Affect Experience in Natural Language Collected With Smartphones

Inaugural-Dissertation
zur Erlangung des Doktorgrades der Philosophie
an der Ludwig-Maximilians-Universität München

vorgelegt von
Timo Koch
aus München

2024

Erstgutachter: Prof. Dr. Markus Bühner

Zweitgutachter: Prof. Dr. Clemens Stachl

Tag der mündlichen Prüfung: 07.07.2023

Acknowledgements

I had the privilege to have an incredibly supportive work and social environment during my journey as a doctoral candidate that helped me overcome all those challenges associated with completing this dissertation.

First, I would like to thank my supervisor, Markus Buehner, for his perpetual support, guidance, and providing me with generous academic freedom. Moreover, a big thank you goes out to my co-supervisor, Clemens Stachl, for the fantastic operational supervision, research inspiration, and mentorship.

I am deeply grateful to the German Academic Scholarship Foundation (Studienstiftung des Deutschen Volkes) for funding this dissertation project, my research visit at Stanford University, and providing me with ample resources and the freedom to grow personally and professionally. This scholarship has changed the course of my life.

Furthermore, I would like to thank CHECK24 for the opportunity to carry out an applied research project at the beginning of my PhD. Particularly, I want to express my gratitude to Bjoern Zollenkop, Dora Simroth, and Florian Weber, who made this project possible. I truly enjoyed working with the amazing Data Science team around Andrea Di Simone (thank you for supervising the project!), Peter Wagenpfeil, Tim Schulz, Mihaela Constantinescu, Fabian Brunn, and Andreas Stephan. Also, I want to thank the entire "Human-AI-Interaction in Healthcare" project team for providing me with a warm and stimulation research environment outside of my dissertation work during the last 1.5 years.

Another big thank you goes to the whole PhoneStudy research group. This dissertation project would not have been possible without this great team effort. In this context, I want to thank the current and prior PhoneStudy researchers: Quay Au, Daniel Buschek, Fiona Kunz, Michelle Oldemeier, Sophia Sakel, Larissa Sust, and many more. Specifically, I want to thank Florian Bemmann for implementing all those technical features into the app and Ramona Schoedel for tirelessly managing everything around the PhoneyStudy. I am standing on the shoulders of giants.

Also, I am grateful for my colleagues at the chair of Psychological Methods at LMU Munich for the pleasant and supportive work environment. In this context, I

want to thank Laura Israel, Elisabeth Kraus, Felix Schoenbrodt, Philipp Sckopke, Lena Schiestel, Lena-Marei Wiessner, and Caroline Zygar-Hoffmann. I am thankful for the great shared office companionship with Ricarda Luebke, Felix Naumann, Thomas Reiter, and Philipp Sterner. Also, I want to specifically thank Florian Pargent and David Goretzko for all their help with any R programming and statistics questions.

I want to express my gratitude for David Stillwell and the entire team of the Psychometrics Centre at the University of Cambridge for inspiring me to pursue a PhD by showing me how cool academic research can be. In this context, I would like to also thank Peter Romero, who helped me with learning to write decent code. Moreover, during my dissertation work and in other research projects, I got to collaborate with so many extremely smart and generally amazing human beings. Here, I'd like to specifically thank Sanaz Talaifar, Johannes Eichstaedt, Lena Frischlich, and Eva Lermer for the great collaboration and passing on of your academia wisdom.

Thank you, Gabriella Harari, for giving me the amazing opportunity to do a research visit at Stanford University that had a profound impact on my personal and professional development. Here, I would also like to express my gratitude for the whole Media and Personality Lab (especially Sumer Vaid, Mahnaz Roshanaei, and Serena Soh) for making my stay such a fun and inspiring time. Thank you Katrin Wheeler for your administrative support during my stay and forwarding me all those great campus events. Also, I am extremely grateful for the whole Alger Drive Crew who has made this such an awesome stay in California.

A big thank you goes out to my friends and family. Here, I would like to express my gratitude for my fellow PhDs Alejandro, Moritz, and Tom. Thank you, Alex, Carina K., Carina W., Jakob, Max, Simon, Tobi, and the many others who made the "life" side of the "work-life balance" so joyful. Also, I'd like to thank the Brudermueller family for all their encouragement and support at the beginning of my PhD endeavor.

I am grateful for my family for their support and for giving me the opportunity to pursue my own dreams. I want to thank my parents, Antje and Thomas, for providing me with more academic opportunities than they had and always encouraging me from early on. Thank you, my sisters, Solvi und Carolin, for bringing me back down to earth when I was high up in the (academic) sky.

Finally, I want to express my deep gratitude for Johanna, my closest companion and PhD partner in crime. With your strength, positivity, and love you helped me find and maintain balance. Thank you for being such an amazing human and always being on my side.

List of Abbreviations

AER	Automatic Emotion Recognition
AI	Artificial Intelligence
API	Application Programming Interface
CI	Confidence Interval
CV	Cross-Validation
EAR	Electronically Activated Recorder
ESM	Experience Sampling Method
LASSO	Least Absolute Shrinkage and Selection Operator
LIWC	Linguistic Inquiry and Word Count
M	Mean
Md	Median
NA	Negative Affect
NLP	Natural Language Processing
OSF	Open Science Framework
PA	Positive Affect
PANAS	Positive and Negative Affect Schedule
RMSE	Root Mean Square Error
SD	Standard Deviation
eGeMAPS	Extended Geneva Minimalistic Acoustic Parameter Set

Table of Contents

List of Tables

List of Figures

Abstract

Recent technological advancements in computerized text and speech analysis as well as machine learning methods have sparked a growing body of research investigating the algorithmic recognition of affect from the ubiquitous digital traces of natural language data and corresponding affect-linked language variations. Also, commercial interest to leverage these new data using AI for affect inferences is on the rise. However, due to the challenges associated with collecting data on subjective affect experience and corresponding language samples, previous research studies and commercial products have mostly relied on data sets from labelled text or enacted speech and, thereby, are focused on affect *expression*. This work leverages new smartphone-based data collection methods to collect self-reports on in-situ subjective affect *experience* and corresponding language samples in the wild to investigate between-person differences and within-person fluctuations in affect experience.

The present dissertation aims to achieve three goals: (1) to investigate if between-person differences and within-person fluctuations in subjective affect experience are associated with and predictable from cues in spoken and written natural language, (2) to identify specific language characteristics, such as the use of specific word categories or voice parameters, that are associated with and predictive of affect experience, and (3) to analyze the influence of the context of language production on the associations and predictions of affect experience from natural language.

This work is comprised of two empirical studies that analyze self-reports on subjective affect experience and natural language data collected with smartphones. Study 1 investigates predictions of between-person differences and within-person fluctuations in subjective momentary affect experience in more than 23000 speech samples from over 1000 participants in two data sets from Germany and the United States. In contrast to voice acoustics, which contain limited predictive information for affective arousal, state-of-the-art word embeddings yield significant above-chance predictions for affective arousal and valence. Moreover, interpretable machine learning methods are used to identify those voice features (i.e., loudness and spectral features) that are most predictive of affect experience. Finally, the work suggests that affect predictions from voice cues from semi-structured free speech are superior to those from read-out predefined sentences and that the emotional sentiment of the spoken content has no effect on affect predictions from voice cues.

Study 2 analyzes patterns in written language data logged through smartphones' keyboards to investigate how between-person differences and within-person fluctuations in affect experience manifest in and are predictable from logged text data across different time frames and communication contexts. From a data set of more than 10 million typed words, features regarding typing dynamics, word use based on word dictionaries, and emoji and emoticon use are computed. From the data, distinct affect-linked language variations across communication contexts (private messaging versus public posting) and time frames (trait, weekly, daily, momentary) are identified (e.g., the use 1st person singular). Predictions of affect experience from machine learning algorithms, however, are not significantly better than chance. Results of this study highlight the challenges of using occurrence-counts, such as word dictionaries, for the assessment of subjective affect experience.

By leveraging novel smartphone-based experience sampling and on-device language data collection in everyday life, the present dissertation shows how characteristics of spoken and written language are associated with and predictive of subjective affect experience. Thereby, this work highlights the utility of smartphones for investigating subjective affect experience in natural language in the wild, overcoming the caveats of prior research methods. Prediction results, however, challenge the optimistic prediction performances reported in prior works on the recognition of affect expression experience. Using statistical methods from the areas of description, prediction, and explanation, the present dissertation also reveals specific affect-linked language characteristics. Finally, results underline the relevance of the context of language production on language characteristics and corresponding affect predictions. The promising applications and potential future directions of this technology come with multiple challenges with regard to the conceptualization of affect, interdisciplinarity, ethics, and data privacy and security. If these challenges can be overcome, natural language analysis based on data collected with smartphones represents a promising tool to monitor affective well-being and to advance the affective sciences.

Zusammenfassung

Die jüngsten technologischen Fortschritte in der computergestützten Text- und Sprachanalyse sowie bei den Methoden des maschinellen Lernens haben eine wachsende Anzahl an Forschungsarbeiten ermöglicht, die sich mit der algorithmischen Erkennung von Affekt aus den allgegenwärtigen digitalen Spuren natürlicher Sprache und den dazugehörigen affektbezogenen Sprachvariationen beschäftigen. Angesichts dieser wissenschaftlichen Fortschritte und dem Aufkommen von Sprachassistenten, Chatbots und anderen text- oder sprachbasierten Diensten, die große Mengen an Text- und Sprachdaten generieren, hat die Gewinnung psychologischer Einblicke in unser Gefühlsleben anhand von Sprache auch zunehmend kommerzielle Aufmerksamkeit auf sich gezogen. Technologieunternehmen bieten sprachbasierte Tools zur Emotionserkennung an, die künstliche Intelligenz (KI) nutzen, um Erkenntnisse über Emotionen von Mitarbeiterinnen und Kundinnen zu gewinnen. Die eingesetzten Algorithmen können dabei bei wichtigen Entscheidungen mitwirken oder diese sogar selbst treffen, beispielsweise wer eingestellt wird, wer eine psychologische Behandlung erhält, oder welches Produkt auf den Markt gebracht wird. Das Innenleben dieser kommerziell eingesetzten Algorithmen und die Daten, auf denen sie trainiert wurden, sind jedoch für die Nutzer in der Regel nicht transparent einsehbar und werden nur selten mit Wissenschaftlern geteilt, was die Untersuchung der Gültigkeit ihrer Vorhersagen erschwert. Im verwandten Bereich der KI-basierten Emotionserkennung anhand von Gesichtsausdrücken zeigten wissenschaftliche Untersuchungen, dass die eingesetzten Tools häufig ungenaue und verzerrte Ergebnisse lieferten. Dies hat dazu geführt, dass mehrere Technologieunternehmen ihre Dienste zur Emotionserkennung aus Gesichtsausdrücken inzwischen wieder eingestellt haben.

