# Inferring Respondents' Emotional States from Text and Voice Answers to Open Questions in a Smartphone Survey


**Claaßen[1,2], Höhne[1,2], Kern[3], & Avetisyan[1]**

[1] German Center for Higher Education Research and Science Studies (DZHW)

[2] Leibniz University Hannover

[3] University of Munich


**WEB DATA OPP Workshop**

Barcelona (Spain) - 18th March 2024

CS3 lab

Computational Survey and Social Science

# Introduction

- Open questions with requests for voice answers are promising
    - *Content: Voice answers are longer and contain more topics* (Gavras et al., 2022; Höhne & Claassen, 2024)
    - *Data Quality: Voice answers have higher criterion validity* (Gavras & Höhne, 2022)
    - *Missing data: High dropout and high item-nonresponse* (Revilla & Couper, 2021; Revilla et al., 2020)
- Voice answers contain tonal cues for inferring emotional states (Höhne et al., 2023)
    - *In-situ inferences of emotional states in contrast to global measures*
    - *Shedding light on engagement and data quality*
- Inferences from tonal cues have limitations
    - *Time-consuming data processing*
    - *Do not consider textual content*
    - *Only possible for voice answers*
- In this study, we therefore analyze both text and transcribed voice answers

# Research Questions (RQs)

- **RQ1**: Do respondents' emotional states inferred through sentiments and transformer models align with each other?

- **RQ2:** How sensitive are inferences of respondents' emotional states to manipulations induced by an environmental treatment?

- **RQ3**: Are respondents' emotional states related to data quality?

# Method: Study Design

- Smartphone survey (N = 1,001) in Germany in November 2021
  - *Cross quota on age, gender plus quota on education*
  - *Average age: **48 years**; female: **49%**; medium education: **42%**; high education: **28%***
- Two open questions
  - ***Q1**: To begin with, we would like to ask you to tell us in your own words how you feel at this moment? Please answer in as much detail as possible.*
  - ***Q2**: What do you think the world will look like in 10 years? Please answer in as much detail as possible.*
- Two independent experiments
  - ***Experiment 1:** Request for text or voice answers*
  - ***Experiment 2**: Picture of unhealthy or healthy environment*
- Voice answers were collected with the open-source SurveyVoice tool (Höhne et al., 2021)

# Method: Study Procedure

**Experiment 1** Question (Q) 1 | **Experiment 2** | **Experiment 1** Question (Q) 2


Text | Voice | Unhealthy | Healthy | Text | Voice

# Method: Analytical Strategy

- Transcription of voice answers via OpenAI's Whisper (Radford et al., 2023)
  - *Manual inspection of 20% of the recordings (n = 120)*
  - *High transcription quality*
- Determining sentiments using SentiWS v2.0 (Remus et al., 2010)
- Determining emotion probabilities using a transformer model
  - *"xlm-roberta-large-xnli" model from Hugging Face (www.huggingface.com)*
  - *Probability with which an emotion follows from an answer*
  - *Emotions: Anger, disgust, fear, joy, sadness, surprise, and contempt* (Ekman & Friesen, 1986)
- Determining the number of words using Quanteda (R) (Benoit et al., 2021)
- Determining the number of topics using STM (R) (Roberts et al., 2014)

# Results: Exemplary Answers

Table 1a. Exemplary **text** answers to Q1 including sentiments and emotion probabilities

| Answer | Sentiment | Anger | Disgust | Fear | Joy | Sadness | Surprise | Contempt |
|---|---|---|---|---|---|---|---|---|
| I am depressed and lack motivation. | -6.16 | 0.22 | **0.81** | 0.47 | 0.00 | **0.99** | **0.86** | 0.74 |
| Healthy and satisfied. | 6.31 | 0.01 | 0.02 | 0.00 | **0.99** | 0.00 | **0.78** | 0.63 |

Note. Emotion probabilities >= 0.7 in bold.

Table 1b. Exemplary **voice** answers to Q1 including sentiments and emotion probabilities

| Answer | Sentiment | Anger | Disgust | Fear | Joy | Sadness | Surprise | Contempt |
|---|---|---|---|---|---|---|---|---|
| I feel good, not stressed, and refreshed after my vacation. | 0.66 | 0.02 | 0.02 | 0.01 | **0.99** | 0.00 | **0.90** | 0.51 |
| Tired, unmotivated, annoyed, not good. | 1.53 | **0.90** | **0.81** | 0.11 | 0.00 | **0.80** | 0.16 | 0.66 |

Note. Emotion probabilities >= 0.7 in bold.

