



Finding your Niche in Research

Oishani and Shivani discuss their research!



Hi, I'm Oishani!

- Co-President of CSSA
- Senior
- Double Major in Cog Sci
(Machine Learning
Specialization) and
Linguistics
- ERC



My Research

- Started as ERC Sophomore Honors Project
- Independent project supervised by professor
- Continued over 3 years
- Currently writing Honors Thesis for Linguistics
Double Major

Testing Voice Assistants' Failure Rates with Different Dialects and Accents (Indian Accents in Particular)

Oishani Bandopadhyay

Professor Will Styler

How many of you have grown up in the US, speaking a mainstream dialect of English?

How many of you use voice assistants?

How many of you don't like your voice assistants?

Coincidence? (Almost) definitely not!

Research Question

What is the overall accuracy difference for speech to text systems across various dialects, especially speakers of Indian dialects of English compared to US dialect speakers?



Hypothesis

The failure rates of speech recognition systems vary according to dialect, with lower accuracy rates for various Indian dialects than Southern California English dialects.



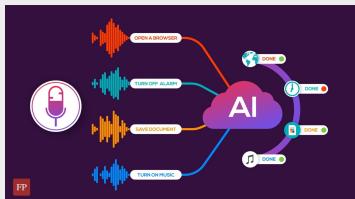
Background

- This exact question doesn't seem to have been addressed (which is...interesting!?)
- Research papers address
 - People with accents self-correcting with Speech to Text systems
 - Accuracies for non-native English speakers being lower, especially regionally

Current Research

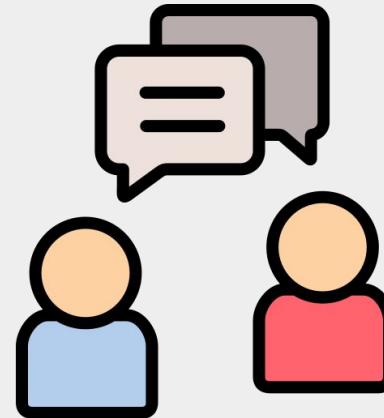


- Speech to Text systems are constantly being upgraded, and getting better at multilingual systems and more diverse language backgrounds
- This means that the Speech to Text systems I'm testing are often updated, but the experiment will continue to be relevant



How do you:

- Get speakers of various dialects of English to speak naturally and get these recordings (without wild violations of privacy as a large corporation)?
- Measure the difference in accuracy between transcriptions of mainstream US dialects vs Indian dialects?



Methods

To collect data, we survey people using FindingFive!

This lets participants record audio and saves the audio files!

The screenshot shows the FindingFive software interface. At the top, there's a navigation bar with links for 'Participate', 'Research', 'Get Help', and 'Hi Oishani'. Below the title 'Testing Voice Assistants' Failure Rates with Different Dialects and Accents (Indian Accents in Particular)' is a note that it's owned by Oishani Bandopadhyay of University of California San Diego. The main area has tabs for 'Trial Templates', 'Procedure', 'Auto-save on', 'Stimuli', and 'Responses'. The 'Stimuli' tab is active, showing a list of items with checkboxes: A_InstructionSlide1, A_InstructionSlide3, A_InstructionSlide2, A_BlockStartInstruction, A_DialPrompt1, A_DialPrompt10, A_DialPrompt11, A_DialPrompt12, A_DialPrompt13, A_DialPrompt2, DialPromptResp1, DialPromptResp10, DialPromptResp11, DialPromptResp12, DialPromptResp13, DialPromptResp2, DialPromptResp3, DialPromptResp4, and DialPromptResp5. The 'Responses' tab also lists these items with checkboxes.

What do we ask them to say?

Methods

4 Different Types of Prompts

‘Easy Prompts’:

- You haven't even been to the In-n-Out in the Outback Steakhouse neighbourhood?
- The complex houses married and single soldiers and their families

Methods

4 Different Types of Prompts

‘Hard Prompts’:

- She flew out to feel flu-flouting flounder few knew to flu flout.
- soldiers married and single their the houses complex families and

Methods

4 Different Types of Prompts

‘Dialect Specific Hard Prompts’:

- They abhor the abolition absconders that speak of aborting arbitrary overlooking.
- Wrap the wrap in aluminium foil wrap and tell that fellow where to go da.

Methods

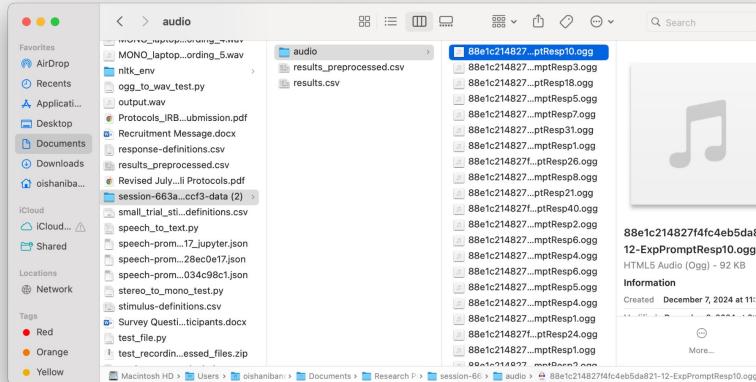
4 Different Types of Prompts

'Local Language Prompts':

- They knew where Rituparno would be on a Friday evening.
- She ate up the naan, confrontation awaited her

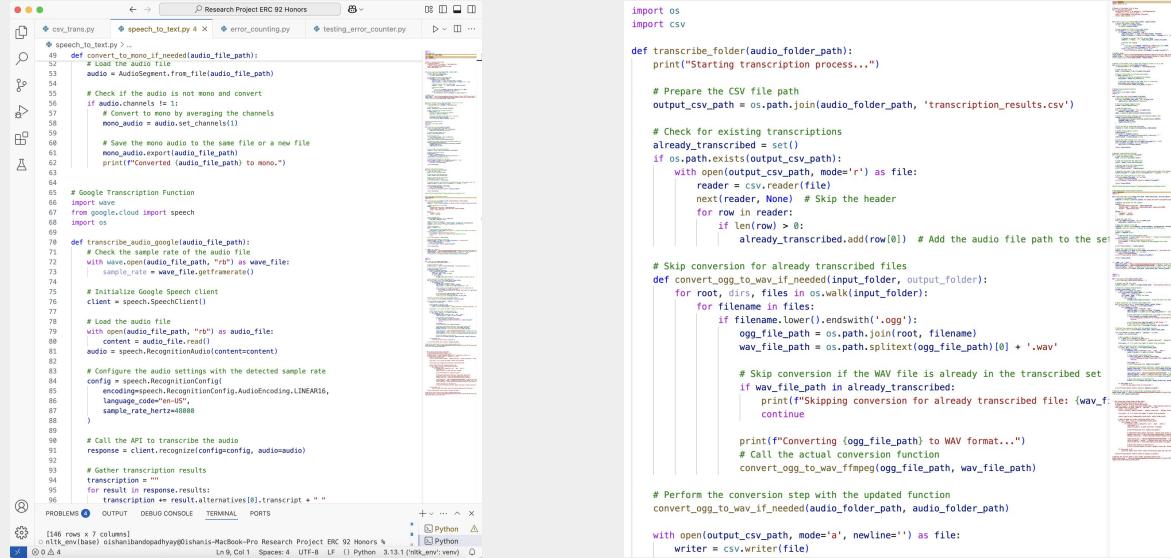
Methods

So, we get these recordings neatly packaged into folders with each anonymized speaker, and need to transcribe the audio files with each prompt's recording!