In der bisherigen Forschung zur Affekterkennung aus Sprache gibt es zwei wesentliche Herausforderungen, welche die vorhergehenden Arbeiten einschränken: Dazu gehört die konzeptuelle Unterscheidung zwischen dem *subjektiven Affekterleben* ("Wie fühle ich mich in diesem Moment?"), welches aus wissenschaftlicher und

praktischer Sicht von großer Relevanz ist, und dem *sichtbaren Affektausdruck* ("Welche Gefühle drücke ich wie aus?"), welcher bereits vermehrt untersucht wurde. Dies liegt an der empirischen Herausforderung, Daten zum subjektiven Affekterleben und entsprechende *zeitnahe* Sprachproben zu sammeln. Aufgrund dieser Herausforderungen basieren bisherige Forschungsarbeiten und kommerzielle Produkte meist auf Datensätzen aus Text- oder Sprachdaten, die entweder von Probanden bezüglich ihres affektiven Gehaltes bewertet oder von Schauspielern eingesprochen wurden. Somit konnten bisherige Forschungsarbeiten und Produkte lediglich den *Ausdruck* von Affekt untersuchen. Um trotz der genannten Herausforderungen das tatsächliche subjektive *Erleben* von Affekt aus Text und Sprache mittels Algorithmen erkennen zu können, benötigen Forscher ein Instrument, welches es ihnen ermöglicht, Daten zum Affekterleben und entsprechende *zeitnahe* Sprachproben, die ein authentisches Affekterleben in Echtzeit festhalten können, zu sammeln. Ein solches Forschungsinstrument wurden den Wissenschaftlern mit dem technologischen Fortschritt bei handelsüblichen Smartphones an die Hand gegeben. Damit ist es möglich, sowohl Selbstberichte über das subjektive Affekterleben per App (durch die sogenannte *Experience-Sampling Methode*), als auch im Alltag entstehende Sprachdaten über die Tatstatur und das eingebaute Mikrofon in großen Mengen zu sammeln. Die vorliegende Arbeit nutzt diese neuen Smartphone-basierten Datenerhebungsmethoden, um Selbstberichte über das subjektive Affekterleben und dazugehörige Sprachproben im Alltagsleben zu erheben und darauf basierend Unterschiede zwischen Personen und Schwankungen innerhalb von Personen im subjektivem Affekterleben zu untersuchen.

Die vorliegende Dissertation verfolgt drei Ziele: (1) zu untersuchen, ob Unterschiede zwischen Personen und Fluktuationen innerhalb von Personen im subjektiven Affekterleben mit Merkmalen in gesprochener und geschriebener natürlicher Sprache assoziiert und vorhersagbar sind, (2) spezifische Sprachmerkmale, wie die Verwendung bestimmter Wörter aus Wortkategorien oder Stimmparameter, zu identifizieren, die mit dem Affekterleben assoziiert und vorhersagbar sind, und (3) den Einfluss des Kontextes der Sprachproduktion auf die Assoziationen mit und den Vorhersagen von Affekterleben aus natürlicher Sprache zu untersuchen.

Diese Arbeit besteht aus zwei empirischen Studien, die Selbstberichte zum subjektiven Affekterleben und gesprochene sowie geschriebene natürliche Sprache, die mit Smartphones erhoben wurden, analysieren. Studie 1 untersucht Unterschiede zwischen

Personen sowie Schwankungen innerhalb von Personen im subjektiven momentanen Affekterleben in mehr als 23000 Sprachproben von über 1000 Studienteilnehmern aus Deutschland (Studie 1.1) und den Vereinigten Staaten von Amerika (Studie 1.2). In Studie 1.1 haben die Teilnehmer vorgegebene Sätze unterschiedlicher emotionaler Valenz (positiv/ neutral/ negativ) in das Mikrofon ihres Smartphones eingesprochen. Aus den extrahierten akustischen Stimmparametern wurde dann das selbstberichtete Affekterleben auf den Dimensionen Valenz und Aktivation durch maschinelles Lernen vorhergesagt. Hierbei war keines der Vorhersagemodelle signifikant besser als der Zufall, wobei die Vorhersage von Aktivation im Durchschnitt etwas genauer ausfiel. In Studie 1.2 haben die Probanden im Rahmen der Aufnahmen frei über ihre aktuelle Situation sowie Gedanken und Gefühle sprechen können. Aus den Sprachaufzeichnungen wurden dann ebenfalls die akustischen Stimmparameter und zusätzlich modernste sogenannte “word embeddings” aus dem gesprochenen Inhalt extrahiert.
Während die Stimmmerkmale lediglich signifikante Vorhersagen für Aktivation lieferten, war der Inhalt prädiktiv für emotionale Valenz (Zufriedenheit und Traurigkeit) und Aktivation. Diese Ergebnisse legen nahe, dass Affektvorhersagen anhand von Stimmparametern aus *freier* Rede (siehe Studie 1.2) denen aus *vorgegebenem* Sprachinhalt (siehe Studie 1.1) überlegen sind. In den trainierten Modellen zeigte der Sprachinhalt ebenfalls eine bessere Vorhersageleistung als die Stimmakustik der Stimme. Die experimentellen (Studie 1.1) und explorativen (Studie 1.2) Ergebnisse deuten außerdem darauf hin, dass der emotionale Gehalt des gesprochenen Sprachinhalts keinen Einfluss auf die Vorhersage von Affekterleben durch die Stimmakustik hat. Das bedeutet, dass der Inhalt keinen Einfluss darauf hat, wie gut der Affekt anhand von Stimmmerkmalen vorhergesagt werden kann. Darüber hinaus wurden Methoden des interpretierbaren maschinellen Lernens eingesetzt, um diejenigen Stimmmerkmale, in diesem Fall Lautstärken- und Spektralmerkmale, zu identifizieren, denen die größte Relevanz in den Vorhersagemodellen für Affekterleben aus Stimmparametern zukommt. Zuletzt wurden die Auswirkungen auf das algorithmische Monitoring von Affekterleben erörtert und Fragen zum Schutz der Rechte der Nutzer auf Privatsphäre diskutiert.

Studie 2 analysiert Sprachmuster in Textdaten, die über die Smartphone-Tastatur aufgezeichnet wurden, um zu untersuchen, ob und inwiefern sich Unterschiede zwischen Personen und Schwankungen innerhalb von Personen im Affekterleben in geschriebener Sprache über verschiedene Zeiträume und Kommunikationskontexte hinweg manifestieren und vorhersagbar sind. Aus einem Datensatz von mehr als 10 Millionen getippten Wörtern von 486 Studienteilnehmern wurden Merkmale bezüglich der Tippdynamik, der Wortverwendung auf der Grundlage von Wortkategorien (z.B. positive Emotionswörter) und der Verwendung von Emoji und Emoticon analysiert. Aus den Daten konnten eindeutige affektbezogene Sprachvariationen in verschiedenen Kommunikationskontexten (private Nachrichten gegenüber öffentlichen Posts) und Zeiträumen (gesamter Studienzeitraum, wöchentlich, täglich, im Moment) identifiziert werden. So korreliert beispielsweise die Verwendung von Wörtern in der ersten Person Singular (z.B., “ich”, “mir”) in öffentlicher Kommunikation, wie zum Beispiel in Posts auf sozialen Medien, deutlich stärker mit einem stabilen negativen Affekterleben als in privater Kommunikation, wie etwa in Chatnachrichten in WhatsApp. Die Vorhersagen von Affekterleben durch diese Textmerkmale mithilfe von maschinellem Lernen waren jedoch nicht signifikant besser als der Zufall. Die Ergebnisse zeigen, dass Methoden, die das Vorkommen von bestimmten Texteigenschaften zählen, wie etwa bestimmte Wortkategorien (z.B., positive Emotionswörter), um auf das subjektive Affekterlebnis zu schließen, insbesondere bei kleinen Zeitfenstern limitiert sind. Schließlich unterstreicht diese Studie die Möglichkeiten des Einsatzes von handelsüblichen Smartphones zur Erhebung und detaillierten Analyse von Textdaten aus dem Alltag.

Die vorliegende Dissertation zeigt auf, inwiefern Merkmale der gesprochenen und geschriebenen Sprache mit subjektivem Affekterleben in Verbindung stehen und dieses vorhersagen können. Im Gegensatz zu bisherigen Forschungsarbeiten, welche Text- oder Sprachdaten verwendeten, die von Probanden bezüglich ihres affektiven Gehaltes bewertet wurden oder von Schauspielern eingesprochen wurden, beruht die vorliegende Arbeit auf Smartphone-basierten Experience-Sampling Daten und zugehörigen Sprachproben. Damit gehören die beiden Studien, welche der vorliegenden Dissertation zugrunde liegen, zu den ersten Forschungsarbeiten, die gesprochene und geschriebene Alltagssprache mithilfe von handelsüblichen Smartphones über einen längeren Zeitraum erheben und wissenschaftlich untersuchen. Mit dieser hier angewandten Methode können Forscher Spuren natürlich verwendeter Sprache sammeln und analysieren. Hierzu hatten bislang nur Technologieunternehmen, die Text- und Sprachdaten durch Sprachassistenten verarbeiten, Zugang.

Die limitierten Vorhersagen unter Verwendung maschinellen Lernens in der vorliegenden Dissertation stellen die optimistischen Vorhersageleistungen früherer Forschungsarbeiten zur automatischen Affekterkennung, welche Sprache als aussagekräftigen Indikator für unser Affekterleben ansehen, in Frage. Außerdem zeigen die vorliegenden Ergebnisse, dass zwar die Form der Sprache, wie beispielsweise Stimmmerkmale und Tippdynamik, wertvolle Informationen über die emotionale Aktivation enthält, allerdings der Inhalt der gesprochenen und geschriebenen Sprache deutlich stärkere Assoziationen und Vorhersagen, insbesondere für die affektive Valenz, ermöglicht. Auch wurden im Vergleich zu vorheriger Forschung vermehrt unter Verwendung statistischer Methoden in den Bereichen Beschreibung, Vorhersage und Erklärung in dieser Dissertation auch spezifische affektbezogene Sprachmerkmale (z.B., bestimmte Wortkategorien oder Stimmparameter) analysiert. Letztlich unterstreichen die Ergebnisse dieser Arbeit die Relevanz des Kontextes der Sprachproduktion für Sprachmerkmale und entsprechende Affektvorhersagen. So ist es bei gesprochener Sprache von Relevanz, ob der Inhalt vorgegeben ist oder frei formuliert wird. Bei geschriebener Sprache hingegen ist es wesentlich, in welchem Kommunikationskontext und über welchen Zeitraum die Daten analysiert werden. Diese Erkenntnisse weisen auf die Notwendigkeit hin, dass Sprachmodelle zur Affekterkennung den Kontext berücksichtigen müssen und entsprechend trainiert und validiert werden müssen.

