

FACTORS INFLUENCING EDA DURING NARRATIVE EXPERIENCES

BY SAM SPAULDING & JULIE LEGAULT

FOR AFFECTIVE COMPUTING, FALL 2013

T

Motivation and Questions

THE STUDY OF AFFECT THROUGH PHYSIOLOGICAL MEASUREMENT HAS BEEN USED BROADLY TO MEASURE AROUSAL AND ENGAGEMENT IN MANY FORMS OF ENTERTAINMENT. THIS PARTICULAR STUDY SOUGHT TO MEASURE AND ANALYZE THE ROLE THAT NARRATIVE VOICE AND LANGUAGE PLAY ON AFFECT IN STORYTELLING.

WE HOPE THAT BY USING ELECTRODERMAL ACTIVITY SENSORS TO STUDY LITERATURE AND STORYTELLING, WE CAN COLLECT DATA TO PROVIDE INFORMATION ABOUT THE READER'S PHYSIOLOGICAL AROUSAL AND, COMBINED WITH A THOUGHTFUL RESEARCH PROTOCOL, IDENTIFY SUBCONSCIOUS COGNITIVE AND AFFECTIVE RESPONSES TO STORYTELLING.

WE SET OUT TO EXAMINE 3 THINGS

(1) CAN EDA BE A RELIABLE MEASURE OF NARRATIVE 'EXCITATION' THROUGHOUT A STORY?

(2) WHAT EFFECT DOES THE EXPRESSIVITY AND 'HUMAN-NESS' OF A VOICE DURING A STORY HAVE ON EDA?

(3) WHAT CAN ONE LEARN ABOUT A STORY FROM EDA MEASUREMENT?

Background and Prior Research

WE STARTED OFF BY THINKING ABOUT VARIOUS WAYS TO TRY TO ANALYZE THE EFFECT OF STORIES ON PEOPLE.

STORIES ARE KNOWN TO MANIPULATE AFFECT AND ARE SOUGHT AFTER FOR THAT EXACT REASON. THE ADVANCES IN TECHNOLOGY AND ENTERTAINMENT ALLOW AND ENCOURAGE US TO LOOK AT THE MANY WAYS IN WHICH STORYTELLING IS A SUCCESSFUL MEDIUM FOR INDUCING AFFECT AND EMPATHY.

WE HOPE THAT OUR EFFORTS TO MEASURE THE PHYSIOLOGICAL RESPONSES TO NARRATIVE VOICE AND LANGUAGE CAN ADVANCE UNDERSTANDING OF THE COMPLEX INTERACTION BETWEEN AFFECT, NARRATIVE AND SENSING TECHNOLOGIES

“EMOTION IS KEY TO THE EFFECTIVENESS OF NARRATIVES AND STORYTELLING. STORIES, EVEN IF FICTIONAL, HAVE THE ABILITY TO INDUCE A GENUINE EMOTIONAL RESPONSE . HOWEVER, THERE ARE NUANCES IN THE EMOTIONAL RESPONSE TO NARRATIVE REPRESENTATIONS COMPARED TO EVERYDAY SOCIAL DIALOGUE AND THEREFORE CONTEXT SPECIFIC MODELS NEED TO BE DESIGNED.”¹

Can fiction promote empathy and improve theory of mind?

WE WERE SPECIFICALLY INTRIGUED BY RECENT EXPERIMENTAL WORK SUGGESTING THAT EXPOSURE TO FICTION CAN MAKE READERS MORE EMPATHETIC AND IMPROVE THE THEORY OF MIND PROCESSES - THE ABILITY TO REPRESENT AND UNDERSTAND ANOTHER'S THOUGHTS AND FEELING.

A RECENT STUDY BY DAVID KIDD AND EMANUELE CASTANO AT THE NEW SCHOOL, NEW YORK, FOUND THAT THE "DEFAMILIARIZATION" THAT OFTEN COMES WITH FICTION READING CAN BE A POWERFUL MODIFIER OF AFFECT AND THAT EXPOSURE TO FICTION CAN LEAD TO GREATER SCORES ON TESTS OF EMOTIONAL INTELLIGENCE AND EMPATHY.

Computational Approaches to Understanding Fiction

ADVANCES IN NATURAL LANGUAGE PROCESSING (NLP) ARE ALSO ENABLING AND INSPIRING NEW METHODS FOR UNDERSTANDING AND ANALYZING FICTION. IN PARTICULAR, WE WERE INSPIRED BY THE WORK OF SAIF MOHAMMAD AT THE INSTITUTE FOR INFORMATION TECHNOLOGY IN OTTAWA.

ONE RECENT PROJECT DEMONSTRATES A SENTIMENT ANALYSIS CLASSIFIER THAT CAN BE USED TO PREDICT THE EMOTIONAL CONTENT OF FAIRYTALES AND LITERARY WORKS. WITH THIS TEXT ANALYZER, ONE CAN SIFT THROUGH DIGITAL TEXTS AND QUANTITATIVELY ESTIMATE THE EMOTIONAL CONTENT OF THE STORY BASED ON WORDS TAGGED AS EMBODYING ONE OR MORE OF THE 8 “BASIC” EMOTIONS (PLUTCHIK 1980). THIS STUDY GOT US INTERESTED IN THE PHYSIOLOGICAL RESPONSES THAT LANGUAGE COULD TRIGGER IN THE CONTEXT OF NARRATIVE MEANT TO STIMULATE IMAGINATION AND AFFECT.

Experimental Methods

QUESTIONS

(1) CAN EDA BE A RELIABLE MEASURE OF NARRATIVE 'EXCITATION' THROUGHOUT A STORY?

(2) WHAT EFFECT DOES THE EXPRESSIVITY AND 'HUMAN-NESS' OF A VOICE DURING A STORY HAVE ON EDA?

(3) WHAT CAN ONE LEARN ABOUT A STORY FROM EDA MEASUREMENT?

STORY 1 - COMPUTER, STORY 2 - UNEXPRESSIVE, STORY 3 - EXPRESSIVE

Experimental Design

WE DESIGNED AN EXPERIMENT TO MEASURE THE EDA FROM PARTICIPANTS READING A STORY UNDER 3 DIFFERENT CONDITIONS - WITH AN EXPRESSIVE HUMAN VOICE, AN UNEXPRESSIVE HUMAN VOICE, AND A COMPUTER GENERATED VOICE.

EACH PARTICIPANT HEARD 3 DIFFERENT STORIES TOLD IN A DIFFERENT 'STYLE' EACH TIME.