# Results: Research Question 1

Table 2a. Correlations between sentiments and emotion probabilities – <u>**Text**</u>

| | Anger | Disgust | Fear | Joy | Sadness | Surprise | Contempt |
|---|---|---|---|---|---|---|---|
| Sentiment (Q1) | **-0.54** | **-0.57** | **-0.52** | **0.67** | **-0.61** | 0.02 | **-0.27** |
| Sentiment (Q2) | **-0.21** | **-0.40** | **-0.39** | **0.33** | **-0.25** | 0.01 | **-0.21** |

Note. Coefficients with $p < 0.05$ in bold.

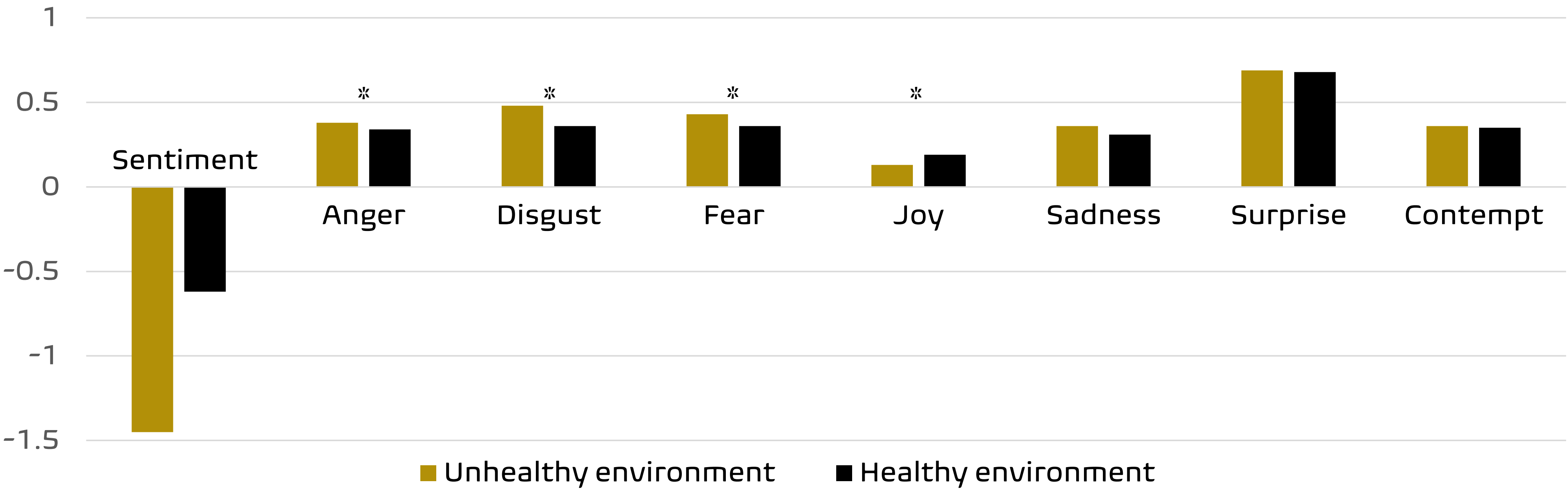
Table 2b. Correlations between sentiments and emotion probabilities – <u>**Voice**</u>

| | Anger | Disgust | Fear | Joy | Sadness | Surprise | Contempt |
|---|---|---|---|---|---|---|---|
| Sentiment (Q1) | **-0.51** | **-0.64** | **-0.58** | **0.71** | **-0.63** | 0.10 | **-0.30** |
| Sentiment (Q2) | **-0.19** | **-0.31** | **-0.35** | **0.29** | **-0.24** | -0.12 | **-0.16** |

Note. Coefficients with $p < 0.05$ in bold.

