	.153	.956	32	.217
GSRMean	0	.166	32	.025	.902	32	.007
	1	.133	32	.158	.952	32	.165
	2	.177	33	.010	.900	33	.005
	3	.117	32	.200*	.912	32	.012
GSRSD	0	.102	32	.200*	.950	32	.140
	1	.134	32	.154	.914	32	.014
	2	.109	33	.200*	.926	33	.026
	3	.114	32	.200*	.901	32	.007
GSRChange	0	.060	32	.200*	.993	32	.998
	1	.205	32	.001	.897	32	.005
	2	.106	33	.200*	.984	33	.902
	3	.112	32	.200*	.952	32	.161

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Figure 4.

Figure 4 shows the results from the Normality test, since the sample size is below 50 (N=33), then the Shapiro-Wilk test is best suited.

Correlation Tests:

Correlations

		GSRChange	POST_Arousal
GSRChange	Pearson Correlation	1	.334**
	Sig. (2-tailed)		<.001
	N	131	129
POST_Arousal	Pearson Correlation	.334**	1
	Sig. (2-tailed)	<.001	
	N	129	129

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 5.

Figure 5 shows the results for the correlation between GSR Change and POST Arousal.

Correlations

		GSRChange	POST_Valence
GSRChange	Pearson Correlation	1	-.037
	Sig. (2-tailed)		.678
	N	131	129
POST_Valence	Pearson Correlation	-.037	1
	Sig. (2-tailed)	.678	
	N	129	129

Figure 6.

Figure 6 shows the results for the correlation between GSR Change and POST Valence.

Correlations

		GSRMean	POST_Arousal
GSRMean	Pearson Correlation	1	.160
	Sig. (2-tailed)		.069
	N	131	129
POST_Arousal	Pearson Correlation	.160	1
	Sig. (2-tailed)	.069	
	N	129	129

Figure 7

Figure 7 shows the results for the correlation between GSR Mean and POST Arousal.

Correlations

		GSRMean	POST_Valence
GSRMean	Pearson Correlation	1	-.018
	Sig. (2-tailed)		.844
	N	131	129
POST_Valence	Pearson Correlation	-.018	1
	Sig. (2-tailed)	.844	
	N	129	129

Figure 8

Figure 8 shows the results for the correlation between GSR Mean and POST Valence.

ANOVA:

ANOVA

GSRChange

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	109.638	3	36.546	7.540	<.001
Within Groups	615.587	127	4.847		
Total	725.225	130			

Figure 9.

Figure 9 above shows the results of the ANOVA, while figure 10 shows the results from the homogeneity of variances test.

Tests of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
GSRChange	Based on Mean	.105	3	127	.957
	Based on Median	.112	3	127	.953
	Based on Median and with adjusted df	.112	3	122.905	.953
	Based on trimmed mean	.102	3	127	.959

Figure 10.

The full SPSS output for the ANOVA is attached in the appendix under the filename '*ANOVA.spv*'.

The Turkey ad-hoc of the ANOVA is shown below in Figure 10.

Multiple Comparisons

Dependent Variable: GSRChange

Tukey HSD

(I) Quad_Cat	(J) Quad_Cat	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
0	1	7.5371593E-1	5.4200202E-1	.508	-6.5731778E-1	2.1647497E+0
	2	6.7270118E-1	5.4200202E-1	.602	-7.3833254E-1	2.0837349E+0
	3	-1.541063E+0*	5.4622000E-1	.028	-2.963078E+0	-1.1904826E-1
1	0	-7.5371593E-1	5.4200202E-1	.508	-2.164750E+0	6.5731778E-1
	2	-8.1014755E-2	5.4200202E-1	.999	-1.492048E+0	1.3300190E+0
	3	-2.294779E+0*	5.4622000E-1	<.001	-3.716794E+0	-8.7276420E-1
2	0	-6.7270118E-1	5.4200202E-1	.602	-2.083735E+0	7.3833254E-1
	1	8.1014755E-2	5.4200202E-1	.999	-1.330019E+0	1.4920485E+0
	3	-2.213764E+0*	5.4622000E-1	<.001	-3.635779E+0	-7.9174944E-1
3	0	1.541063E+0*	5.4622000E-1	.028	1.1904826E-1	2.9630776E+0
	1	2.294779E+0*	5.4622000E-1	<.001	8.7276420E-1	3.7167936E+0
	2	2.213764E+0*	5.4622000E-1	<.001	7.9174944E-1	3.6357788E+0

*. The mean difference is significant at the 0.05 level.