Methods

Next, we transcribe! This involves Speech to Text APIs from Microsoft, Google, and OpenAI...



```
import os
import csv

def transcribe_folder(audio_folder_path):
    print("Starting transcription process...")

    # Prepare the CSV file path
    output_csv_path = os.path.join(audio_folder_path, 'transcription_results.csv')

    # Check for existing transcriptions
    already_transcribed = set()
    if os.path.exists(output_csv_path):
        with open(output_csv_path, mode='r') as file:
            reader = csv.reader(file)
            next(reader, None) # Skip the header
            for row in reader:
                if len(row) > 0:
                    already_transcribed.add(row[0]) # Add the audio file path to the set

    # Skip conversion for already transcribed files
    def convert_ogg_to_wav_if_needed(input_folder, output_folder):
        for root, dirs, files in os.walk(input_folder):
            for filename in files:
                if filename.lower().endswith('.ogg'):
                    ogg_file_path = os.path.join(root, filename)
                    wav_file_path = os.path.splitext(ogg_file_path)[0] + '.wav'

                    # Skip conversion if the WAV file is already in the transcribed set
                    if wav_file_path in already_transcribed:
                        print(f"Skipping conversion for already transcribed file: {wav_file_path}")
                        continue

                    # Call the conversion function
                    convert_ogg_to_wav_ffmpeg(ogg_file_path, wav_file_path)

    # Perform the conversion step with the updated function
    convert_ogg_to_wav_if_needed(audio_folder_path, audio_folder_path)

    with open(output_csv_path, mode='a', newline='') as file:
        writer = csv.writer(file)
```

```
import wave
from google.cloud import speech
import os

# Google Transcription Function
def transcribe_audio_google(audio_file_path):
    # Check the sample rate of the audio file
    with wave.open(audio_file_path, "rb") as wave_file:
        sample_rate = wave_file.getframerate()

    # Initialize Google Speech client
    client = speech.SpeechClient()

    # Load the audio file
    with open(audio_file_path, "rb") as audio_file:
        content = audio_file.read()
    audio = speech.RecognitionAudio(content=content)

    # Configure the audio settings with the detected sample rate
    config = speech.RecognitionConfig(
        encoding=speech.RecognitionConfig.AudioEncoding.LINEAR16,
        language_code="en-US",
        sample_rate_hertz=sample_rate
    )

    # Call the API to transcribe the audio
    response = client.recognize(config=config, audio=audio)

    # Gather transcription results
    transcription = ""
    for result in response.results:
        transcription += result.alternatives[0].transcript + "
```

```
def convert_to_mono_if_needed(audio_file_path):
    # Load the audio file
    with wave.open(audio_file_path, "rb") as audio:
        audio = AudioSegment.from_file(audio_file_path)

    # Check if the audio is not mono and convert
    if audio.channels != 1:
        # Convert to mono by averaging the channels
        mono_audio = audio.set_channels(1)

    # Save the mono audio to the same file or a new file
    mono_audio.export(audio_file_path, format="mono")
    print(f"Converted {audio_file_path} to mono.")

# Load the audio file
with open(audio_file_path, "rb") as audio_file:
    content = audio_file.read()
audio = speech.RecognitionAudio(content=content)

# Configure the audio settings with the detected sample rate
config = speech.RecognitionConfig(
    encoding=speech.RecognitionConfig.AudioEncoding.LINEAR16,
    language_code="en-US",
    sample_rate_hertz=sample_rate
)

# Call the API to transcribe the audio
response = client.recognize(config=config, audio=audio)

# Gather transcription results
transcription = ""
for result in response.results:
    transcription += result.alternatives[0].transcript + "
```

Methods

This looks like a lot of code - but it's basically calling existing Speech-to-Text APIs and making them store transcriptions into tables and text files!