PARTICIPANTS IN GROUP A HEARD:

STORY 1 - **COMPUTER**, STORY 2 - **UNEXPRESSIVE**, STORY 3 - **EXPRESSIVE**

PARTICIPANTS IN GROUP B HEARD:

STORY 1 - **UNEXPRESSIVE**, STORY 2 - **EXPRESSIVE**, STORY 3 - **COMPUTER**

PARTICIPANTS IN GROUP C HEARD:

STORY 1 - **EXPRESSIVE**, STORY 2 - **COMPUTER**, STORY 3 - **UNEXPRESSIVE**

The Stories

IN ORDER TO GET MEASURABLE, INTERESTING REACTIONS FROM THE PARTICIPANTS IN THE TIME ALLOCATED FOR OUR STUDY, WE CHOSE THREE SHORT, SCARY STORIES FROM *MORE SCARY STORIES TO TELL IN THE DARK* BY ALVIN SCHWARTZ

STORY 1 : **“SOMETHING WAS WRONG”**

STORY 2 : **“THE CAT’S PAW”**

STORY 3 : **“OH, SUSANNAH”**

THE PRIMARY MOTIVATION BEHIND CHOOSING THESE STORY WAS NOT ONLY THEIR LENGTH (EACH APPROX. 75 SECONDS LONG IN RECORDING), BUT ALSO THEIR STRONG NARRATIVE CLIMAX AND THE LIKELIHOOD OF THEM PRODUCING MEASURABLE AROUSAL THROUGH. THESE STORIES WERE ALSO UNLIKELY TO BE KNOWN TO THE PARTICIPANT PRIOR TO THE STUDY.

THESE 3 STORIES RELY ON SURPRISE AS AN AFFECTIVE MODIFIER, AND THEREFORE ALSO SATISFY THE PRINCIPLE OF **‘DEFAMILIARIZATION’** USED IN THE PROMOTION OF THEORY OF MIND PROCESSES IN LITERARY FICTION, AS PER THE STUDY BY DAVID KIDD AND EMANUELE CASTANO.

EACH STORY WAS RECORDED IN 3 DIFFERENT NARRATIVE VOICES, AND OVERLAID ONTO A VIDEO OF THE SCROLLING TEXT. EACH PARTICIPANT LISTENED TO ONE VERSION OF EACH STORY WHILE THEIR EDA WAS BEING MONITORED THROUGH THE SENSORS ATTACHED TO THEIR NON-DOMINANT HAND.

Monitoring Equipment

WE USED THE ACQKNOWLEDGE SENSOR SYSTEM MANUFACTURED BY BIOPAC BECAUSE OF ITS REPUTATION FOR HIGH-FIDELITY READINGS. THE SYSTEM USES A WRIST-BASED MONITOR THAT CONNECTS 2 ELECTRODES TO BE PLACES ON THE PARTICIPANTS FINGERPADS AND WIRELESSLY TRANSMITS THE DATA TO SOFTWARE.

THE SOFTWARE SUPPORTS VIDEO RECORDING AND SYNCED PLAYBACK, WHICH WE ULTIMATELY DECIDED NOT TO INCORPORATE THAT INTO OUR ANALYSIS BECAUSE OF THE RELUCTANCE OF PARTICIPANTS TO BEING FILMED AND THE OUR LEVELS OF COMFORT WITH THE SENSOR SYSTEM.

WE MEASURED EDA FROM EACH PARTICIPANT'S INDEX AND MIDDLE FINGERTIPS ON THEIR NON-DOMINANT HAND. THIS WAS RECORDED WIRELESSLY IN THE ACQKNOWLEDGE SOFTWARE ON A BOOTCAMP VERSION OF WINDOWS ON A MACBOOK PRO. THE DATA WAS TAGGED WITH THE BEGINNING AND ENDING OF EACH ACTIVITY AND THEN SAVED AND REVIEWED FOR ANALYSIS. THIS ANALYSIS ALSO INCLUDED THE DATA COLLECTED FROM THE FOLLOW-UP SURVEYS AND EMOTIONAL TEXT ANALYSIS BY THE PARTICIPANTS.

Study Design

Phase A) Listening and measuring

- 1) THE PARTICIPANT COMES IN, SIGNS THE RELEVANT DOCUMENTS.
- 2) THE SENSORS ARE THEN INSTALLED ON THEIR NON-DOMINANT HAND AND, WHEN THEY ARE INSTALLED COMFORTABLY AND THE SENSORS HAVE BEEN TESTED FOR RESPONSE (BY MAKING THE PARTICIPANT TAKE A DEEP BREATH, TALK OR LAUGH), THE PARTICIPANTS PUTS ON THE HEADPHONE AND STARTS THE VIDEO WHICH IS ALREADY IN FULLSCREEN MODE.
- 3) 5 MINUTE MEDITATION EXERCISE (IN ORDER TO ATTEMPT TO CONTROL FOR PERSONAL, EXTERNAL VARIATIONS IN EDA (STRESSFUL DAY, PREOCCUPIED THOUGHTS, ETC.)
- 4) STORIES START: PARTICIPANTS HEAR 3 STORIES, APPROXIMATELY 75 SECONDS EACH, EACH READ IN A DIFFERENT VOICE DEPENDING ON EXPERIMENTAL CONDITION:

A : STORY 1 – COMPUTERIZED, STORY 2 – UNEXPRESSIVE HUMAN, STORY 3 – EXPRESSIVE HUMAN
B : STORY 1 – UNEXPRESSIVE HUMAN, STORY 2 – EXPRESSIVE HUMAN, STORY 3 – COMPUTERIZED
C : STORY 1 – EXPRESSIVE HUMAN, STORY 2 – COMPUTERIZED, STORY 3 – UNEXPRESSIVE HUMAN
- 5) AFTER THE STORIES ARE FINISHED, THE SENSORS ARE REMOVED FROM THE PARTICIPANT.

got there or where he had been earlier. He did not even know what time it was.