Zusammenfassend kann festgehalten werden, dass die Erkennung von subjektivem Affekterleben aus natürlich vorkommender Sprache, die mit Smartphones gesammelt werden kann, viel Potential und Zukunftsperspektiven sowohl für weitere Forschung als auch kommerzielle Anwendungen bietet. Insbesondere in den Bereichen der multimodalen Affekterkennung, welche auf einer Kombination von mehreren Datentypen (z.B., Sprache und physiologische Daten) basiert, Zeitreihenanalysen und idiografischen Modellen (Vorhersagemodelle für einzelne Personen) besteht noch reichlich Entwicklungspotenzial. Um aber dieses Potenzial voll ausschöpfen zu können, müssen zunächst die Herausforderungen einer präzisen Konzeptualisierung von Affekt, der Interdisziplinarität der Forschungseinheiten, der Ethik sowie des Datenschutzes und der Datensicherheit bewältigt werden. Wenn diese Herausforderungen überwunden werden können, stellt die Analyse natürlicher Sprache, welche durch Smartphone erhoben wird, ein vielversprechendes Instrument für Rückschlüsse zum subjektiven Affekterleben im Alltag (z.B., über das affektive Wohlbefinden) und für die Weiterentwicklung der affektiven Wissenschaften dar.

Chapter 1

Introduction

"Language is a mirror of the mind in a deep and significant sense." (Chomsky, 1975)

Humans use language - be it spoken or written - as the fundamental channel to communicate with each other. For example, we share ideas, coordinate with one another, and express how we feel through language. Thereby, language represents a dynamic window into the human mind and our day-to-day emotional experience (Jackson et al., 2021). When communicating with others, we are also intuitive language analysts: We constantly and automatically monitor, process, and interpret *what* and *how* others communicate (Scherer, 1986, 2003). Hereby, we process the plain content of the words one speaks or writes, but also monitor, for instance, the tone of the voice when talking to someone or read between the lines in text messages to infer emotional meaning.

Due to the intuitive association of language and our emotional experience, thinkers have made inferences about properties of the human mind from language for centuries. While the early works on voice-affect associations in rhetoric (e.g., by Cicero) and evolutionary biology (e.g., by Darwin) were anecdotal in nature (Darwin, 1886; May, 2001), the first half of the 20th century saw a surge in empirical evidence coming from psychologists who used electroacoustic analysis to analyze voice characteristics systematically (Fairbanks & Pronovost, 1939; Skinner, 1935). The study of language's content had also undergone an evolution during that time: Unscientific early analytical approaches focusing on revealing the contents of people's subconscious processes (Freud & Strachey, 1901) were steadily replaced by more sophisticated methodologies as more records of human language became accessible (Allport, 1942). This led to the development of the General Inquirer (Stone, Bales, Namenwirth, & Ogilvie, 1962), a computer system that could automatically count words from dictionaries in an input text file for psychological analysis.

With new technological means emerging over the course of the following decades of the 20th century, spoken and written language records could be transmitted, stored, and reproduced. After the turn to the 21st century, behavioral scientists have been facing rapid technological changes: an ever-growing amount of human-created text and speech data has become available to be analyzed through advanced computer software and algorithms with constantly increasing computational power (Iliev, Dehghani, & Sagi, 2015). These new methods have sparked a growing body of research investigating language variations based on individual differences in peoples' stable trait characteristics, such as age, gender, or personality, as well as fluctuating attributes, such as their affective states (Eichstaedt & Weidman, 2020; Kosinski, Stillwell, & Graepel, 2013). While confined to a niche group of psychologists in its early days, the scientific investigation of language has evolved into a research field spanning across disciplines, such as psychology, computer science, linguistics, and phonetics.

In recent years, affect, such as short-termed emotions and longer lasting moods, moved into researchers' focus due to its striking explanatory power for our thinking and behaving. In the same manner as behaviorism and cognitivism, some scientists even call for a new era of affectivism (Dukes et al., 2021). In line with this growing scientific interest in affect, research concerning the automated recognition of affect from language has also been evolving in the last two decades (Schuller, 2018). Hereby, researchers investigated associations of, for instance, trait affect (i.e., stable individual predispositions to experience certain affective states) and enduring mood disorders with language (Eichstaedt et al., 2018). Also, with more fine-grained longitudinal language data becoming available, research on momentary affective states (i.e., transient fluctuations in affect) and language also flourished. For instance, researchers have predicted momentary emotions from Facebook posts (Preoţiuc-Pietro et al., 2016). However, for this endeavor, researchers face the challenge of collecting data of momentary *subjective affect experience* and corresponding language data and, therefore, mostly rely on actors enacting given emotions or raters labeling speech or text samples.

Given the scientific progress in affect recognition from language and the rise of voice assistants, chat bots, and other text- or speech-based services, which generate vast amounts of text and speech data, gaining psychological insights into one's emotional life from language has also drawn increasing commercial attention. A range of tech companies, such as Amazon and Hume AI, are offering language-based emotion recognition tools leveraging artificial intelligence (AI) (Parthasarathy, Rozgic, Sun, &

Wang, 2019; Wiggers, 2022). These algorithms are often deployed to make important decisions, for example who is hired for a job, who is treated how in mental health care, or which product is launched. However, the inner workings of commercially applied algorithms and what data they had been trained on is usually not transparent to users and rarely shared with the research community, hindering investigations of the validity of their predictions. In the related domain of AI-based emotion recognition from facial expressions, scientific investigation revealed that the deployed tool can be inaccurate and biased (Barrett, 2022; Barrett, Adolphs, Marsella, Martinez, & Pollak, 2019) leading multiple tech companies to take down their facial emotion recognition services as a result (Hill, 2022).

The present dissertation investigates the associations of between-person differences and within-person fluctuation in self-reported subjective affect experience with and predictions from language characteristics from digital records of spoken and written language collected using off-the-shelf smartphones. The work is organized as follows: This introductory chapter lays the foundation to establish an understanding of the theory of affect, how affect is communicated through language, and the promises and challenges of affect recognition from language. Two empirical studies are described in detail in chapter 2 and 3. Chapter 4 finalizes the work by providing a general discussion of findings, contributions, limitations, and implications for theory and practice.

1.1 Affect Theory

Affect is the "mental counterpart of internal bodily sensations". Thereby, "affect" is a theory-neutral umbrella term that refers to anything emotional that is something's effect or internal state (Barrett & Bliss-Moreau, 2009). However, ever since Darwin's seminal work on emotion expression in human and animals (Darwin, 1886), there has been much scientific debate about how to conceptualize affect and, consequently, how to measure it. In the following, the two prominent theoretical frameworks of affect, namely discrete emotion categories and dimensional core affect, that are most prevalent in language-based affect research, are presented.

Categorical emotion approaches propose the existence of four to up to 22 fundamental emotions, for example happiness or anger (Ekman, 1992; Russell & Barrett, 1999). They are based on the hypothesis that instances of a set of basic emotions have unique (facial) expressions and exist across cultures and, therefore, are viewed as universal. It is assumed that unique programs of the autonomic nervous system ("fingerprints") constitute to these emotion categories (Siegel et al., 2018). These prototypical emotional categories involve complex biological processes and are assumed to be time restricted.

On the contrary, the concept of core affect suggests that we always experience some degree of elementary consciously accessible affective feelings (Russell & Barrett, 1999). Core affect can be mapped in a two-dimensional space with the dimensions of valence (i.e., pleasure) and arousal (i.e., physical and psychological activation) (Posner, Russell, & Peterson, 2005; Russell, 1980; Russell & Barrett, 1999). The aforementioned discrete prototypical emotions involve core affect and can be mapped onto this two-dimensional space (see Figure 1.1). The present dissertation is based on the concept of core affect since it allows to cover a broad range of affective experience at any given time.

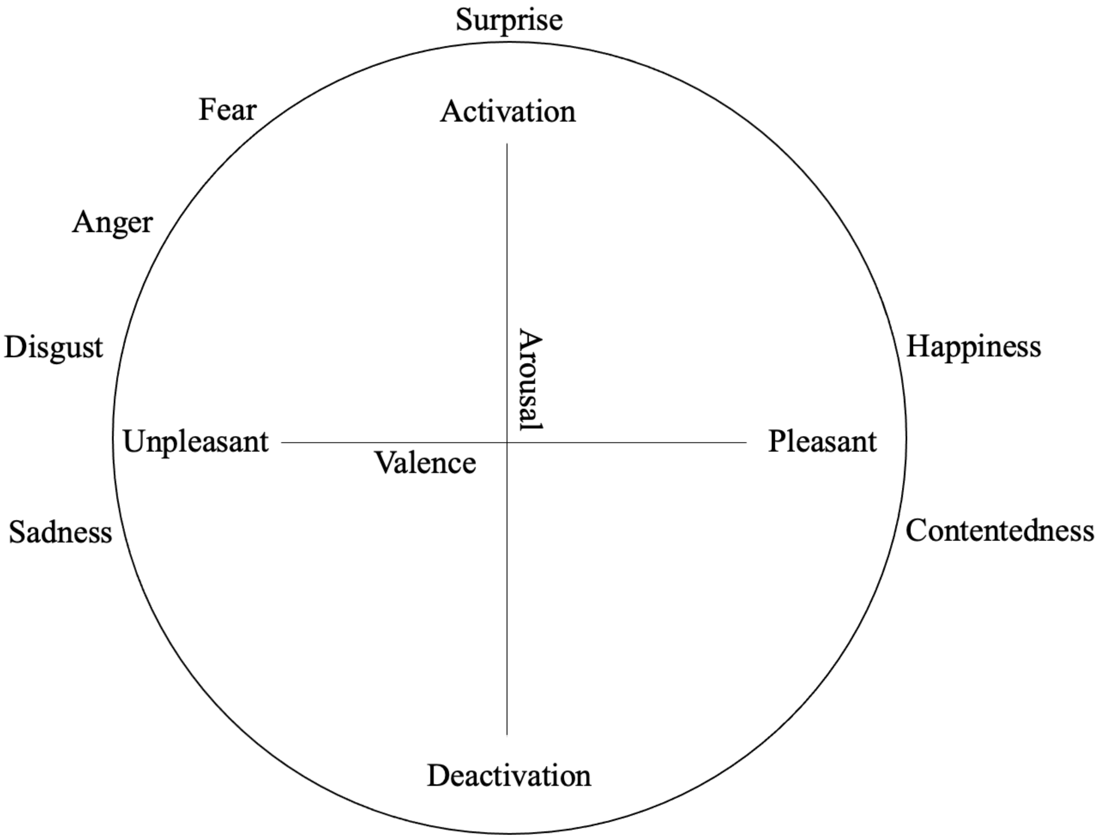


Figure 1.1: Circumplex model of affect. Depicted are the two core dimensions of valence and arousal along discrete emotion categories mapped onto the circumplex. The figure has been adapted from Russell (1999).