A	B	C	D	E
1 audio_filepath	google_transcript	whisper_transcript	azure_transcript	
2 /Users/oishaniba	please call Stella asked her to bring these things with her from the	Please call Stella, ask her to bring these things with her from the store, please call stella ask her to bring these things with her from the store six spoons of fr		
3 /Users/oishaniba	the beige Hue on the Waters of the Loch impressed all including tr	The beige hue on the waters of the lock impressed all, including the Fr the beige hue on the waters of the loch impressed all including the french queen befc		
4 /Users/oishaniba	not wind and the sun were disputing which was Stronger when a tr	North Wind and the Sun were disputing which was stronger when a tra north wind and the sun were disputing which was stronger when a traveller came alo		
5 /Users/oishaniba	he is leaving for maleshwaram tomorrow	He is leaving for Maleshwaram tomorrow.	he is leaving for maleswaram tomorrow	
6 /Users/oishaniba	he's leaving from Malaysia tomorrow	He is leaving from Malaysia tomorrow.	he is leaving from malleswaram tomorrow	
7 /Users/oishaniba	the beige you on the loft impressed all including the French Qt	The beige hue on the loft impressed all including the French Queen bet the beige hue on the loft impressed all including the french queen before she hea		
8 /Users/oishaniba	she gives it a 1 of 10 when often it deserves more	She gives it a 1 of 10 when often it deserves more.	she gives it a one of ten when often it deserves more	
9 /Users/oishaniba	he took a walk from the Chinese restaurant at the end of his walk	He took a walk from the Chinese restaurant at the end of his walk.	he took a walk from the chinese restaurant at the end of his walk	
10 /Users/oishaniba	you haven't even been to the In-N-Out in the Outback Steakhouse	You haven't even been to the end and out in the Outback Steakhouse	if you haven't even been to the in and out in the outback steakhouse neighborhood	
11 /Users/oishaniba	she gives it a 1 of 10 when often it deserves more	She gives it a one of 10 when often it is so small.	she gives it a one of ten when often it deserves more like	
12 /Users/oishaniba	Guru help out boss I'm running out of money	Guru, help on boss, I am running out of money.	guru help out boss I'm running out of money	
13 /Users/oishaniba	she is leaving for Gore tomorrow	She is leaving for Gorakhpur tomorrow.	she is leaving for gorakhpur tomorrow	
14 /Users/oishaniba	it's not all fun and games with the all subject for them	It's not all fun and games with y'all subg for them.	it's not all fun and games with the all subject for them	
15 /Users/oishaniba	Guru help me out boss I'm running out of money	Guru, help me out boss, I am adding out of money.	guru help me out boss I'm running out of money	
16 /Users/oishaniba	she is leaving for Gore tomorrow	She is leaving for Gorakpoor tomorrow.	she is leaving for gorakhpur tomorrow	
17 /Users/oishaniba	it's not all fun and games with your subbing firm	It's not all fun and games with the all subbing for them.	it's not all fun and games with y'all subbing forum	
18 /Users/oishaniba	everyone expected instant metanoia when meta launched and me	Everyone expected instant metanoia when meta launched and met a mere man to m	everyone expected instant metanoia when meta launched and met a mere man to m	
19 /Users/oishaniba	she's leaving for Houston tomorrow	She's leaving for Houston tomorrow.	she is leaving for houston tomorrow	
20 /Users/oishaniba	he is living for Newton tomorrow	He is leaving for Newton tomorrow.	he is leaving for newton tomorrow	
21 /Users/oishaniba	those 2 are dancing full-on the wedding	Those two are dancing full-on the wedding.	those two are dancing full-on the wedding	
22 /Users/oishaniba	his living for Houston tomorrow	his living for Houston tomorrow.	she's leaving for houston tomorrow	
23 /Users/oishaniba	those 2 were dancing full-on at the wedding	Those two were dancing full-on at the wedding.	those two were dancing full-on at the wedding	
24 /Users/oishaniba	he is leaving for Newton tomorrow	He is leaving for Newton tomorrow.	he is leaving for newton tomorrow	

Methods

Easy Prompt Example for Speaker 1:

- Target Sentence: You haven't even been to the In-n-Out in the Outback Steakhouse neighbourhood?
- Google: you haven't even been to the In-N-Out in the Outback Steakhouse neighborhood, Word Errors: 1
- Whisper: You haven't even been to the end and out in the Outback Steakhouse neighborhood, Word Errors: 4
- Microsoft: you haven't even been to the in and out in the outback steakhouse neighborhood, Word Errors: 4

Methods

Local Prompt Example for Speaker 2:

- Target Sentence: They knew where Rituparno would be on a Friday evening.
- Google: then you were rutu parno would be on a Friday evening, Word Errors: 5
- Whisper: The new Routu Parnu would be on a Friday evening., Word Errors: 4
- Microsoft: they knew where rituparno would be on a friday evening, Word Errors: 0

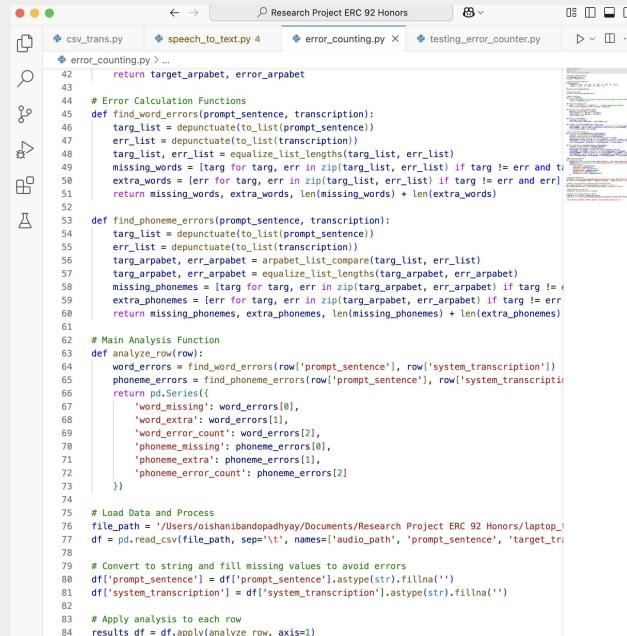
Methods

So - how do we count these errors?

- Word Errors
- Phoneme Errors

Methods

- Word Errors
 - Missing words
 - Substitutions
 - Insertions
- Phoneme Errors
 - Same as above!
 - Think of these as individual speech sounds



The screenshot shows a code editor window with several tabs open. The main tab contains a Python script for error calculation. The script includes functions for finding word errors, phoneme errors, and a main analysis function that processes rows of data. It uses pandas for data manipulation and lists for comparing target and transcription lists. The code is well-structured with comments explaining the logic.

```
csv_trans.py    speech_to_text.py 4  error_counting.py  testing_error_counter.py ...  
error_counting.py > ...  
42     return target_arpabet, error_arpabet  
43  
44 # Error Calculation Functions  
45 def find_word_errors(prompt_sentence, transcription):  
46     targ_list = depunctuate(to_list(prompt_sentence))  
47     err_list = depunctuate(to_list(transcription))  
48     targ_list, err_list = equalize_list_lengths(targ_list, err_list)  
49     missing_words = [targ for targ, err in zip(targ_list, err_list) if targ != err and targ is not None]  
50     extra_words = [err for targ, err in zip(targ_list, err_list) if targ != err and err is not None]  
51     return missing_words, extra_words, len(missing_words) + len(extra_words)  
52  
53 def find_phoneme_errors(prompt_sentence, transcription):  
54     targ_list = depunctuate(to_list(prompt_sentence))  
55     err_list = depunctuate(to_list(transcription))  
56     targ_arpabet, err_arpabet = arpabet_list_compare(targ_list, err_list)  
57     targ_arpabet, err_arpabet = equalize_list_lengths(targ_arpabet, err_arpabet)  
58     missing_phonemes = [targ for targ, err in zip(targ_arpabet, err_arpabet) if targ != err and targ is not None]  
59     extra_phonemes = [err for targ, err in zip(targ_arpabet, err_arpabet) if targ != err and err is not None]  
60     return missing_phonemes, extra_phonemes, len(missing_phonemes) + len(extra_phonemes)  
61  
62 # Main Analysis Function  
63 def analyze_row(row):  
64     word_errors = find_word_errors(row['prompt_sentence'], row['system_transcription'])  
65     phoneme_errors = find_phoneme_errors(row['prompt_sentence'], row['system_transcription'])  
66     return pd.Series({  
67         'word_missing': word_errors[0],  
68         'word_extra': word_errors[1],  
69         'word_error_count': word_errors[2],  
70         'phoneme_missing': phoneme_errors[0],  
71         'phoneme_extra': phoneme_errors[1],  
72         'phoneme_error_count': phoneme_errors[2]  
73     })  
74  
75 # Load Data and Process  
76 file_path = 'Users/oishanibandopadhyay/Documents/Research Project ERC 92 Honors/laptop_labeled.csv'  
77 df = pd.read_csv(file_path, sep='\t', names=['audio_path', 'prompt_sentence', 'target_transcription'])  
78  
79 # Convert to string and fill missing values to avoid errors  
80 df['prompt_sentence'] = df['prompt_sentence'].astype(str).fillna('')  
81 df['system_transcription'] = df['system_transcription'].astype(str).fillna('')  
82  
83 # Apply analysis to each row  
84 results_df = df.apply(analyze_row, axis=1)
```