He saw a woman walking towards him and stopped

Study Design

Phase B) Self-Assessment

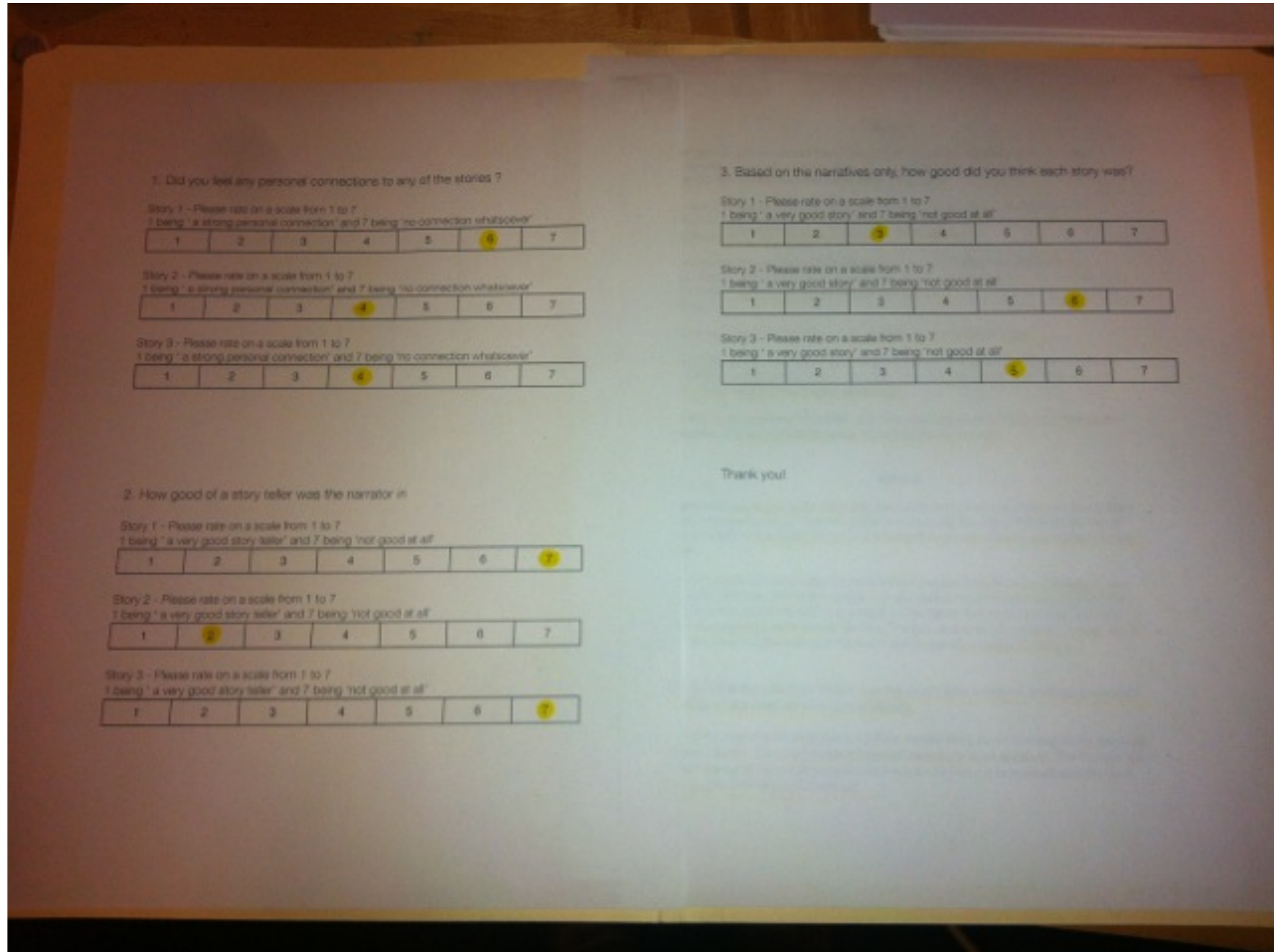
SENSORS ARE REMOVED FROM THE PARTICIPANT, AND PARTICIPANTS ARE ASKED TO FILL OUT TWO SURVEYS ABOUT THEIR EXPERIENCE WITH THE STORIES. THESE SURVEYS ALLOW FOR MAPPING OF THE PEAKS OF AROUSAL TO THEIR SELF-REPORTED DATA, AS WELL AS FOR DETECTION OF FALSE SENSING/ARTIFACTS/ PEAKS UNRELATED TO THE STORIES.

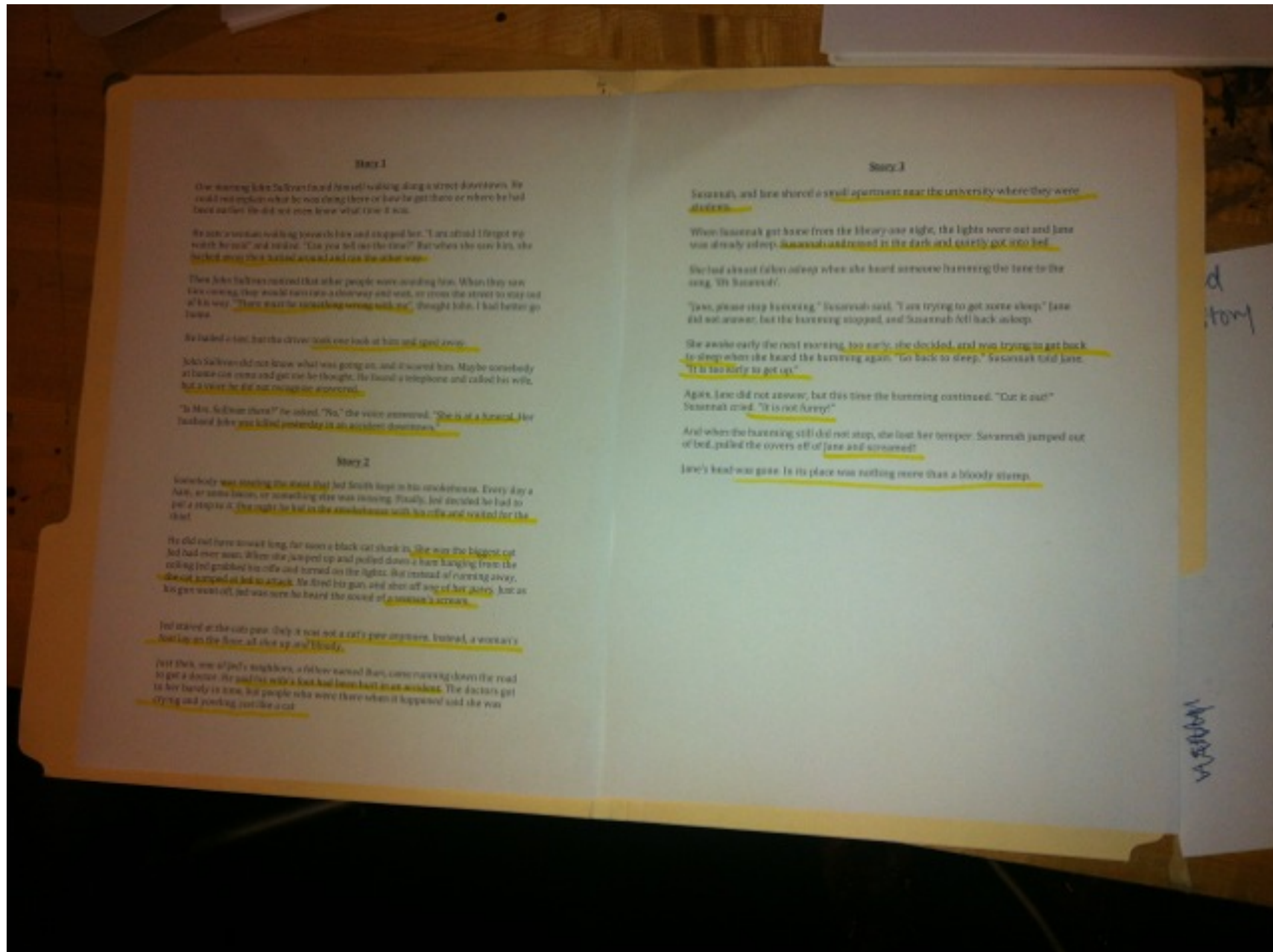
1) SELF-REPORTED AROUSAL SURVEY.