Table 1.1: Delimitation of different types of affect. Symbols (ranging from "0" to "+++") indicate to what the degree the features are present. Arrows indicate hypothetical ranges. This table has been adapted from Scherer (2000).

Affect	Intensity	Duration	Event focus	Rapidity of change	Behavioral impact
Emotion	++ —> +++	+	+++	+++	+++
Mood	+ —> ++	++	+	++	+
Trait	0 —> +	+++	0	0	+

Moreover, it is helpful to use distinct terminology to differentiate between different kinds of affect experience. In this manner, one can contrast different types of affect on a number of characteristics, such as their intensity, duration, event focus, rapidity of change, and behavioral impact (Scherer, 2000, 2003). Table 1.1 illustrates such a delimitation of different types of affect. Accordingly, emotions, moods, and trait affect, decrease in intensity, event focus, rapidity of change, and behavioral impact and increase in duration in the aforementioned order. In line with this, emotions, are short-lived and directed affective states that fluctuate rapidly over time. In contrast, trait affect is relatively stable over time and is not focused on a specific event.

1.2 Affective Language

Since humans are unable to directly access the internal affective states of one another, we have to make use of observable cues, for example facial expressions, body posture, and language, to build our own mental model of others' states. Historically, there had been a lot of focus on facial expressions, but more recently, the importance of language as an important means to communicate affect has moved into the research focus.

1.2.1 Channels in Spoken and Written Language

Language has two channels: *What* we communicate (i.e., the *content*) and *how* we communicate it (i.e., the *form*). Both channels convey valuable affective information (Scherer, 1986, 2003). The following section lays out how the distinct channels of content and form in spoken and written language as illustrated in Table 1.2 can be scientifically analyzed.

Table 1.2: Language channels in spoken and written language that can be used to communicate affectively with corresponding examples.

	Speech		Text	
	Spoken words	Voice acoustics	Written words	Typing dynamics
Example	word use	pitch, loudness	word use	typing speed

Spoken and Written Content

The words we write or speak transmit valuable information about our inner psychological workings (Boyd & Schwartz, 2021; Eichstaedt, Kern, et al., 2021; Kennedy, Ashokkumar, Boyd, & Dehghani, 2021). In order to quantify language content, predefined dictionaries, such as ones implemented in the Linguistic Inquiry and Word Count (LIWC) software, are often used (Boyd, Ashokkumar, Seraj, & Pennebaker, 2022). Hereby, words fall into different categories (e.g., first person singular, emotion words, family-related words) and the share of that category in the whole text is computed. With regard to affect, the use of positive (e.g., "happy") and negative emotion words (e.g., "sad") is most prominently studied (Kross et al., 2019). However, these dictionary-based approaches do not consider the context those words are used in. Advanced natural language processing (NLP) methods, such as topic or word embedding models, are able to capture those context effects (Eichstaedt, Kern, et al., 2021; Kennedy et al., 2021). A comprehensive overview of research on the associations of spoken language content and affect is provided in study 1 (see chapter 2) and on the associations of written language content and affect in study 2 (chapter 3).

Voice Acoustics

The voice signal represents a curve that can be quantified through different statistical measures that represent, for example pitch, loudness, or speaking rate (Vogt, André, & Wagner, 2008). There is software, such as OpenSMILE, available to automatically extract these voice characteristics from a speech record (Eyben, Weninger, Gross, & Schuller, 2013). A comprehensive overview of findings related to affect in voice acoustics can be found in study 1 (see chapter 2).

Typing Dynamics

There are many ways on *how* to write a piece of text: Rather quickly, without using many words, and correcting oneself multiple times or writing slowly and not correcting oneself at all. This kind of information is usually not visible to the communication partner, but it can be logged though a device's keyboard, for example from a computer or a smartphone. For instance, typing speed has been most often found to be related to affect (Ghosh, Ganguly, Mitra, & De, 2017). An exhaustive summary of findings related to affect in typing dynamics is provided in study 2 (see chapter 3).

1.3 Inferring Affect from Language

We intuitively use the channels of form and content in written and spoken language to communicate affectively and gain insights into others' emotional lives. Scientist in the field of affective computing have been working on deciphering this process to automatically infer affect from language (Picard, 2000). The ultimate goal is to automatically infer how people feel by utilizing those language cues in the same manner as a human would do, for instance, when monitoring the day-to-day affect experience of people suffering from mood disorders (Muaremi, Gravenhorst, Grünerbl, Arnrich, & Tröster, 2014). The timely recognition of affect from language also holds many promises in human-machine interaction, where the artificial agent could use the recognized affective information to craft an appropriate emotional response.

Inferring affect from language offers multiple advantages over the traditional psychological assessment via self-report questionnaires, particularly for fluctuating affective states. Researchers traditionally rely on self-reports by asking people verbally or through questionnaires to gain insights into people's behaviors, thoughts, and feelings. However, this comes with many caveats that can be overcome by using language instead. First, it eliminates the general downsides of questionnaire assessment, such as social desirability and response biases (Demetriou, Ozer, & Essau, 2015). Second, it offers an unobtrusive mean to track fluctuating states, such as affect, over time, which would otherwise require repeated assessment. Here, self-report methods are particularly impractical because reporting one's emotions can alter those emotions and the number of self-reports that one can complete in a given time period of interest is limited (Kassam & Mendes, 2013; Kuppens, Oravecz, & Tuerlinckx, 2010).

1.3.1 Data Collection

To associate affect with and infer it from language, researchers have to collect data on affect and corresponding language samples. Here, scientists have two options to gather these data: They can either collect language samples in a controlled setting in a laboratory, for instance by asking actors to enact a given emotion and recording them or inducing an emotion and letting participants write a piece of text, or in the wild (e.g., collecting language data when it is known that a person is in a specific affective state).

In the Lab

In the past, scientists mostly relied on language samples that had been collected in the lab because of the challenges associated with collecting language samples in the wild (e.g., getting language data from the exact moment a participant experiences a specific state in their everyday life). For example, if interested in traits, they would invite participants to the lab who would write stream of consciousness essays (Pennebaker & King, 1999). Alternatively, they would aim to elicit desired emotions, for instance by giving participants a task to complete, and record their speech (Batliner et al., 2004).

Another approach, that is mainly used to collect speech data, is to hire actors who are then instructed to enact predefined affective states and then create records of their language. For instance, researchers would tell them to read out a given sentence in different emotional modalities, such as happy, angry, and sad. Alternatively, scientists would use excerpts of emotion expressions from TV shows. Here, they would cut out those scenes that they believe represent a specific emotion and run their analyses on them.

While this lab approach allows to acquire data on very specific, salient, short-termed affective states, it comes with multiple caveats. First, due to the effort of recruiting participants or hiring actors the resulting data sets are comparably small. Second, the focus of these works can only be on short-lived affective states, such as emotions. With this approach, it is not possible to investigate lower intensity everyday moods. Third, when using actors to enact emotions, the desired affective state may not be authentically acted out as it may be driven by how the actor believes the respective emotion should be expressed and it remains unclear if the desired affect has been authentically experienced (Schuller, 2018). Fourth, replying on lab data also limits the external validity of findings. Finally, one can only obtain one-time language data in the lab which does not allow to investigate fluctuations in affect over time.

In the Wild

Alternatively, researchers use existing records of language that had been created “in the wild”, i.e., in natural everyday settings, such as diary entries (Tov, Ng, Lin, &

Qiu, 2013). With the advent of digital technologies, an increasing amount of digital language footprints have become available to researchers. Especially on social media, users produce vast amounts of textual data that can be analyzed for psychological insights (Kosinski, Matz, Gosling, Popov, & Stillwell, 2015). By donating their data for research, volunteers have helped create large data sets for scientific research, for example the well-known MyPersonality Facebook data set or chat corpora from instant messaging services (Koch, Romero, & Stachl, 2022; Verheijen & Stoop, 2016). This naturalistic approach allows to collect participants' language data over a period of time in order to also investigate within-person fluctuations in language, instead of only collecting one-time language samples analyzed in previous lab studies. Further, it provides a higher external validity than lab-created language samples since it contains language data from naturalistic settings.

When using naturally occurring language samples to investigate affect, researchers need to obtain corresponding *ground truth* data, i.e., the information that is assumed to be fundamentally true, on participants' affect experience. Therefore, they have two options: Asking participants themselves or hiring raters to assign affect labels to the collected language data. For trait affect or mood disorders that are assumed to be relatively stable, self-reports on the constructs of interest can be obtained before or after the language data collection period. For fluctuating states, however, researchers need affective self-reports from those exact moments the corresponding language, such as a social media post, had been produced. Asking participants in hindsight would yield imprecise results, since one could not accurately recall how they had felt when posting on Facebook a few weeks or months ago. Also, it would be very labor intensive for single participants to add self-reports to each of their posts. Therefore, researchers often rely on raters to label collected language samples, such as social media posts (Preoţiuc-Pietro et al., 2016). For the labeling, either research assistants or participants from online platforms, such as mturk, are recruited. The labels of multiple raters are then averaged out and used as the ground truth. As a consequence, the sample size of those labelled data sets is often limited due to the labor intensity of labeling and limited access to raters.

1.3.2 Statistical Modeling: From Associations to Predictions

The collected data on enacted, labelled, or self-reported affect and corresponding spoken or written language samples are then statistically analyzed. Traditional descriptive studies mostly utilized in-sample-associations (e.g., correlation coefficients) to investigate and describe how language is related to affect. For example, how voice pitch (Banse & Scherer, 1996) or the use of positive emotion words is related to affect (Tov et al., 2013). This approach had been sufficient when the number of language features is limited, and one is interested in specific linear associations.