Methods

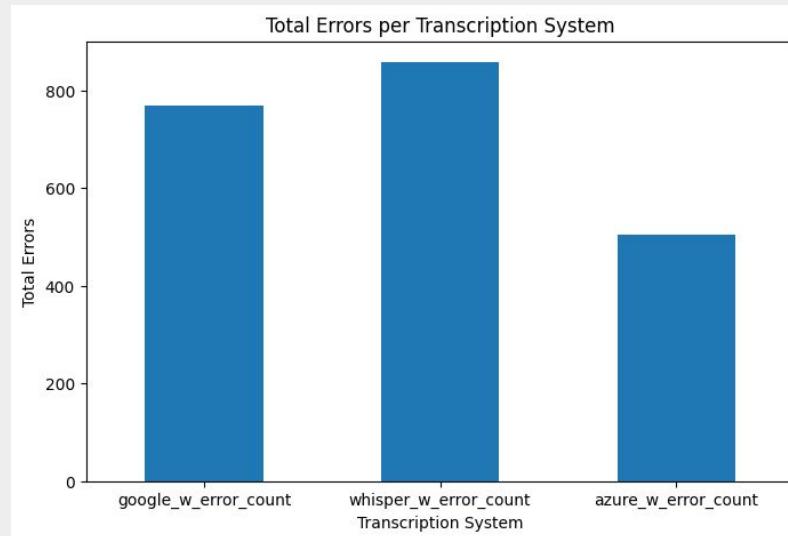
Next Steps:

Seeing how these errors vary across accents and language models!

Then, we can analyze the data and look for a wide variety of patterns - across language backgrounds, different systems, types of prompts, etc!

Results

Here are some results based off of super new data!
(still working on this - but it's interesting at this
preliminary stage as well)



Expanding This Project

- More dialects
- More variation in language backgrounds
- More speech recognition systems
- More ways of error calculation
- More languages
- More data! Equity! Possibly letting big companies with speech to text systems know where they're not doing so good
- And potential improvements to these systems using more diverse data

Where I'm Going Next With This

- Completing this research project as an Honors Thesis and writing a paper about it
- Ideally, publishing and heading to a conference
- Probably grad school!



There's many other ways of getting into research...

Here's Shivani's journey!



Hi, I'm Shivani!

- Co-President of CSSA
- Junior
- Cog Sci ML & NC
- Sixth College



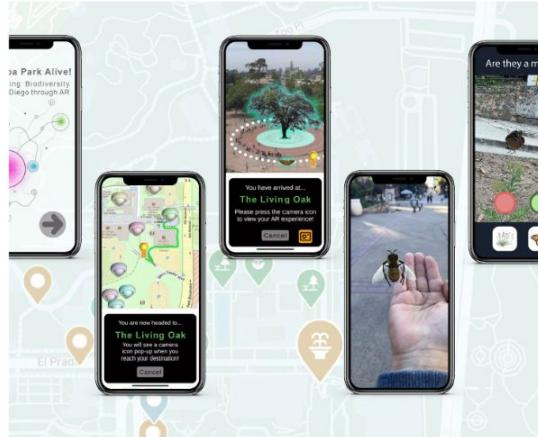
Featured Projects



Embodied Coding

How can augmented and virtual reality support human-centered, collaborative computing?

[Learn more](#)



Balboa Park Alive!

This project aims to foster care for our region's biodiversity by promoting ecological literacy with AR.

[Learn more](#)



Visual Search in VR

We model ecologically valid visual search processes in 3D space using virtual reality.

[Learn more](#)

The Overview Effect: The Impact of Awe on the Brain and Heart

Presented by Shivani Kedila
P.I. Ying Choon Wu
Professor Andrea Chiba



OVERVIEW EFFECT

Awe - Keltner and Haidt

1. Perceived Vastness

2. Need for Accommodation

Research Question

Can a VR simulation of orbiting the earth elicit the Overview Effect?

Autonomic Nervous System

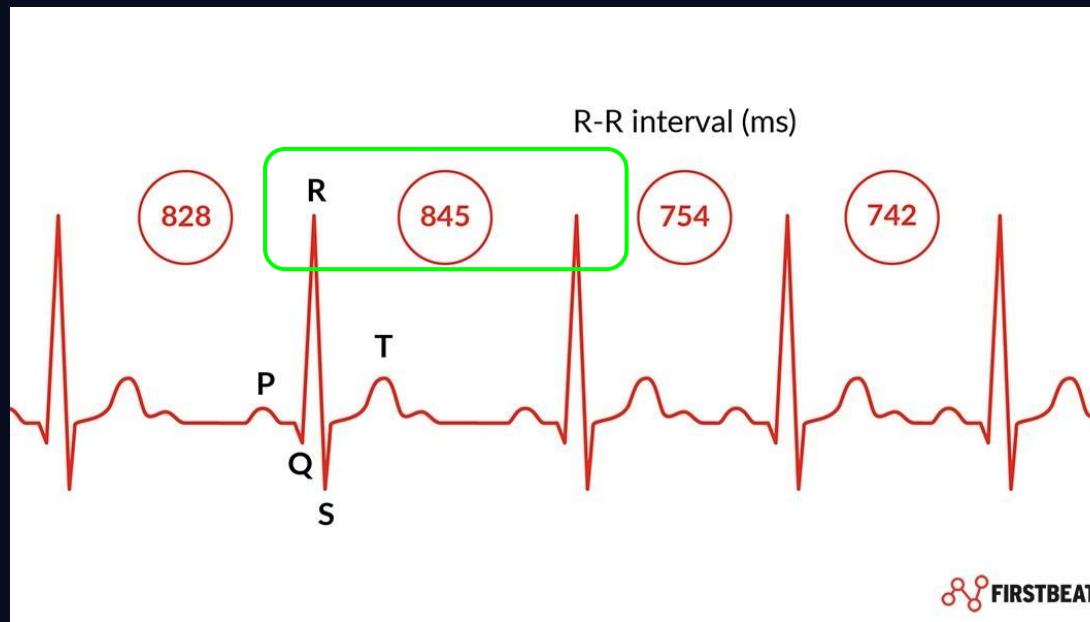
SYMPATHETIC
NERVOUS SYSTEM



PARASYMPATHETIC
NERVOUS SYSTEM



Heart Rate Variability (HRV)