PARTICIPANTS ARE GIVEN TEXT OF EACH STORY AND ASKED TO HIGHLIGHT PARTS OF THE STORY THAT THEY REMEMBER TRIGGERING FEELINGS OF EXCITEMENT, AROUSAL, OR FEAR.

2) SELF-REPORTED NARRATIVE ASSESSMENT SURVEY.

PARTICIPANTS ARE ASKED TO RATE EACH STORY BASED ON *PURELY* NARRATIVE ELEMENTS AS WELL AS ON *PURELY* PRESENTATION (VOICE) QUALITIES.





Controls and Manipulations

WE DESIGNED A RESEARCH PROTOCOL THAT WOULD ALLOW US TO GATHER THE RELEVANT DATA IN A FAIRLY NEUTRAL ENVIRONMENT WHILST BEING ENJOYABLE FOR THE PARTICIPANT.

IN THIS CASE, THE CHOICE OF STORIES HELPED MAKE THE EXPERIENCE PLEASANT FOR THEM. SEVERAL PARTICIPANTS ALSO COMMENTED ON THE MEDITATION EXERCISE, NOTING THAT IT WAS A NICE BREAK

WE TRIED TO MINIMIZE THE OUTSIDE INFLUENCES ON THE DATA BY CREATING A SPACE WITH LITTLE TO NO DISTRACTION, PROVIDING HIGH QUALITY HEADPHONES AND A 5 MINUTE RELAXATION EXERCISE PRIOR TO STORY TIME.

WE ALSO MINIMIZED THE INTERACTION BETWEEN US AND THE PARTICIPANTS BY HAVING THE VIDEOS ALL PLAY FROM A SINGLE CONTINUOUS FILE AND LEAVING THE IMMEDIATE AREA WHERE THE PARTICIPANT WAS SITTING.

PARTICIPANTS DID NOT RECEIVE COMPENSATION FOR THEIR PARTICIPATION, BUT ANALYSIS OF THE SURVEY DATA DEMONSTRATED THAT PARTICIPANTS CLEARLY UNDERSTOOD THE INSTRUCTIONS AND GAVE REASONABLE ANSWERS.

N = 6, ALL PARTICIPANTS WERE MIT STUDENTS FLUENT IN ENGLISH. AGES RANGED FROM 21 TO 35

Data Analysis

How well do people's self-reported ratings of which parts of the story were arousing correspond to EDA measurements?

Here, we examined similarities between participants' self-reported moments of excitement and their measured EDA.

Participants tended to report more moments of excitement than were actually recorded on the sensors.

S1

Avg. # of **recorded** peaks = **4.0**

Avg. # of **reported** peaks = **6.0**

S2

Avg. # of **recorded** peaks = **4.5**

Avg. # of **reported** peaks = **7.167**

S3

Avg. # of **recorded** peaks = **3.667**

Avg. # of **reported** peaks = **5.833**

What does this mean?

2 possibilities:

1: Despite the instructions, participants did not reflect on their actual feelings, but instead on re-reading and highlighting emotionally charged words
I.e., Participants have better recall than precision when asked to self-reflect on affective tasks

2: The BioPac sensor, or perhaps EDA as a metric itself, is not accurate enough to capture all the emotional peaks
I.e., The sensors are not able to capture every arousal peak, and therefore have better precision than recall

Conclusion: It's probably a mix of the two. The EDA signal probably represents a subset of the actual emotional responses to the story. However, there were certain regularities in the peaks that DID show up and are likely indicative of response to narrative.

What effect does “expressivity” and “human-ness” of voice have on EDA data in a narrative setting?

EACH STORY WAS READ TO EVERY PARTICIPANT IN ORDER.

WHAT VARIED BY EXPERIMENTAL CONDITION WAS THE LEVEL OF EMOTION AND DEGREE OF HUMAN-NESS IN THE VOICE. PARTICIPANTS HEARD ALL THREE VOICES, BUT READING DIFFERENT STORIES AND IN DIFFERENT ORDERS.

PARTICIPANTS IN GROUP A HEARD:

STORY 1 - **COMPUTER**, STORY 2 - **UNEXPRESSIVE**, STORY 3 - **EXPRESSIVE**

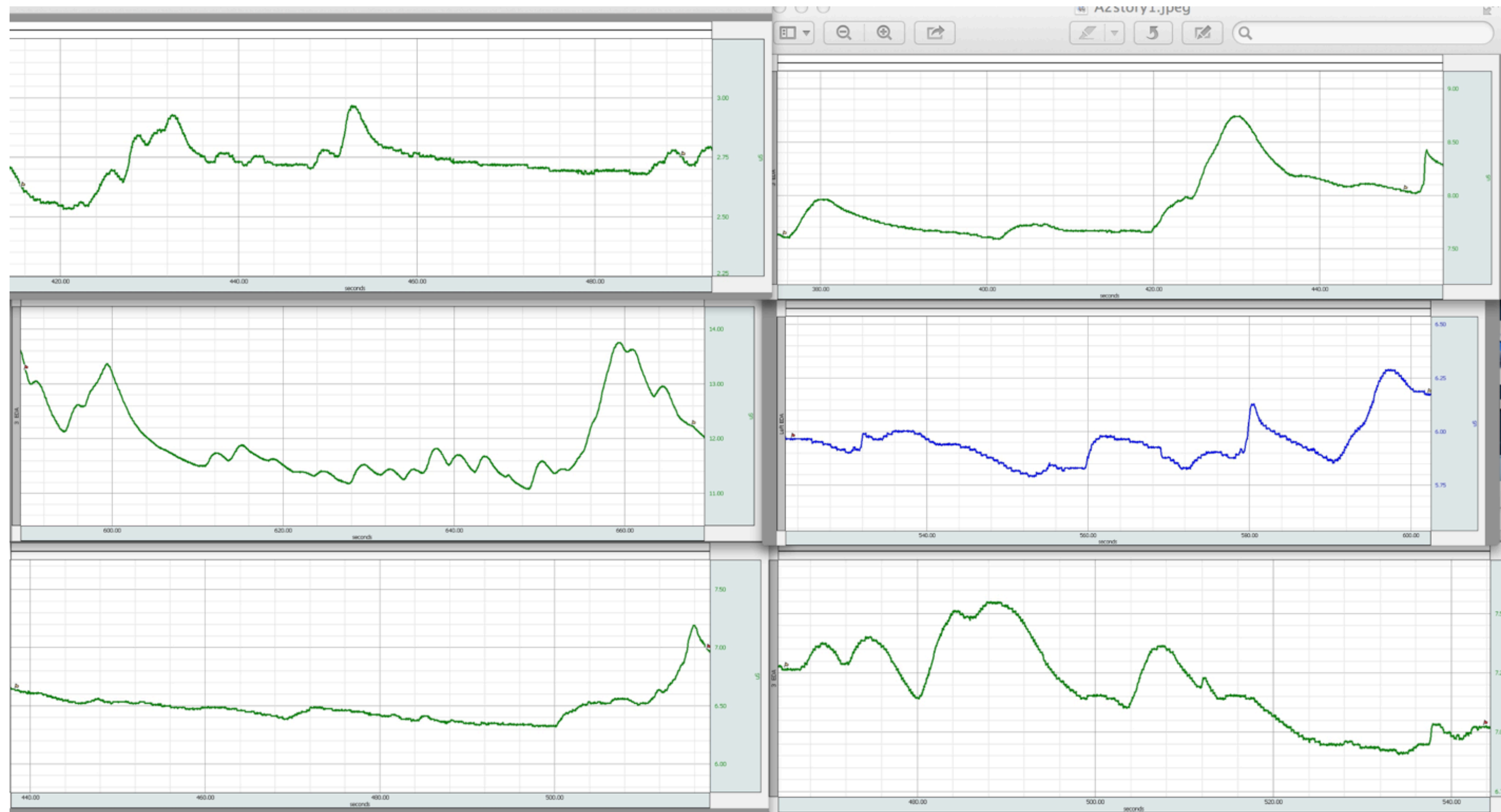
PARTICIPANTS IN GROUP B HEARD:

STORY 1 - **UNEXPRESSIVE**, STORY 2 - **EXPRESSIVE**, STORY 3 - **COMPUTER**

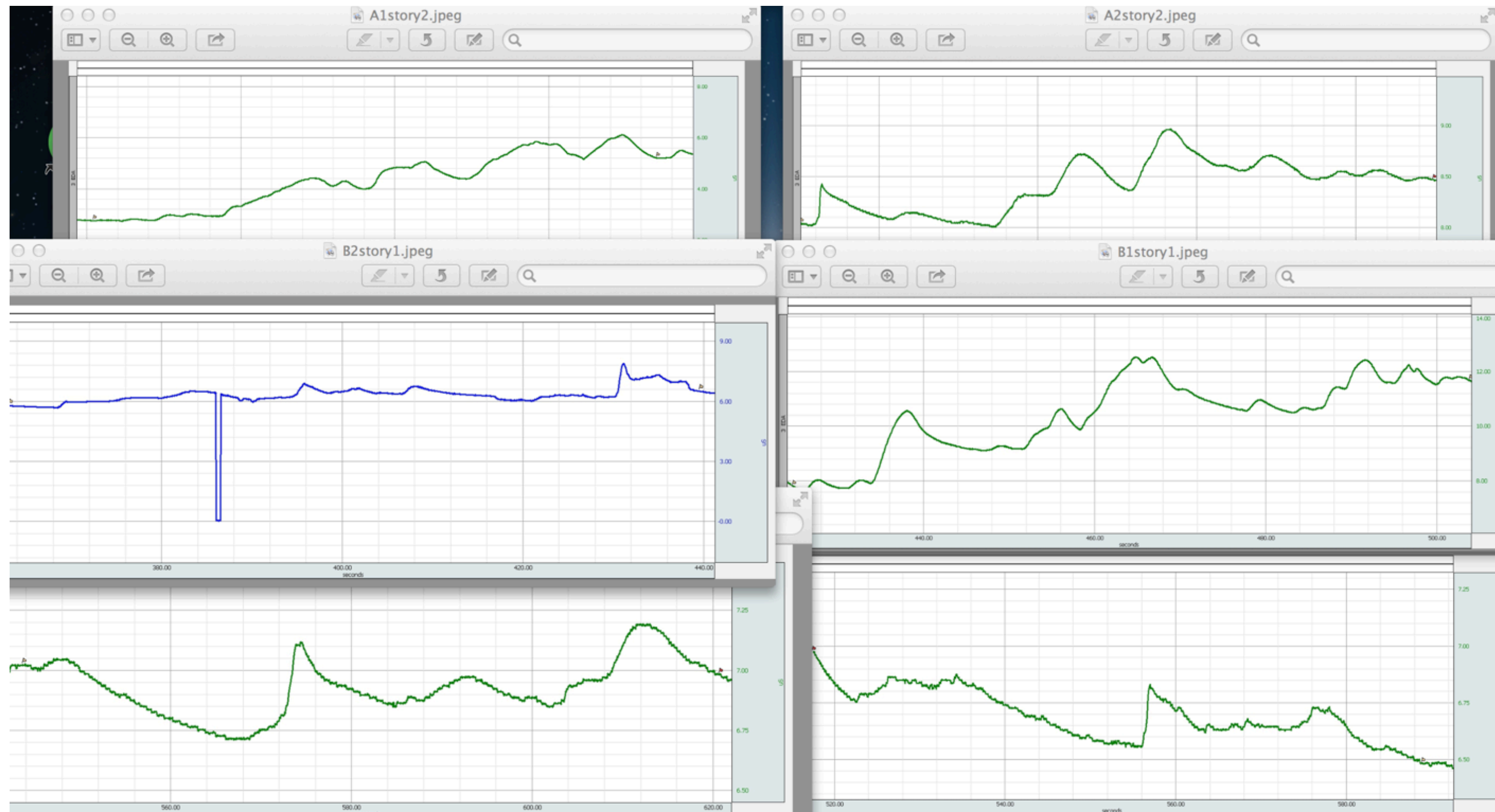
PARTICIPANTS IN GROUP C HEARD:

STORY 1 - **EXPRESSIVE**, STORY 2 - **COMPUTER**, STORY 3 - **UNEXPRESSIVE**

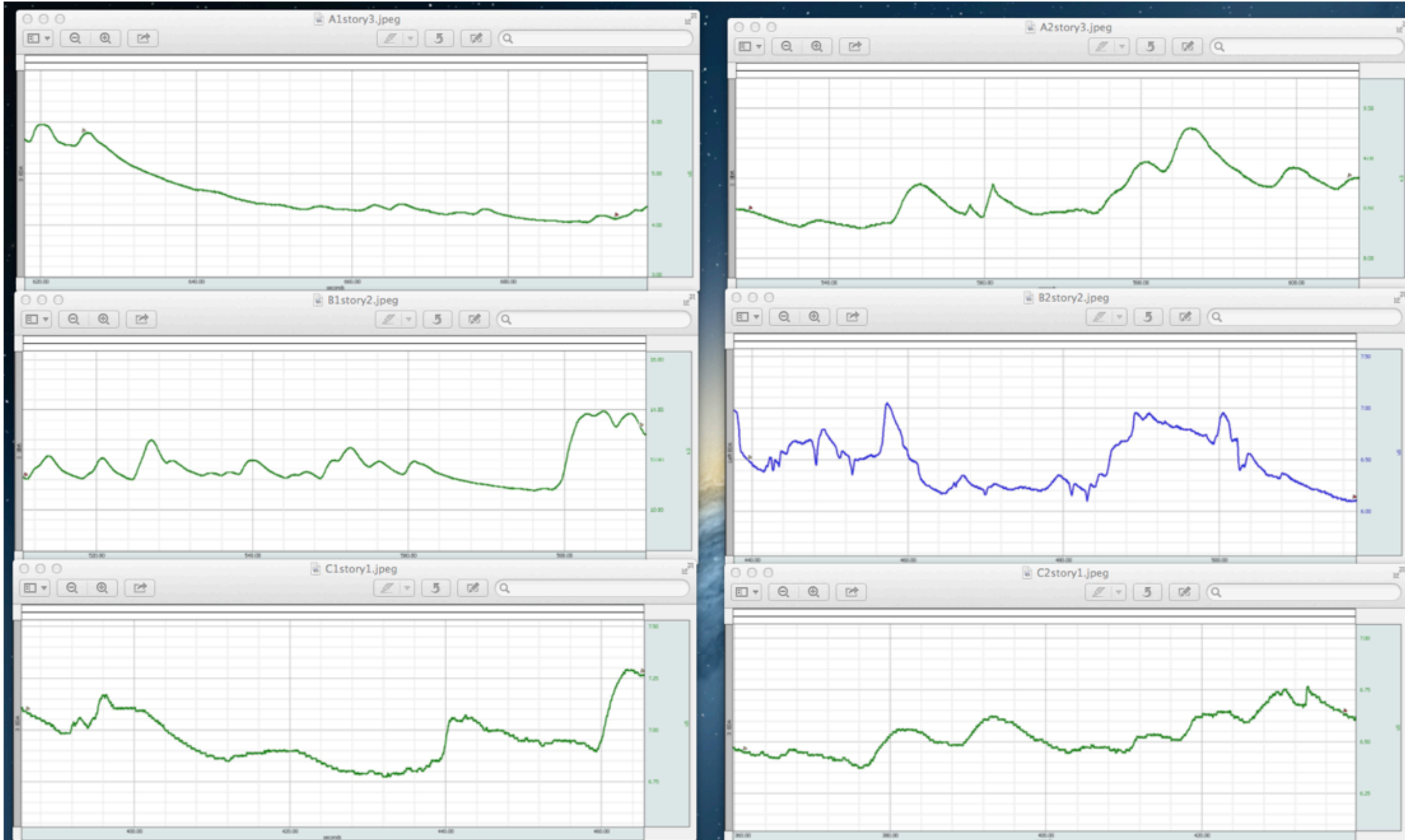
Computer-Generated Voice



Human-Unexpressive Voice



Human-Expressive Voice



What effect does “expressivity” and “human-ness” of voice have on EDA data in a narrative setting?

RESULTS: FEWER PEAKS IN UNEXPRESSIVE HUMAN CASE.

AVG. PEAKS PER STORY:

COMPUTER - **4.16667**

UNEXPRESSIVE HUMAN - **3.5**

EXPRESSIVE HUMAN - **4.333**

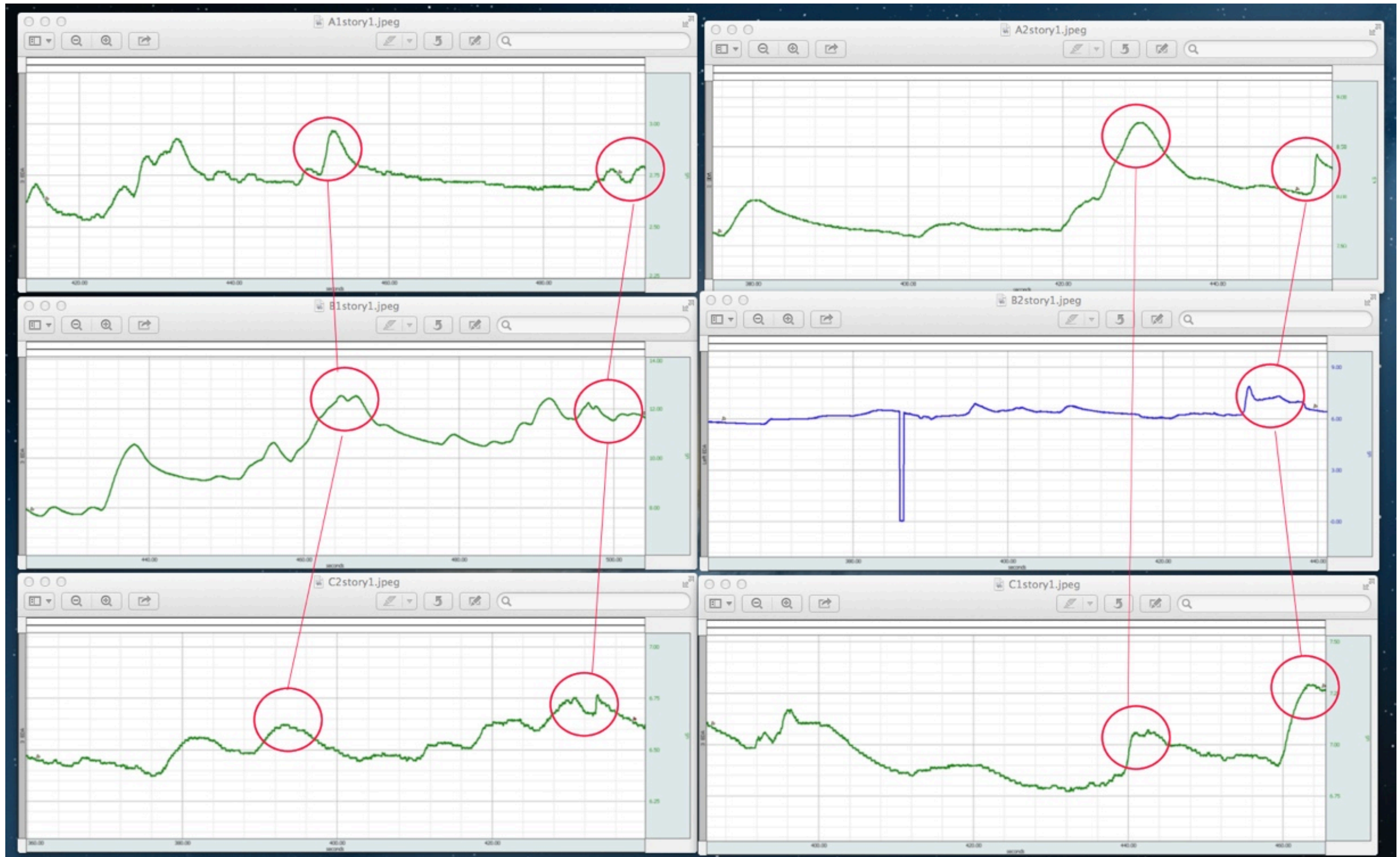
DISCUSSION: WE HYPOTHESIZE THAT PARTICIPANTS WERE DISTRACTED BY A HUMAN VOICE THAT LACKED ANY EXPRESSION. HUMAN EXPRESSIVE IS THE LEAST DISTRACTING, THEREFORE THE SIGNAL IS STRONGEST.