With the collected data sets getting bigger and more dimensional (e.g., by leveraging digital footprints) in combination with the paradigm shift in psychological research from explanation to prediction (Yarkoni & Westfall, 2017), new predictive methods have moved into the focus. These new methods from the area of machine learning running on potent computational instances can handle large data sets with many cases and features. Further, they are able to detect complex interaction effects as well as non-linear associations. For example, machine learning has been used to predict affect from social media posts (Eichstaedt & Weidman, 2020; Jaidka et al., 2020). Machine learning models are often referred to as "black boxes", however there are new methods to gain insights from models and advance theory (Molnar, 2019; Shrestha, He, Puranam, & von Krogh, 2021).

1.3.3 Challenges

There are two main challenges in the research field concerning language and affect that are limiting prior work: First, being conceptually precise regarding the difference between affect experience and affect expression. Second, the empirical challenge of collecting data of affect experiences and timely corresponding language samples. After describing these two challenges in detail, novel mobile data collections methods using off-the-shelf smartphones are presented and how these can help researchers overcome the aforementioned challenges.

Conceptual: Affect Experience versus Affect Expression

Prior studies on affect recognition from language often use imprecise terminology and differ in the choice of emotion model and, as a consequence, make an aggregation and comparison of findings across studies challenging. For instance, one needs to differentiate if the research goal is to automatically recognize short-termed elicited emotions or longer lasting self-reported moods. Moreover, prior research in this field differs in the underlying theoretical framework: Some studies are based on the idea of core affect (Eichstaedt & Weidman, 2020; Preoţiuc-Pietro et al., 2016), while others focus on specific discrete emotions (J. Sun, Schwartz, Son, Kern, & Vazire, 2020). Based on these theoretical decisions, the operationalization (i.e., what items are used) differs, which in turn affects results.

Furthermore, there is a fundamental difference between what affect we express and what affect we actually experience, even though there is an overlap to a varying degree. Affect *expression* represents the emotional expressive behavior based on our internal affect *experience*. However, "feeling is not always revealing", i.e., one does not necessarily express what one experiences affectively or might even express something completely different (Gross & John, 1997; Gross, John, & Richards, 2000). Furthermore, the way one expresses affect might be perceived and interpreted differently. A helpful framework to illustrate the communication of affect through language is the tripartite emotion expression and perception (TEEP) model that is based on the Brunswick Lens model (Bänziger, Hosoya, & Scherer, 2015; Brunswik, 1956; Scherer, 2003; Vinciarelli & Mohammadi, 2014). An adaption of the TEEP to affective states is depicted in 1.2. In this framework, individuals express their experienced affective states (or parts of them) through distal cues (e.g., use of certain words, specific voice intonation) that can be objectively measured. During the transmission process of this information from the sender to the observer noise might be added which alters the information. Then, subjective proximal percepts (e.g., voice quality impressions) initiate the impression formation, where the observer creates a perceptual judgment of the affective state of the sender. Consequently, what affect one experiences and how an observer judges someone's affective state might be two different things.

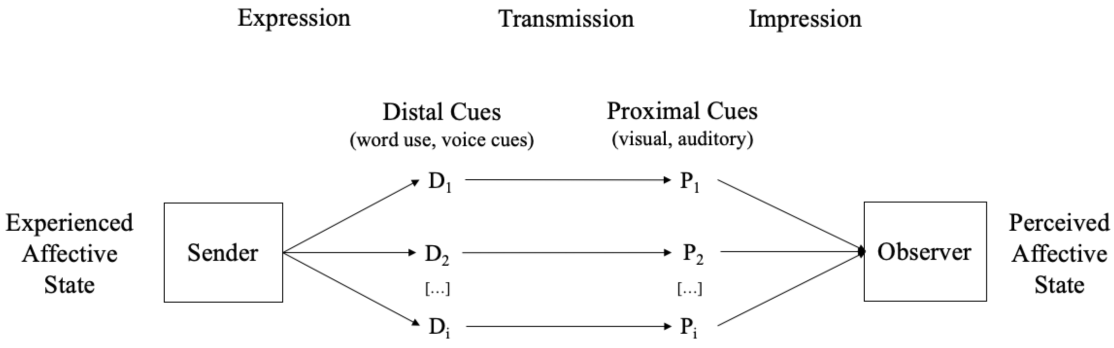


Figure 1.2: Brunswik lens model for the communication of affect experience through language

This important conceptual distinction between affect experience (i.e., what one truly experiences), affect expression (i.e., how one expresses), and how an observer perceives someone's affective state has consequences for the implications of research and deployed commercial algorithms trained on data from actors and raters: Those works have been focused on the automated recognition of affect *expression* instead of affect *experience*. Specifically, when actors are prompted to enact specific emotions and their distal cues (e.g., voice loudness) are recorded in order to train algorithms, it remains unclear if the actor truly experienced the target emotion. Further, how the target emotion is enacted is dependent on the mental model of the actor of how that specific emotion should be expressed. The desired state may not be authentically acted out and it may be driven by how the actor believes the respective emotion should be expressed (Schuller, 2018). Further, actors might not truly feel the targeted emotion and overact (Wilting, Krahmer, & Swerts, 2006). Similarly, when using raters to label others' affective expression, for example, speech samples or social media posts, there is an ambiguity of ground truth due to the subjective nature of labeling because raters tend to disagree to some extent as to what the affective state should be expressed in the language of others (Schuller, 2018). Therefore, these labels represent perception processes (perceived affect) rather than production processes (felt affect) (Schuller, Batliner, Steidl, & Seppi, 2011). Further, it remains unclear what the speaker or the author of a post experienced affectively when producing that piece of language since raters base their judgment on the proximal cues related to people's affect expression. Also, specifically for social media language, people might manage their self-image on social media leading to inaccurate labeling (Bazarova & Choi, 2014). Consequently, research that uses actors or raters is concerned with affect *expression*. Even though many of these studies are summarized under the term of *Automatic Affect Recognition* that targets the inference of affect experience using machine detectable distal cues, what they actually investigate is *Automatic Affect Perception* (algorithmically inferring affect expressions an observer attributes to a given individual from proximal cues). However, it is the subjective affect experience - what affect a person is truly experiencing in a particular moment - with its challenges in assessment (see chapter 4) that is of high relevance for research and applied science. Models trained on expressed affect contain some information for the recognition of affect experience as affect expression and experience overlap to a varying degree (Kross et al., 2019), but it remains unclear if findings are truly generalizable. For the application of those algorithms, such as in mental health care, it would be important to have accurate models for the recognition of subjective affect experience.

Empirical: Timely Pairings of Affect Experience and Language Data

In order to investigate affective language, researchers need relevant data. For stable trait affect or affective disorders, such as depression, participants' language is usually aggregated across a given time period and then associated with self-reports. Past research has, for instance, followed this approach and predicted self-reported depression scores from social media text (Eichstaedt et al., 2018). However, if the goal is to investigate the language of fluctuating affective states, one needs data on the affective state just from that moment the person has produced a text or speech sample. Since one cannot ask people in hindsight what their affective state was at a given point in time, raters are often used to label those samples, like social media posts (Eichstaedt & Weidman, 2020; Preoţiuc-Pietro et al., 2016). As a consequence, through this labeling approach, researchers obtain data on affect expression, but not on what affect people actually experienced during language production (see previous section on the conceptual difference of affect experience and affect expression). Moreover, specifically for social media data, there are often gaps in the data steam when users do not post for a while, leaving scientists blind for that time period, even though not posting might be an important indicator for one's emotional condition itself.

The Promises of Mobile Data Collection

In order to overcome these challenges, researchers need a tool that allows them to collect in-situ records of self-reported affect experience and corresponding language samples. Recent technological progress, particularly in smartphones, has equipped researchers with new research tools to collect both, in-situ self-reports on subjective affect experience and corresponding language data in large quantities, in the wild.

The Electronically Activated Recorder (EAR) (Mehl, 2017; Mehl, Pennebaker, Crow, Dabbs, & Price, 2001) is a small recording device that is attached to one's clothing and takes audio records of participants' everyday lives in predefined intervals. After retrieving the files from the EAR, the records can be analyzed with regard to language form and content. For textual analysis, the records must be transcribed first, either manually or using a transcription algorithm. The EAR has the advantage, that records might also contain valuable information about the environments participants spent time in (e.g., sitting in a cafe). Further, it allows to collect longitudinal data to study not just between-person differences but also within-person fluctuations. Multiple studies have used the EAR to collect speech samples to associate them with affect experience (J. Sun et al., 2020; Weidman et al., 2020). However, using the EAR comes with privacy challenges because other people might be recorded without their knowledge and consent. Plus, managing the device distribution and transcribing the recorders is laborious. Finally, the records can contain much noise and little speech even though the EAR could be set in a way that it only records human speech

(Lazarevic, Bjekic, Zivanovic, & Knezevic, 2020).

Another emerging mean to collect language data is to use off-the-shelf smartphones. Here, recent advancements in leveraging mobile sensing technology, i.e., the built in sensors and logs, have been used to create records of everyday behaviors (Ferreira, Kostakos, & Dey, 2015; Harari et al., 2016). In order to collect spoken language, participants are usually asked to make an intended record of their voice using the smartphone's microphone (Marrero, Gosling, Pennebaker, & Harari, 2022; Petrizzo & Popolo, 2021). Another option is to eavesdrop on calls or randomly turn on the smartphone's microphone, but this approach comes with serious privacy challenges (Faurholt-Jepsen et al., 2016; Muaremi et al., 2014; Wang et al., 2014). In order to collect written language, either a special keyboard is used to capture what participants had typed into their phone (Bemmann & Buschek, 2020; Buschek, Bisinger, & Alt, 2018) or screenshots are taken regularly (Brinberg et al., 2021). Recent studies have leveraged these tools to collect language data and associate them with affect (Carlier et al., 2022; Tony Liu et al., 2021). These language footprints can also be combined with other mobile sensing data (e.g., app use, GPS data) that can also provide valuable insights into the contexts when language had been produced.

Off-the-shelf smartphones also represent a useful platform to collect self-reports on participants' momentary affective states in an ecologically valid way over a period of time by using the Experience Sampling Method (ESM) (Bolger & Laurenceau, 2013; Csikszentmihalyi & Larson, 2014). ESM is a method to collect participants' self-reports on their activities, emotions, and other situational variables. It is seen as the gold standard for collecting data on in-vivo experiences (Conner, Tennen, Fleeson, & Barrett, 2009). Hereby, participants are asked multiple times per day to respond to a short questionnaire on how they feel. Smartphones represent a fruitful medium to administer ESM assessments (Van Berkel, Ferreira, & Kostakos, 2017). Thereby, information on one's affect experience can be collected in situ during normal everyday life. Multiple recent studies have used the ESM on mobile devices to collect data on momentary affect experience and associate it with corresponding spoken or written language data (Kross et al., 2019; J. Sun et al., 2020; Weidman et al., 2020).