Figure 10

***Note**: The Quad_Cat numbers correspond to:*

- *0 High Arousal, High Valence*
- *1 Low Arousal, High Valence*
- *2 Low Arousal, Low Valence*
- *3 High Arousal, Low Valence*
- *4 Baseline*

Discussion

- Interpretation is insightful with respect to the data and **links back to the research question**

Finding a reliable proxy for emotional state is imperitative to so much psychological and sociological research. A fullproof method of collecting data of a participant's emotional state allows for studies that explore the elicitation of emotions, the effects of emotion on the body, the power of emotion in decision making, and endless more, to faithfully present findings based on concrete evidence with reduced room for error. A mismatch between Galvanic Skin Response (GSR) and participant's self-reported data could potentially call into question both methods of emotional data collection. Perhaps GSR sensitivity vary greatly from individual to individual, and self assessment may require introspection that again, individuals have differing levels of.

The results in the previous section indicated a significant relationship between GSR change, and self-reported arousal. This suggests that GSR accurately reflects the intensity of their emotional activation. This aligns with what is currently understood about GSR, in that it corresponds to the arousal of the body's sympathetic nervous system. To contrast, there was not a significant relationship between GSR change, and self-reported valence. This would then imply that GSR does not distinguish between positive and negative emotion. Similar correlations were performed to explore the relationship between GSR mean instead of GSR range, and no significant correlations were found. This is to be expected due to mean being a poor signifier of the elicited skin response due to the 360 video clip. A participant may, for example, be relaxed at the start of a clip and as it continues become more and more stressed or activated. For this reason, GSR range conveys a lot more about how the participant responded to the clip's peaks than mean does.

The ANOVA revealed that significant differences exist in GSR response across the quadrants. The ad-hoc then revealed that the significant differences all occured in quadrant 3 (High Arousal, Low Valence), as the change in GSR was far greater than in the other quadrants.

Comparison	Significance
0-1	.508
0-2	.602
0-3	**.028***
1-0	.508
1-2	.999
1-3	**<.001***
2-0	.602
2-1	.999
2-3	**<.001***

Significant comparisons marked with asterix.

The emotional states associated with this quadrant would be things like anger, fear, or anxiety. These negative yet highly activated emotions appear to, in this dataset, drive the most pronounced changes in skin sweat. As said earlier, the known association between Galvanic Skin Response (GSR) to the sympathetic nervous system aligns with this finding. The sympathetic nervous system is a powerful system in the body designed to kickstart particularly rapid and intense response to threat. Therefore, a large response from this system when put under a negative, highly activating, 360 video clip is in accordance with the current understanding of how the sympathetic nervous system operates and how intensley it reacts to differing situations.

While valence did not correlate with GSR, the ANOVA result shows that it does still influence activation enough to significantly differ the *highly activated, low valence* quadrant to the other quadrants. The negative affect intensifies the physiological reaction to a highly activating clip.

To conclude, the exploration of the VREED dataset found that GSR does indeed correlate to a participant's self assessment of their emotional state, but only the arousal axis of the emotion, not the valence. However, the effect of a highly arousing clip is increased when said clip is negative valence.

References:

1: Russell, James. (1980). *A circumplex model of affect*. Journal of Personality and Social Psychology. 39. 1161-1178. 10.1037/h0077714.

2: *SPSS Shapiro-Wilk Test – quick tutorial with example* Available at: https://www.spss-tutorials.com/spss-shapiro-wilk-test-for-normality