ECG and Respiration

Immersive Awe Experience



Higher Parasympathetic Activity

Hypothesis

During the simulation, heightened parasympathetic activity may indicate when people are entering states consistent with the Overview Effect (e.g. awe, compassion, empathy)

Edgar Mitchell Experience (EMXVR)



Edgar Mitchell Experience (EMXVR)



Post-VR Surveys

AWE-S

Item	Factors					
	1	2	3	4	5	6
I sensed things momentarily slow down.	0.86
I noticed time slowing.	0.86
I felt my sense of time change.	0.78
I experienced the passage of time differently.	0.76
I had the sense that a moment lasted longer than usual.	0.68
I felt that my sense of self was diminished.	.	0.79
I felt my sense of self shrink.	.	0.76
I experienced a reduced sense of self.	.	0.76
I felt my sense of self become	.	0.75				

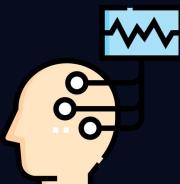
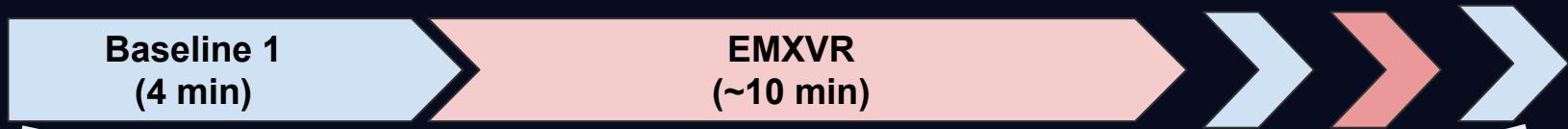
JSE

Number of item	Jefferson Scale of Physician Empathy	Females vs males		Psychiatrists vs other specialties	
		z	p	z	p
2	My patients feel better when I understand their feelings.	-2.429	0.015	-0.119	0.906
3	It is difficult for me to view things from my patients' perspectives	-2.108	0.035	-0.521	0.602
7	I try not to pay attention to my patients' emotions in history taking or in asking about their physical health.	-1.003	0.316	-3.586	0.000
8	Attentiveness to my patients' personal experiences does not influence treatment outcomes.	-2.866	0.004	-3.583	0.000
10	My patients value my understanding of their feelings which is therapeutic in its own right.	-2.236	0.025	-0.190	0.849

Compassion

	almost never 1	2	3	4	almost always 5
1	when people cry in front of me, I often don't feel anything at all				
2	sometimes when people talk about their problems, I feel like I don't care				
3	I don't feel emotionally connected to people in pain				
4	I pay careful attention when other people talk to me				
5	I feel detached from others when they tell me their tales of woe				
6	If I see someone going through a difficult time, I try to be caring toward that person				
7	I often tune out when people tell me about their troubles				
8	I like to be there for others in times of difficulty				
9	I notice when people are upset, even if they don't say anything				
10	when I see someone feeling down, I don't like I can't relate to them				
11	everyone feels down sometimes, it is part of being human				
12	sometimes I am cold to others when they are down and out				

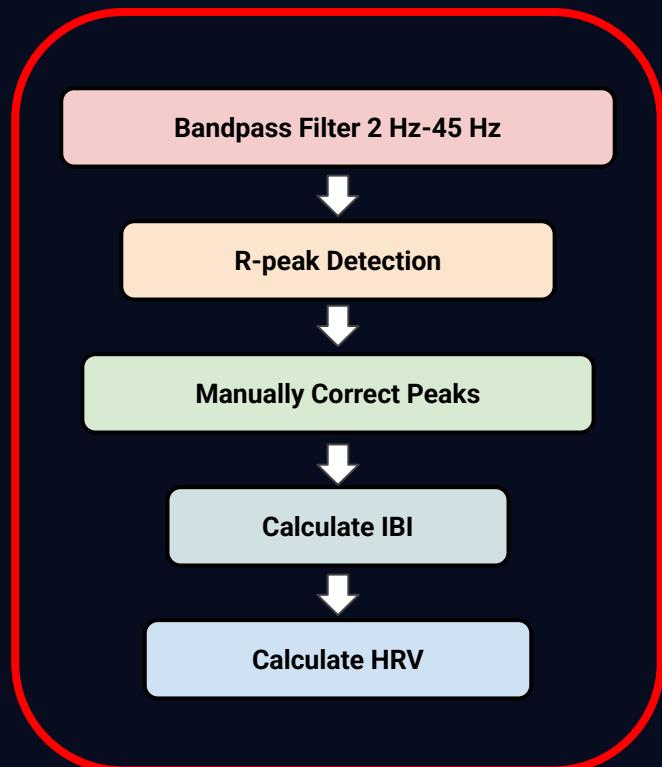
Methods: Data Collection



- Biosemi
- 512 Hz
- 64 Channels
- 2 ECG (L/R)