1.4 The Present Dissertation

Digital traces of natural language data have become ubiquitous. As a results, scientific and commercial interest to leverage these new data to infer peoples' affect is on the rise. Due to the challenges associated with collecting data on subjective affect experience and corresponding language samples, previous research studies and commercial products have mostly relied on actors or labelled data, and, thereby, are focused on affect expression. The present dissertation leverages new smartphone-based data collection methods to collect self-reports on in-situ subjective affect experience and corresponding language samples in the wild to investigate between-person differences and within-person fluctuations in affect experience. In doing so, the present work aims to achieve the following three research goals.

1.4.1 Research Goals

First, the present dissertation investigates if between-person differences and within-person fluctuations in subjective affect experience are associated with and predictable from cues in spoken and written natural language. Therefore, traditional descriptive statistical methods (e.g., correlation analyses) and state-of-the-art predictive machine learning algorithms are employed.

The second goal is to identify specific language characteristics, such as the use of specific word categories or voice parameters, that are associated with and predictive for between-person differences and within-person fluctuations in affect experience. Again, descriptive statistics (e.g., correlation coefficients) and tools from the field of interpretable machine learning (e.g., feature importance) are combined.

Third, the present dissertation analyzes the influence of the context of language production on the associations with and predictions of affect experience from natural language. Specifically, for spoken language, the impact of the degree of freedom of the content being spoken about (predefined sentences to read aloud versus prompted free speech) and the emotional sentiment of the content on affect recognition from speech are analyzed. With regard to written language, the influence of the time window (trait, weekly, daily, momentary) language has been aggregated on and the communication channel (private messaging versus public posting) on affect recognition is investigated.

1.4.2 Overview of the Studies

This dissertation is comprised of two empirical studies that contribute to the aforementioned three research goals. Both studies couple self-reports on subjective affect experience with natural language data collected with smartphones and draw on data collected in a representative smartphone sensing panel study conducted in 2020 in Germany (Schoedel & Oldemeier, 2020).

Study 1 investigates between-person differences and within-person fluctuations in subjective momentary affect experience in over 23000 speech samples collected with smartphones in two data sets from Germany and the United States. In the predictive models using machine learning, the predictive power of voice acoustics and state-of-the-art word embedding in the prediction of momentary subjective affect experience is considered separately and combined. Also, those voice features that are most strongly associated with and predictive of affect experience are identified using descriptive measures and interpretable machine learning methods. Further, the influence of having participants read out aloud predefined sentences or letting them talk freely about their current situation and feelings on affect predictions is analyzed. Finally, the effect of the emotional sentiment of the spoken content on affect prediction from voice cues is investigated.

In Study 2, patterns in written language data logged through smartphones' keyboards are used to investigate how between-person differences and within-person fluctuations in affect experience manifest in and are predictable from logged text data across different time frames (trait, weekly, daily, momentary) and communication contexts (private messaging versus public posting). From the keyboard logs, general typing characteristics, language use captured with different (sentiment) dictionaries, and metrics on emoji and emoticon use are extracted for descriptive and predictive analyses. Finally, those text features that are most strongly associated with affect experience in varying contexts are identified.

The research conducted as part of this dissertation adheres to open science principles (Kathawalla, Silverstein, & Syed, 2021). Both studies had been pre-registered before analyzing the data and the relevant study materials (e.g., aggregated data sets, code scripts) will be made available openly on the Open Science Framework (OSF) once they are accepted for publication.

Chapter 2

Affect Experience in Speech

2.1 Abstract

Advances in the area of artificial intelligence (AI) and the ubiquity of speech data, for example coming from voice assistants, have created numerous commercial products that claim to be able to automatically recognize emotions from human speech. However, the employed algorithms have often been trained solely on enacted or labelled speech samples from artificial lab settings representing affect *expression* and are used to infer everyday subjective affect *experience*. In the present study, we investigate if machine learning algorithms can truly recognize subjective affect experience from speech samples collected in the wild. In two studies, we extract acoustic voice parameters and state-of-the-art word embeddings from 23632 speech samples with corresponding experience-sampled self-reports on momentary affect experience from 1066 participants collected using off-the-shelf smartphones. While voice acoustics provide limited predictive information of affective arousal, speech content is predictive of arousal as well as valence (sadness and contentedness). Further, experimental and explorative findings suggest that emotional speech content does not affect predictions from voice acoustics (i.e., what someone talks about does not affect how well emotions can be predicted from voice cues alone). We discuss implications for the algorithmic monitoring of affect experience from speech in everyday life.

2.2 Introduction

Research findings on the algorithmic recognition of affective states (e.g., emotions) and related affective disorders from speech offer promising applications, for instance in health care, human-machine interaction, education, and marketing (Hildebrand et al., 2020; Milling, Pokorny, Bartl-Pokorny, & Schuller, 2022). The advances in algorithmic affect recognition from speech leveraging AI and the ubiquity of available speech data due to the rise of voice assistants, for example Amazon's Alexa and Apple's Siri, have created an increasing commercial interest in the field. Here, tech companies aim to leverage speech data to, for instance, recognize what momentary affect their customers experience in order to develop personalized user interfaces or make product recommendations (Knight, 2016; Mandell, 2020; Vlahos, 2019). Most of the prediction models used in research and in the corresponding commercial tools are trained on enacted or labelled speech samples from artificial lab settings that represent emotion *expressions*. However, those algorithms are often deployed to detect people's subjective affect *experience* in everyday life. Further, many of the commercial algorithms are not transparent with regard to how well their predictions work and how (biased) predictions are being made. These issues raise questions regarding the promises of emotion-detecting speech technology and the protection of user privacy in setting where speech data can be analyzed, for example, when using voice assistants. The present work investigates the algorithmic recognition of between-person differences and within-person fluctuations in subjective self-reported momentary affect *experience* from speech samples collected with smartphones.

2.2.1 Predicting Affect from Speech

Researchers have successfully predicted affective states from a range of speech data, such as labelled TV clips (Grimm, Kroschel, & Narayanan, 2008), phone calls, and enacted speech samples from the lab (Bänziger, Mortillaro, & Scherer, 2012; Burkhardt, Paeschke, Rolfes, Sendlmeier, & Weiss, 2005; Schuller, 2018; Vogt et al., 2008). They report on impressive prediction performances for the automatic recognition of emotions (i.e., correlations between true scores and predicted scores of up to .81 for arousal and .68 for valence predictions) (Weninger, Eyben, Schuller, Mortillaro, & Scherer, 2013). However, one has to keep in mind that in those works the enacted target emotion or the rater labels serve as ground truth. Thereby, these works predict affect *expression*, which is considered easier to algorithmically recognize than real-life *experienced* affect (Vogt et al., 2008). Moreover, the prediction performance varies greatly across studies due to a varying choice of emotion targets (i.e., discrete emotion versus core affect), conceptualizations of affect (e.g., short-termed elicited emotions versus moods), and employed algorithms (e.g., supervised versus unsupervised machine learning).

Also, these prior studies on algorithmic affect recognition often offer no insights into how predictions in their "black box" models were being made. For instance, it frequently remains unclear which specific speech characteristics are particularly predictive of a given affective state. Prior descriptive research reported on associations of specific acoustic features and affective states. For example, voice pitch and intensity were found to be associated with affective arousal (Vogt et al., 2008; Weninger et al., 2013). Two recent studies provide a remarkable non-technical summary of voice features (Hildebrand et al., 2020) and a comprehensive overview of associations of word use with affect in spoken language (J. Sun et al., 2020). Recent developments in the area of interpretable machine learning can help gain insights into the inner working of machine learning algorithms and, consequently, aid with detecting speech features that are especially predictive of affective states (Molnar, 2019).

Due to the challenge of obtaining speech data with corresponding affect labels in-vivo, most prior research on affect recognition from speech has used actors or labelled samples. This comes with a set of downsides, such as actors potentially overacting and the ambiguity of ground truth due to the subjective nature of labeling (see section 1.3.3) (Batliner et al., 2011; Schuller, 2018; Wilting et al., 2006). As a consequence, studies investigating predictions of subjective affect experience from speech are rare. Recent works have collected everyday speech samples using the Electronically Activated Recorder (EAR) (Mehl, 2017). Hereby, speech data can be collected over a period of time which allows researchers to not only investigate between-person differences in affect (i.e., is this person sad?), but also assess within-person fluctuations (i.e., is this person sadder than other days?) (Huang & Epps, 2018; J. Sun et al., 2020; Weidman et al., 2020). Using the EAR, however, can be privacy invasive since potentially non-consenting persons may be recorded, too. Moreover, handling the EAR recorders and transmitting the collected data can be tedious for participants and researchers. Here, off-the-shelf smartphones represent a useful platform to collect experience samples on momentary affect experience over time and make corresponding speech records using the build-in microphone (Carlier et al., 2022; J. Sun et al., 2020; Weidman et al., 2020).

2.2.2 Content-Form Interactions in Affect Recognition from Speech

Prior research has shown that voice form (prosody) and the lexical content of the produced words (semantics) work together when transmitting affective information through speech (Ben-David, Multani, Shakuf, Rudzicz, & van Lieshout, 2016). Moreover, studies suggest that there is a prosodic (i.e., from voice cues) dominance in the perception of affect based on lab experiments (Ben-David et al., 2016; Lin, Ding, & Zhang, 2020), but not (yet) using speech data from the wild (Schwartz & Pell, 2012). Moreover, while this research field had focused on the interplay of prosody and semantics in the recognition of affect by human raters, there are, to our knowledge, no studies on content-form interactions in algorithmic affect detection. Hence, it is currently unclear if what users talk about (i.e., the emotional content) has an effect on voice acoustics that impact automated affect recognition. In an applied setting, for example, the question is if an algorithm could recognize affective states regardless of what the person talks about, may it be a mundane topic, such as the weather or ordering pizza, or does one need to talk about an emotional topic (e.g., meeting a loved one).