# Results: Research Question 2

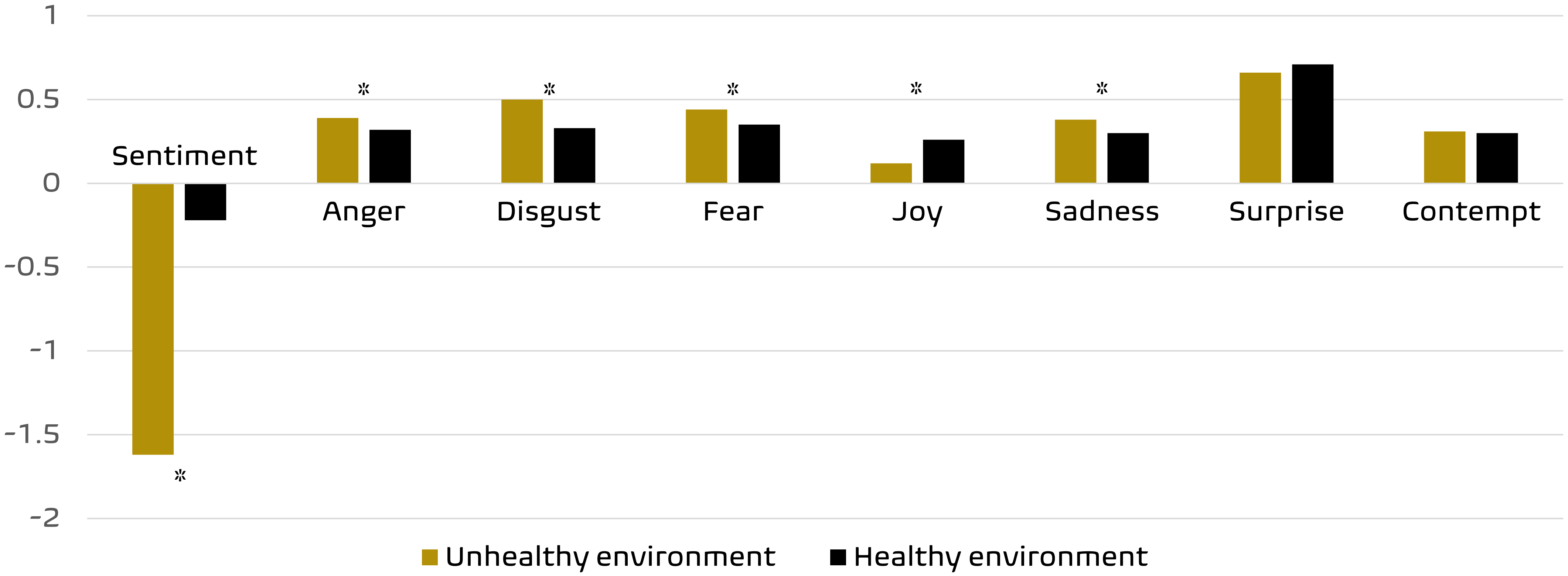
Figure 1a. Emotional state differences between picture conditions (Q2) – **Text**




Note. * $p < 0.05$. We conducted t-tests for independent groups.

# Results: Research Question 2

Figure 1b. Emotional state differences between picture conditions (Q2) - <u>Voice</u>




Note. * $p < 0.05$. We conducted t-tests for independent groups.

# Results: Research Question 3

Table 3a. Correlations between sentiments, emotion probabilities, and answer length - **Text**

| | Sentiment | Anger | Disgust | Fear | Joy | Sadness | Surprise | Contempt |
|---|---|---|---|---|---|---|---|---|
| Answer length (Q1) | **-0.11** | -0.02 | **0.19** | **0.21** | **-0.13** | 0.06 | -0.00 | **-0.13** |
| Answer length (Q2) | 0.00 | -0.04 | 0.04 | 0.04 | **-0.11** | -0.04 | **-0.12** | **-0.15** |

Note. Coefficients with $p < 0.05$ in bold.

Table 3b. Correlations between sentiments, emotion probabilities, and answer length - **Voice**

| | Sentiment | Anger | Disgust | Fear | Joy | Sadness | Surprise | Contempt |
|---|---|---|---|---|---|---|---|---|
| Answer length (Q1) | **-0.18** | **0.21** | **0.31** | **0.30** | -0.06 | **0.21** | -0.03 | **0.21** |
| Answer length (Q2) | 0.10 | **0.23** | **0.21** | **0.18** | **0.12** | **0.20** | 0.02 | **0.18** |

Note. Coefficients with $p < 0.05$ in bold.

# Results: Research Question 3

Table 4a. Correlations between sentiments, emotion probabilities, and number of topics - **Text**

| | Sentiment | Anger | Disgust | Fear | Joy | Sadness | Surprise | Contempt |
|---|---|---|---|---|---|---|---|---|
| Topic number (Q1) | **-0.14** | -0.02 | -0.05 | -0.06 | -0.07 | 0.03 | 0.01 | **-0.10** |
| Topic number (Q2) | 0.09 | 0.01 | -0.02 | -0.04 | -0.01 | **-0.10** | 0.07 | -0.06 |

Note. Coefficients with $p < 0.05$ in bold.