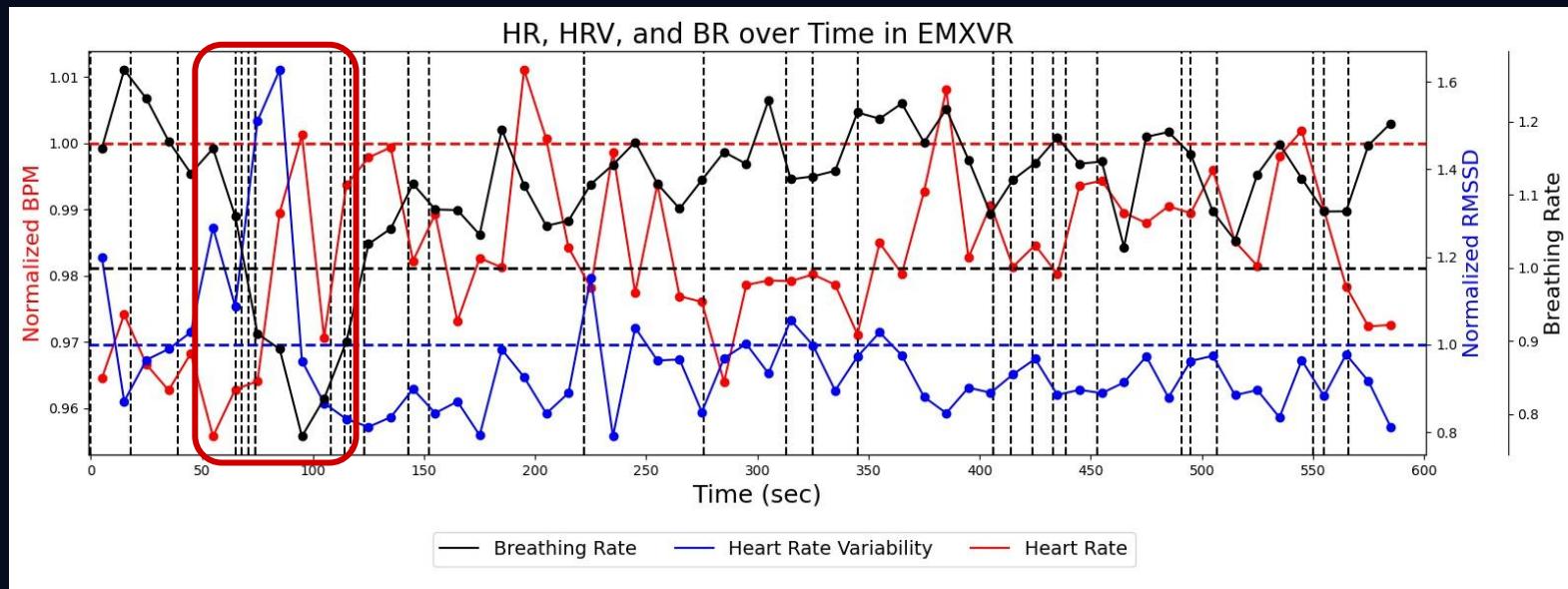
Methods: Preprocessing ECG



Predictions

- ECG and Respiration
 - Increase in Heart Rate Variability
 - Decrease in Breathing Rate
- Survey Scores
 - AWE-S
 - Compassion
 - Empathy

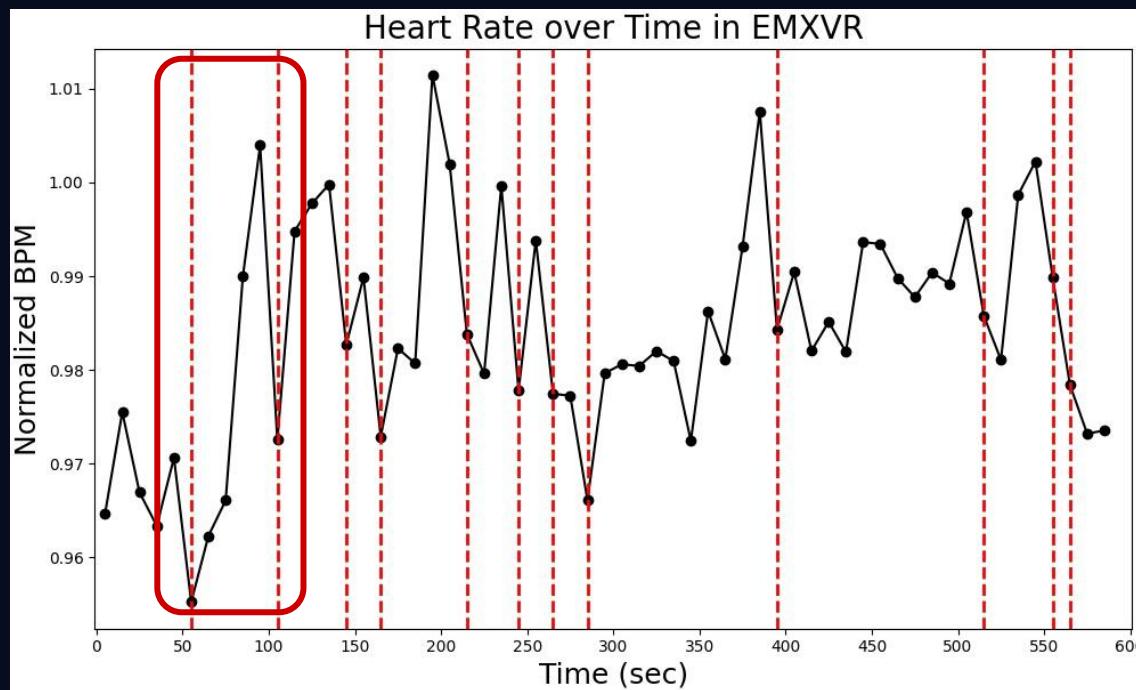
Results: Heart Rate, HRV, and Breathing Rate