The present work leverages methodological advances in the area of smartphone-based data collection methods to investigate the prediction of between-person differences and within-person fluctuations in subjective momentary affect experience from speech. In two large-scale studies, we train cross-validated machine learning models on acoustic voice cues and state-of-the-art word embeddings from speech samples collected in the wild. Moreover, for predictive models, we investigate which voice cues are most predictive. Further, we experimentally and exploratively investigate the effects of the emotionality of speech's content on algorithmic affect recognition from voice acoustics. Thereby, we aim to inform potential applications and promises in automatic affect recognition from speech signals and advise the discussion on the protection of user privacy rights.

2.3 Study 1.1

2.3.1 Method

Smartphone-Based Voice Data Collection and Privacy-Preserving On-Device Acoustic Feature Extraction

Data collection for this study was part of a large six-month panel study (from May until November 2020) using the *PhoneStudy* research app at Ludwig-Maximilian-Universität München (Schoedel & Oldemeier, 2020). Data collection was approved by the responsible IRB board. The study comprised two two-week experience sampling phases (July 27, 2020, to August 9,2020; September 21, 2020, to October 4, 2020) during which participants received two to four short questionnaires per day. Here, self-reported valence and arousal were assessed in two separate items on six-point Likert scales among other psychological properties as part of an experience sampling procedure. Furthermore, for each experience sampling instance, we computed the fluctuation of assessed momentary affect in valence and arousal from one's (median) affect baseline (for participants with at the five experience sampling days) across all experience sampling instances. For example, if a participant had a valence baseline of "3" and reports a "6" in a particular moment, this fluctuation score of "+3" indicated that this person had been a lot more happy than usual.

The last experience sampling questionnaire of each day included an additional instruction, where participants were asked to read out a series of predefined emotional sentences while making an audio recording of their voice. The sentences presented to the participants are based on a set of validated German neutral and affective sentences (Defren et al., 2018) and differ in their emotional content: positive (e.g., "My team won yesterday."), negative (e.g., "Nobody is interested in my life."), and neutral (e.g., "The plate is on the round table."). These three emotional categories are presented consecutively in each audio logging task. The order of the categories was randomized per experience sampling questionnaire. For each emotional content category, three sentences were randomly drawn from respective sets of sentences in the database. The use and experimental manipulation of these emotional semantic categories allowed us to control for the content spoken by our participants and at the same time enabled us to conduct a privacy-friendly study. The audio recording was started by the participants via a button on the screen. Participants could stop the recording manually after a minimum of four seconds. Alternatively, the recording was stopped automatically after twelve seconds. We chose these lower- and upper-time thresholds because this is the minimum and maximum time required to read out the three sentences per condition. Once the audio record had been completed, we used the widely adopted *OpenSMILE* open-source algorithm (Eyben, Wöllmer, & Schuller, 2010) to automatically extract acoustic features directly on the participant's device. Here, we used the extended Geneva Minimalistic Acoustic Parameters Set (eGeMAPS) that is comprised of 88 acoustic features (Eyben et al., 2016). Theses voice feature have been used in a range of prior studies on affect recognition from speech. After feature extraction, the voice records were automatically deleted and only extracted voice features were stored on our servers.

With this procedure, we collected 11199 audio logs from 586 participants. We excluded 214 voice logs because the respective acoustic features (mean voicing score, voiced segments per second, mean voiced segment length) indicated that no human voice was recorded. Moreover, we excluded a total of 997 samples without corresponding self-reports on valence and arousal, from participants with less than five experience sampling days, and those participants who had no variance in all their valence and arousal scores across all their experience samples. This left us with a final data set of 9908 voice samples with corresponding acoustic features from 3381 experience sampling instances for valence and arousal from 499 participants (48.5% female, M(Age) = 42.97 years). Overall self-reported valence was positive ($M = 4.72$, $SD = 1.03$) and overall arousal was slightly geared towards activity ($M = 3.68$, $SD = 1.35$). The distribution of valence and arousal as well as fluctuations from the baseline is provided in Figure 2.7 this chapter's appendix in section 2.10.

In the final sample, voice samples were not equally distributed across emotional sentence conditions ($chi^2(2) = 36.48$, $p < .001$): 3219 came from the positive condition, 3110 from the neutral condition, and 3579 from the negative condition.

Predictive Modelling

We trained multiple supervised machine learning models on the extracted acoustic voice features for the prediction of self-reported valence and arousal and their fluctuations. Here, we compared the predictive performance of linear regularized regression models (LASSO) (Zou & Hastie, 2005) with those of a non-linear tree-based Random Forest model (Breiman, 2001; Wright & Ziegler, 2017), and a baseline model. The baseline model would predict the respective mean values for valence and arousal of the respective training set for all cases in a test set. Additionally, we included the prediction of participants' age and gender as a benchmark. Models were evaluated using a ten-fold cross-validation scheme (Bischl, Mersmann, Trautmann, & Weihs, 2012). We blocked participants in the resampling procedure ensuring that for each train/test set pair the given participant is either in the training set or in the test set.

We evaluated the predictive performance of the models based on the coefficient of determination (R^2) and Spearman's (rank) correlation (r) between the predicted scores and participants' self-reported scores. To determine whether a model was predictive beyond chance ($alpha = 0.05$), we carried out variance-corrected (one-sided) t-tests comparing the R^2 measures of all prediction models with those of the baseline models (Nadeau & Bengio, 2003). We adjusted for multiple comparison ($n = 8$) via Holm correction.

All data processing and statistical analyses in this work were performed with the statistical software R version 4.1.1 (R Core Team, 2021). For machine learning, we used the *mlr3* framework (Lang et al., 2019). Specifically, we used the *glmnet* (Friedman, Hastie, & Tibshirani, 2010) and *ranger* (Wright & Ziegler, 2017) packages to fit prediction models. To quantify the impact of single predictors in Random Forest prediction models, we computed (out-of-bag) permutation feature importance using the *DALEX* package (Biecek, 2018; Wright, Ziegler, & König, 2016). We preregistered the present study as a transparent account of our work (Koch & Schoedel, 2021).

2.3.2 Results

Recognizing Affect Experience from Acoustic Voice Cues

Overall, none of the employed machine learning algorithms predicted affect experience on the dimensions of valence and arousal significantly better than chance, even though, on average, predictions of raw arousal scores ($R^2 = 0$, $r = 0.13$) and arousal fluctuations ($R^2 = 0.01$, r $= 0.12$) from Random Forest models were slightly better than the baseline models' predictions. On the contrary, valence predictions did not yield any predictive information. Figure 2.1 provides an overview of the performance of all employed machine learning algorithms across prediction tasks. In the arousal prediction tasks, Random Forest models performed slightly better than LASSO models, suggesting non-linear relationships between voice cues and affective arousal outcomes in the present study. Finally, benchmark predictions of speaker age ($R^2 = 0.11$, $r = 0.39$) and gender (prediction accuracy = 91.4%) suggest that voice acoustics from the collected read-out sentences in study 1.1 contain valuable information about speaker demographics.

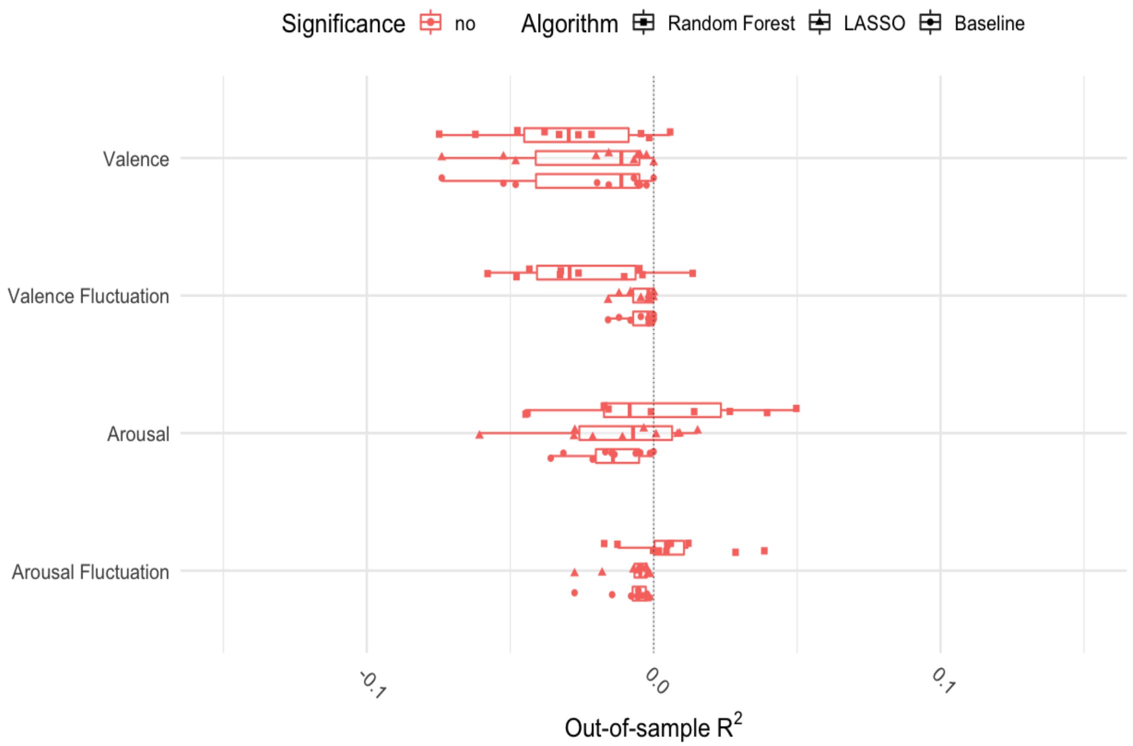


Figure 2.1: Box and whisker plot of prediction performance measures from 10-fold cross-validation for prediction models.

Figure 2.2 shows the five most important features in the Random Forest model (based on permutation feature importance) for the prediction of arousal from acoustic voice features. Overall, features related to spectral flux (i.e., how quickly the power spectrum of the voice signal is changing) and loudness were most important. This observation is in line with descriptive correlations of voice features and self-reported affect experience (see figures 2.9 and 2.10 in section 2.10 in the chapter's appendix), where spectral flux and loudness features also had the highest (Spearman) correlation coefficients. Together, these findings indicate that a louder voice that has a quickly changing spectrum is indicative of the experience of heightened arousal.