CURIOSLY, COMPUTER VOICE GENERATES NEARLY AS MANY PEAKS AS HUMAN EXPRESSIVE. THIS MAY BE DUE TO THE FACT THAT ALL PARTICIPANTS WERE STUDENTS MAJORING IN TECHNOLOGICAL FIELDS AND THEREFORE MAY HAVE BEEN MORE COMFORTABLE WITH COMPUTER GENERATED SPEECH

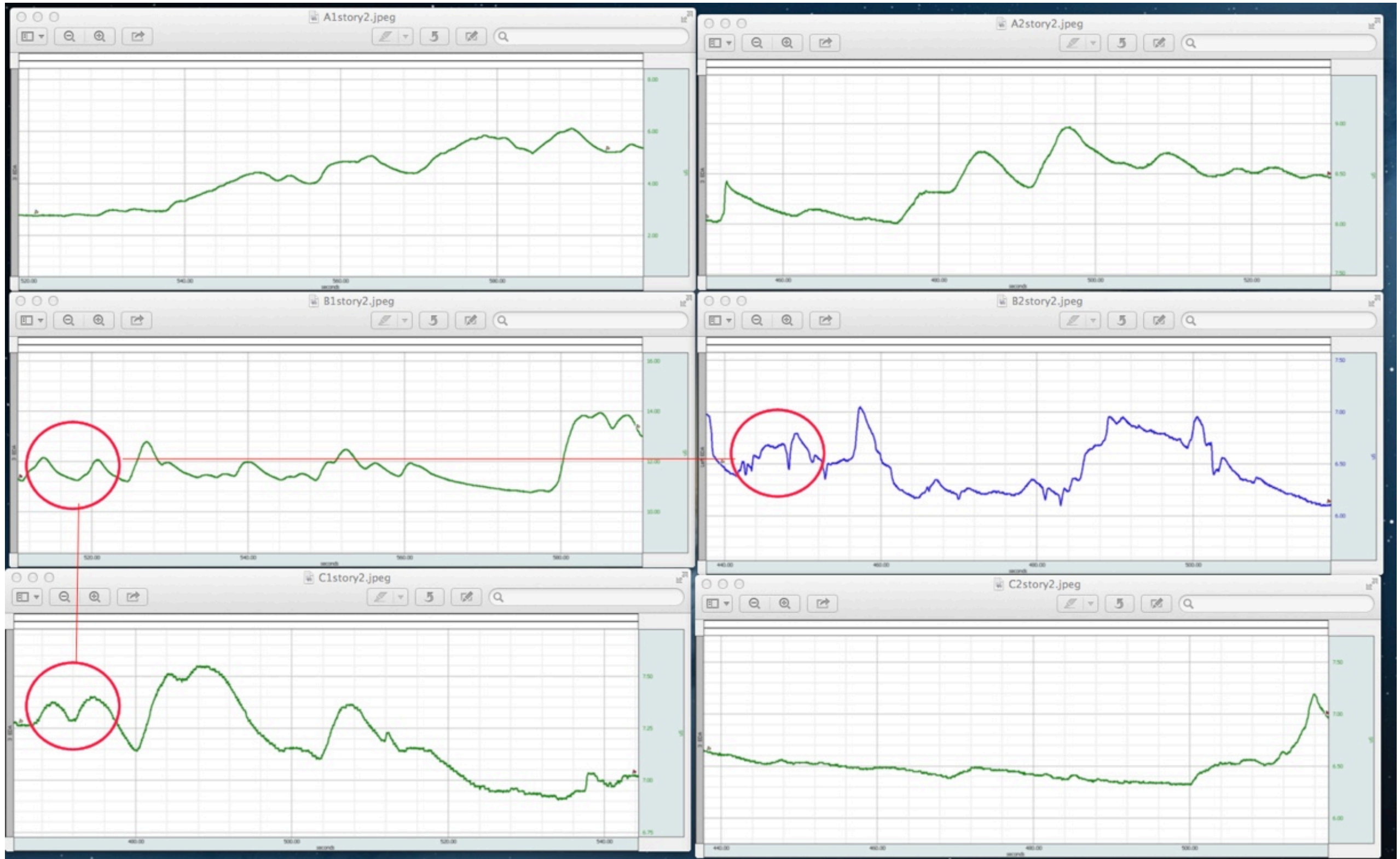
What can one learn about a story from EDA measurement?

IN ORDER TO ISOLATE THE RESPONSE TO TO THE NARRATIVE ASPECTS OF THE EXPERIENCE AS MUCH AS POSSIBLE, WE EXAMINED **SAME-STORY CROSS-CONDITION SIMILARITIES IN EDA RESPONSE**