Table 4b. Correlations between sentiments, emotion probabilities, and number of topics - **Voice**

| | Sentiment | Anger | Disgust | Fear | Joy | Sadness | Surprise | Contempt |
|---|---|---|---|---|---|---|---|---|
| Topic number (Q1) | **-0.26** | 0.04 | 0.11 | 0.06 | **-0.23** | **0.13** | **-0.14** | 0.06 |
| Topic number (Q2) | 0.04 | -0.03 | -0.06 | -0.06 | 0.07 | -0.00 | -0.05 | -0.07 |

Note. Coefficients with $p < 0.05$ in bold.

# Discussion and Conclusion

- Moderate to strong correlations between sentiments and emotion probabilities
    - *Stronger correlations for question on in-situ feelings (Q1)*
    - *Patterns hold for both text and voice answers*
- Emotional states are sensitive to environmental treatment – text and voice
    - *<u>Negative</u> sentiments and emotions are more prevalent in <u>unhealthy</u> environment condition*
    - *<u>Positive</u> sentiments and emotions are more prevalent in <u>healthy</u> environment condition*
- Moderate correlations between emotional states and answer length
    - *Strength and direction differs between text and voice answers*
- Few correlations between emotional states and topic number – text and voice
- Take home message: Emotional states can be inferred from text and transcribed voice answers and they inform about answer behavior

# Many thanks for your attention!

claassen@dzhw.eu
@JClaass

# Questions for Discussion

1) What alternative approaches can be used to infer sentiments and discrete emotions?

2) What additional data quality indicators should be considered?

3) How can we combine transcribed voice answers with tonal features to infer emotional states?

# Literature I


- Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., Matsuo, A., Lowe, W., & Müller, C. (2018). Quanteda: R package for quantitative analysis of textual data. https://cran.rproject.org/web/packages/quanteda/quanteda.pdf
- Ekman, P., & Friesen, W. V. (1986). A new pan-cultural facial expression of emotion. Motivation and Emotion, 10(2), 159-168. doi: 10.1007/BF00992253
- Gavras, K., & Höhne, J. K. (2022). Evaluating political parties: Criterion validity of open questions with requests for text and voice answers. International Journal of Social Research Methodology, 25, 135-141. doi: 10.1080/13645579.2020.1860279
- Gavras, K., Höhne, J.K., Blom, A., & Schoen, H. (2022). Innovating the collection of open-ended answers: The linguistic and content characteristics of written and oral answers to political attitude questions. Journal of the Royal Statistical Society (Series A), 185, 872-890. doi: 10.1111/rssa.12807
- Höhne, J. K., & Claassen, J. (2024). Examining final comment questions with requests for written and oral answers. International Journal of Market Research. doi: 10.1177/14707853241229329
- Höhne, J. K., Gavras, K., & Qureshi, D. D. (2021). SurveyVoice (SVoice): A comprehensive guide for collecting voice answers in surveys. GitHub. https://github.com/JKHoehne/SVoice/tree/v1.0.0
- Höhne, J.K., Kern, C., Gavras, K. & Schlosser, S. (2023). The sound of respondents: Predicting respondents' level of interest in questions with voice data in smartphone surveys. Quality and Quantity. doi: 10.1007/s11135-023-01776-8
- Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023). Robust speech recognition via large-scale weak supervision. Proceedings of the 40th International Conference on Machine Learning, 28492–28518. doi: 10.5555/3618408.36 19590
- Remus, R., Quasthoff, U. & Heyer, G. (2010) SentiWS - a Publicly Available German-Language Resource for Sentiment Analysis. Proceedings of the 7th International Language Resources and Evaluation, 1168–1171.
- Revilla, M., & Couper, M. P. (2021). Improving the use of voice recording in a smartphone survey. Social Science Computer Review. doi: 10.1177/0894439319888708

# Literature II


- Revilla, M., Couper, M. P., Bosch, O. J., & Asensio, M. (2020). Testing the use of voice input in a smartphone web survey. Social Science Computer Review, 38(2), 111-244. doi: 10.1177/0894439318810715
- Roberts, M.E., Stewart, B.M., & Tingle, D. (2014). stm: R package for structural topic models. https://cran.r-project.org/web/packages/stm/vignettes/stmVignette.pdf