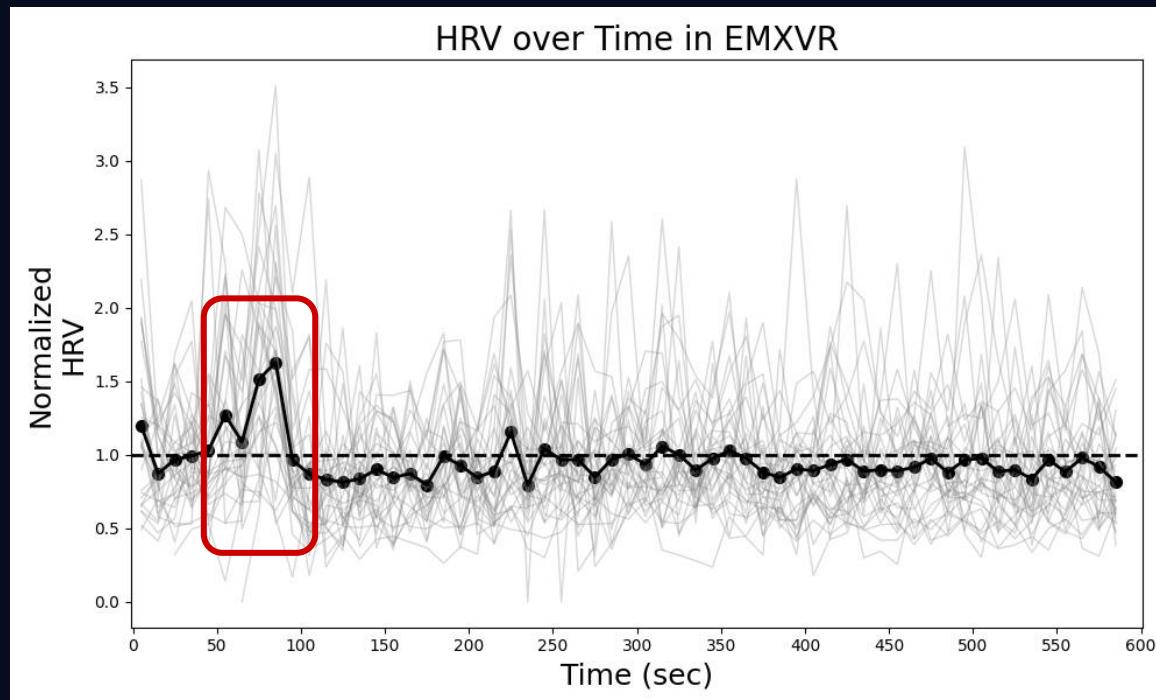
EMXVR - 80 to 100 seconds



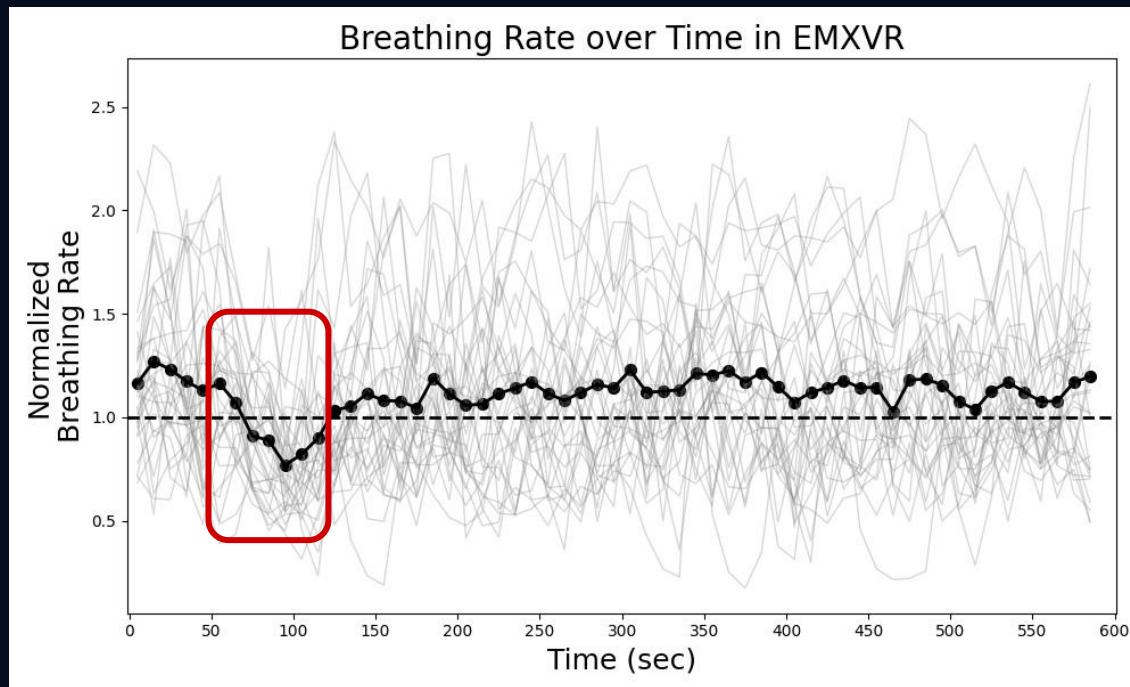
Results: Increase in Heart Rate



Results: Increase in HRV



Results: Drop in Breathing Rate



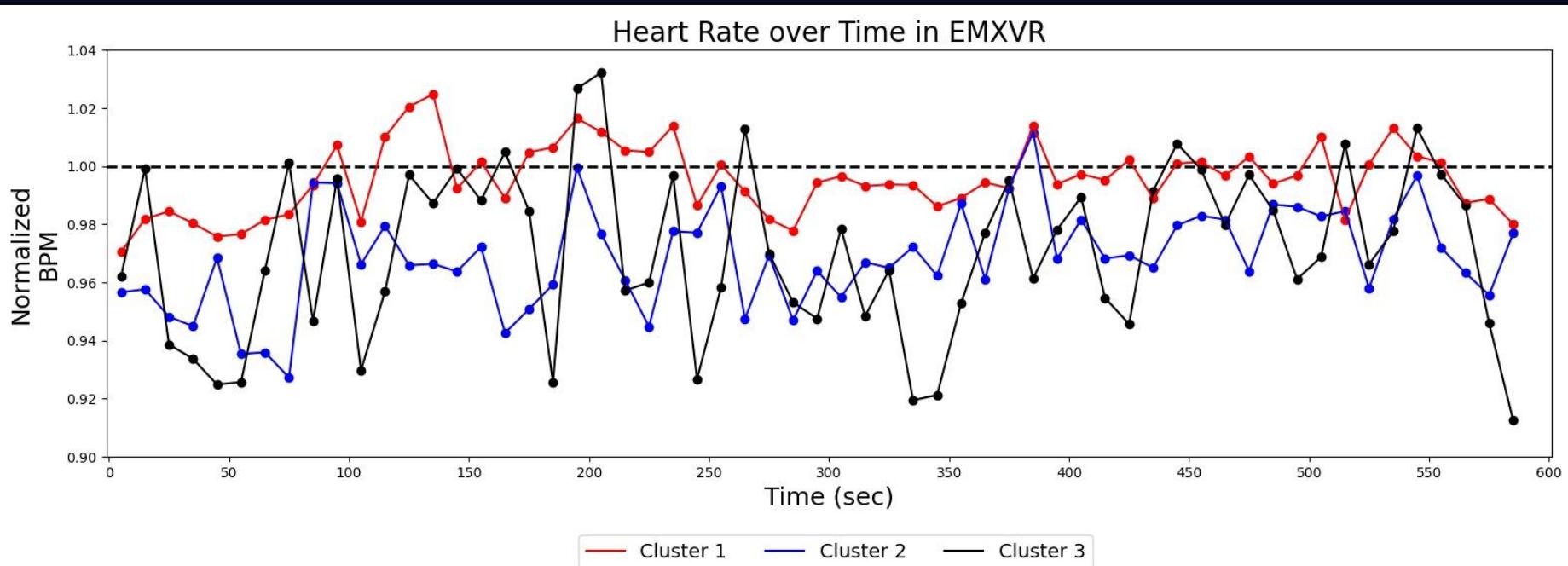
Hierarchical Clustering: Survey Data



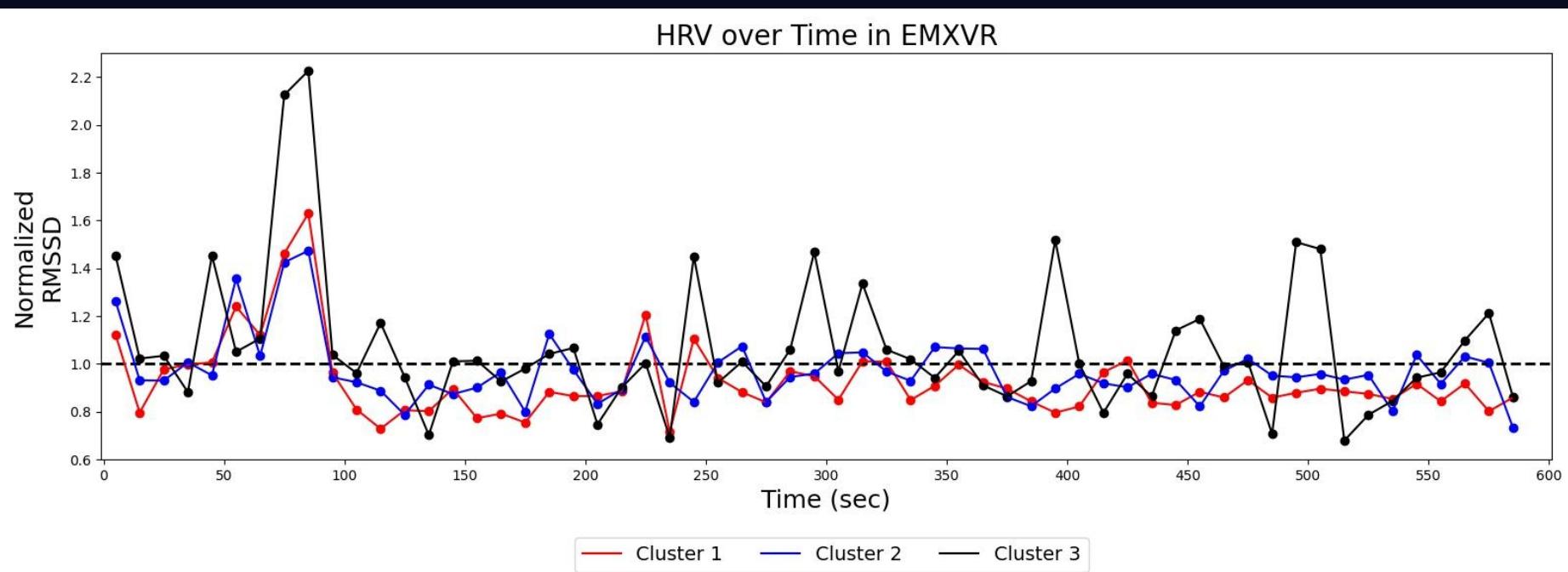
Results: Cluster Scores

	Cluster 1	Cluster 2	Cluster 3
Awe	66.41	50.42	30.00
Compassion	4.11	4.07	4.15
Empathy	29.06	27.58	29.67

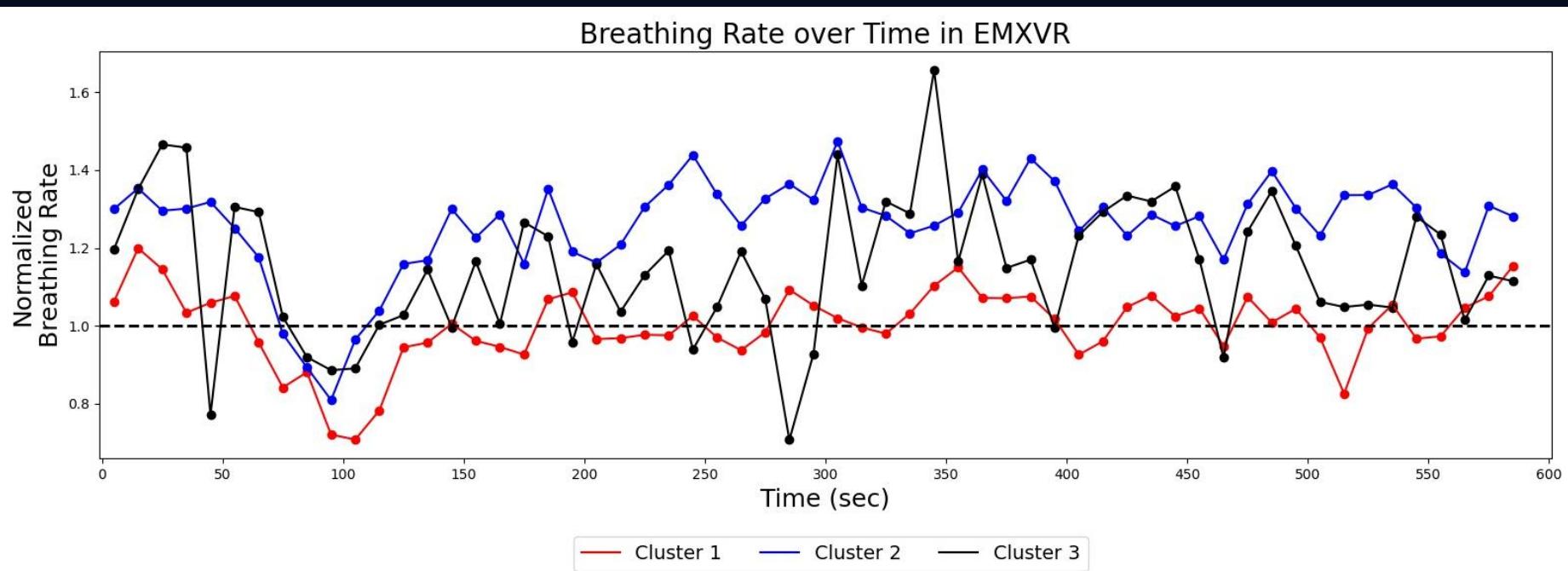
Results: HR



Results: HRV



Results: Breathing Rate



Results: Summary

- ECG
 - Certain time segments cause an increase in HRV and decrease in breathing rate, indicating parasympathetic nervous system activity
 - Decrease in breathing rate was accompanied with an increase in heart rate

Results: Summary Cont.

- Survey
 - No significant correlation in awe scores and compassion and empathy scores

Conclusion

- Certain segments of EMXVR elicit physiological responses associated with the Overview Effect
- Cluster results do not indicate that high awe scores were associated with increased compassion and empathy scores

What does this mean?

Future Directions

- Further analysis and data collection is necessary to support our findings
- With more study, VR could be a beneficial tool to replicate awe experiences

Thank you!