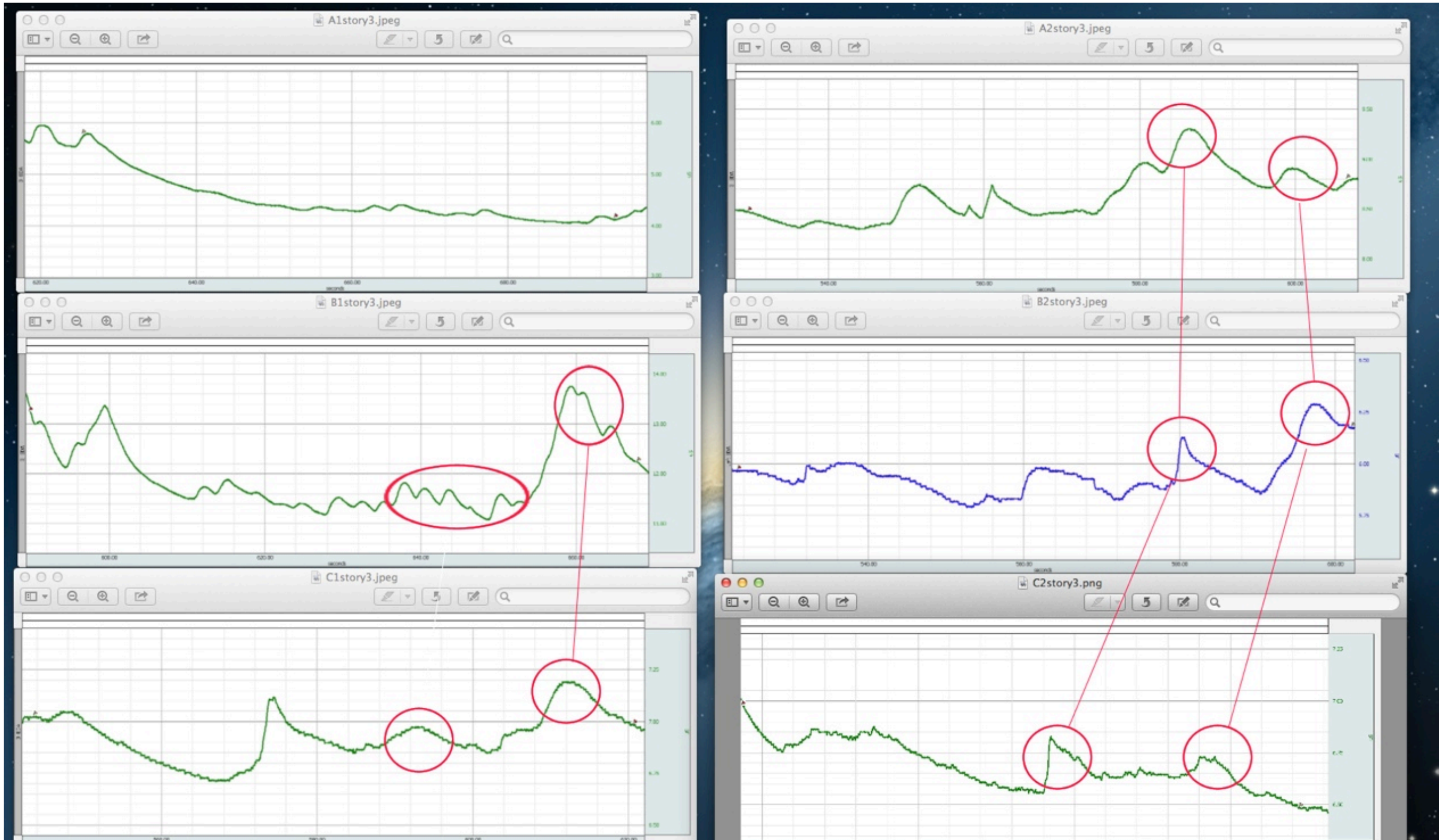
Story I



Story 2



Story 3



What can one learn about a story from EDA measurement?

RESULTS:

STRENGTH OF SAME-STORY, CROSS-CONDITION SIMILARITY IS RELATED TO **ASSESSMENT OF NARRATIVE STRENGTH**

“BASED ON NARRATIVE STRENGTH **ONLY**, HOW GOOD DID YOU THINK EACH STORY WAS?”

RATED FROM 1-7 (1 IS ‘A VERY GOOD STORY’ 7 IS ‘NOT GOOD AT ALL’)

STORY 1: **3.00**

STORY 2: **4.83**

STORY 3: **3.667**

Discussion

RESULTS:

IT IS LIKELY POSSIBLE TO EXTRACT A “NARRATIVE PROFILE” FROM EDA RESPONSE, BUT THE MEASURED STRENGTH OF THE EXCITATION MAY BE PROPORTIONAL TO THE PERCEIVED NARRATIVE VALUE TO THE AUDIENCE

IF ONE OF THE ROLES STORIES PLAY IN SOCIETY IS TO DELIBERATELY PROVOKE EMOTIONAL RESPONSES, IT MAKES SENSE THAT A STORY WITH A STRONGER AFFECTIVE RESPONSE WILL BE PERCEIVED AS “BETTER”

Conclusions

(1) CAN EDA BE A RELIABLE MEASURE OF NARRATIVE 'EXCITATION' THROUGHOUT A STORY?

YES, BUT HOW THE REFLECTIVE ASSESSMENT IS CONDUCTED CAN ACT AS A CONFOUND, ESP. WHEN THE SENSORS ARE NOISY.

(2) WHAT EFFECT DOES THE EXPRESSIVITY AND 'HUMAN-NESS' OF A VOICE DURING A STORY HAVE ON EDA?

WE HYPOTHESIZE THAT A **DISTRACTING VOICE CAN DAMPEN THE MEASURED RESPONSE** TO THE STORY

(3) WHAT CAN ONE LEARN ABOUT A STORY FROM EDA MEASUREMENT?

A STORY-SPECIFIC RESPONSE PROFILE CAN BE GENERATED, BUT THE **STRENGTH OF THE RESPONSE IS LIKELY INFLUENCED BY HOW "GOOD" THE STORY IS PERCEIVED TO BE**

Future Work

(1) BOOSTING THE RESPONSE SIGNAL

THE STRENGTH OF THE EMOTIONAL RESPONSE IS TIED TO THE NARRATIVE STRENGTH OF THE STORY. WHAT OTHER FACTORS COULD BOOST THE STRENGTH OF THE EMOTIONAL RESPONSE?

(2) FURTHER EXPLORATION OF VOICE EFFECTS

WE OBSERVED THAT A DISTRACTING VOICE DAMPENED THE EMOTIONAL RESPONSE. MIGHT IT ALSO CREATE AN ENTIRELY DIFFERENT SIGNAL?

(3) EVERYTHING DONE HERE SHOULD BE WAY MORE RIGOROUS. IN CASE YOU DIDN'T NOTICE, NOTHING WAS STATISTICALLY SIGNIFICANT! IT'S HARD WHEN YOU ONLY HAVE N=6!

References

- D. Comer Kidd, E. Casatano. (2013). Reading Literary Fiction Improves Theory of Mind, *Science*, 342 (6156), 377-380.
- Cress, Susan W.; Holm, Daniel T. (2000). Developing Empathy through Children's Literature, *Education*, 120, 593-596.
- K.F. Morrison. (1992). I Am You:: The Hermeneutics of Empathy in Western Literature, Theology and Art, *Speculum*, 67(4), 1016-1020.
- S.Mohammad. (2011). From Once Upon a Time to Happily Ever After: Tracking Emotions in Novels and Fairy Tales, *LaTech 11 Proceeding of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, 105-114.
- M. Bayer Mail, W. Sommer, A. Schacht. (2012). Font Size Matters—Emotion and Attention in Cortical Responses to Written Words, *Plos One*, 7 (5).
- Burgess, C., & Rosen, A. (1988). Priming of emotional words in the cerebral hemispheres. *Fifth Annual Linguistic Studies Conference on Language and Communication*, Syracuse, NY.
- A. Gerrads-Hess, K.Spies, F. W. Hesse. (1994) Experimental inductions of emotional states and their effectiveness. *British Journal of Psychology*, 85(1), 55-78.
- D. McDuff. (2012). Affective Storytelling: Automatic Measurement of Story Effectiveness from Emotional Responses Collected over the Internet. PhD Proposal in Media Arts & Sciences Affective Computing Group, MIT Media Lab.