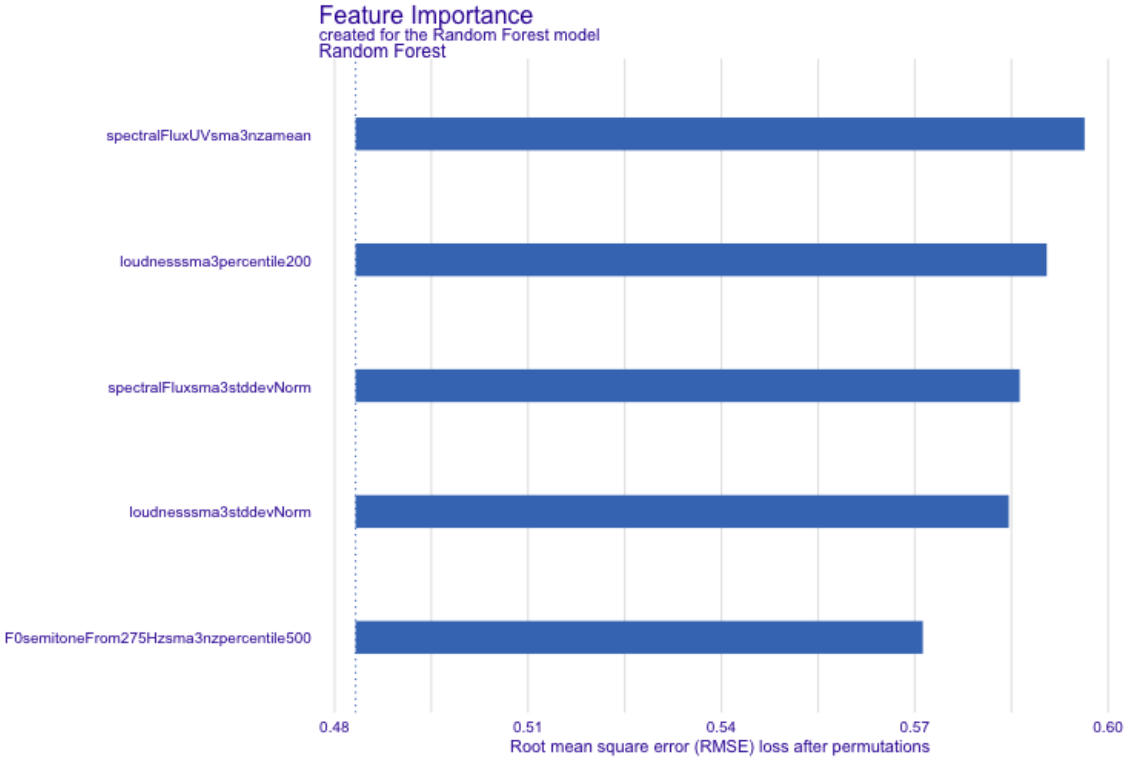


Figure 2.2: Permutation feature importance for the five most predictive features in the Random Forest model for the prediction of arousal. Permutation feature importance represents the decrease in the model's prediction performance (as measured by RMSE) after permuting a single variable.

Content Effects on Affect Predictions from Voice Acoustics

Finally, we analyzed if the experimentally altered emotional content (positive/ neutral/ negative) of the predefined sentences that had been read out by participants had an effect on affect predictions from voice acoustics. Here, we focused on predictions of between-person differences in valence and arousal since these showed a better, even though not significant, prediction performance than those models for within-person fluctuations. There were no significant differences in prediction errors across the three sentence conditions for valence ($F(2,11873) = 0.09$, $p > .05$) and arousal ($F(2,11873) = 0.39$, $p > .05$) predictions suggesting that sentences' emotional valence did not influence affect predictions from voice acoustics (see Figure 2.3).

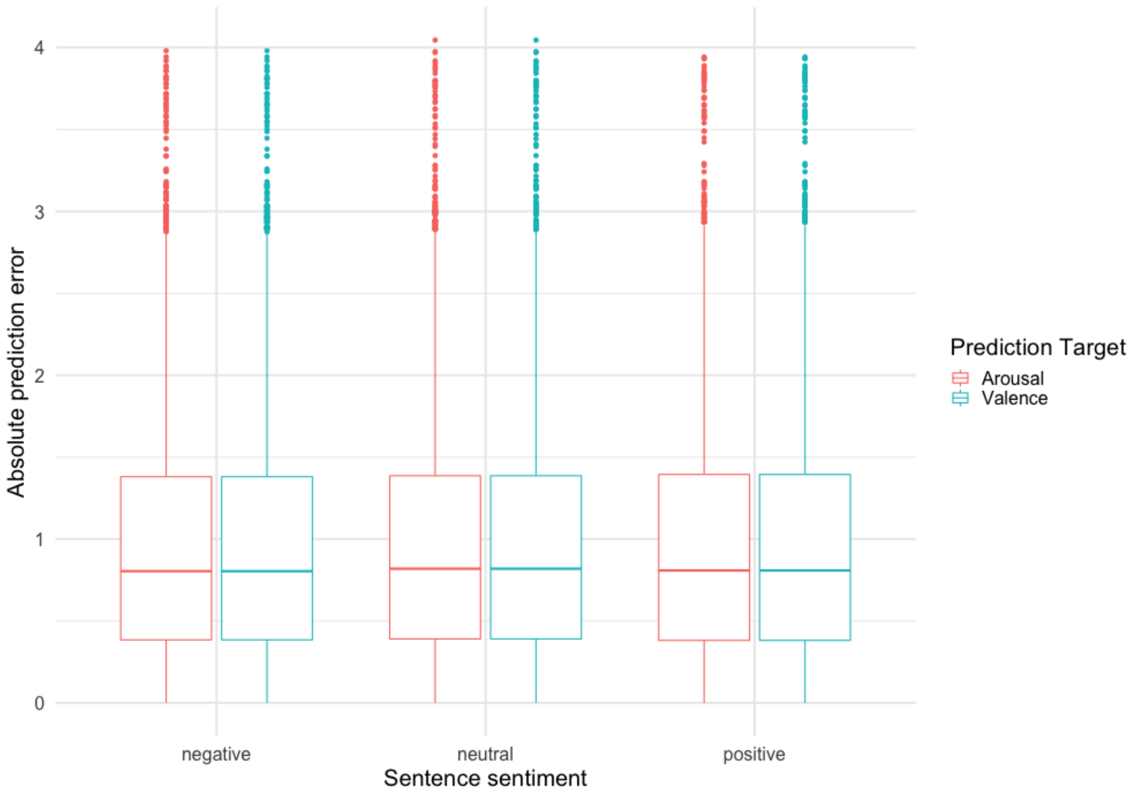


Figure 2.3: Prediction error for valence and arousal prediction in Random Forest models from voice acoustics in different emotional sentence conditions.

2.4 Study 1.2

2.4.1 Method

Smartphone-Based Speech Data Collection

Data collection for this study was part of the UT1000 Project at the University of Texas at Austin in the United States in fall 2018 (Wu et al., 2021). During a three-week self-tracking assignment using their smartphones, students from a Synchronous Massive Online Course (SMOC) in introductory psychology received four short experience sampling questionnaires per day where they could also make records of their speech at the end. Here, self-reported arousal (assessed on a five-point Likert scale), contentedness, and sadness were assessed in separate items on four-point Likert scales among other psychological properties as part of an experience sampling procedure. Thereby, in study 1.2, we captured emotional valence on two items (contentedness and sadness) instead of one as done in study 1.1. According to the affect grid (see Figure 1.1), contentedness and sadness have a comparable low level of arousal and an opposing emotional valence. Furthermore, for each experience sampling instance, in line with study 1.1., we computed the fluctuation of assessed momentary affect in arousal, contentedness, and sadness from one's (median) affect baseline (for participants with at least five experience sampling days) across all experience sampling instances. For the audio records, participants received the following instruction: "Please record a short audio clip in which you describe the situation you are in, and your current thoughts and feelings. Collect about 10 seconds of environmental sound after the description." The responses to this prompt are analyzed in the present study. Any parts of the record that did not contain speech were cut out before further analysis since the focus of this work is affect in human speech. The collected speech samples had also been used in another research project that describes the data collection procedure in more detail (Marrero et al., 2022).

With this procedure, we collected 23482 audio logs from 980 participants. We followed the same procedure to select speech records as in study 1.1 and, to ensure comparability of the two studies with regard to the length of speech samples, we retained all speech transcripts that contained at least 15 words and were more than four seconds long which is equivalent to the length of the sentences that had been read out in study 1.1: We removed records where the respective acoustic features indicated that no human voice was recorded. Moreover, we excluded audio samples without corresponding affect self-reports, participants with less than five experience sampling days, and those participants who had no variance in all their valence and arousal scores across all their experience samples.

This procedure left us with a final data set of 13724 speech samples with corresponding experience-sampled self-reports on momentary affect experience from 567 participants (64.9% female, M(Age) = 18.57 years). Overall participants reported balanced experienced contentedness (M = 1.65, *SD* = 0.85) and low sadness (M = 0.53, *SD* = 0.77). Overall arousal was balanced out (M = 1.95, *SD* = 0.95). The distribution of arousal, contentedness, and sadness as well as respective fluctuations from the baseline is provided in figure 2.8 in the chapter's appendix in section 2.10.

In the same manner as in study 1.1, we extracted the extended Geneva Minimalistic Acoustic Parameters Set (eGeMAPS) from the collected audio files using the OpenSMILE algorithm (Eyben et al., 2016, 2010). In study 1.2, those features were extracted from the raw recorded audio files after data collection and not directly on participants' smartphones as in study 1.1.

We transcribed all raw audio records using the Google Speech-to-text API. Then, we extracted state-of-the-art word embeddings from speech transcripts using the *text* R package (Kjell, Giorgi, & Schwartz, 2021). Word embeddings are vector representations of words in a high-dimensional space, which capture their meaning and relationships with other words. Specifically, for predictive modeling, we used the second to last layer (layer 23) from the language model "RoBERTa large" as recommended in prior work (Y. Liu et al., 2019; Matero, Hung, & Schwartz, 2022).

Predictive Modelling

For predictive modelling, we applied the same machine learning pipeline using the same R packages as used in study 1.1 to predict self-reported sadness, contentedness, and arousal as well as their deviations from the respective person's baseline levels. Moreover, in addition to extracted acoustic features, we also used extracted word embeddings as features. To investigate and compare the predictive power of speech form (voice cues) and content (word embeddings), we ran predictions on all features combined as well as acoustic features and word embeddings separately.

2.4.2 Results

Prediction of Affect Experience from Speech

The employed machine learning models trained on voice acoustics (speech form) and word embeddings (speech content) predicted between-person differences and within-person fluctuations in the subjective experience of momentary affect experience significantly better than chance. Figure 2.4 provides an overview of the performance of all learners across prediction tasks while in this section we report on the best performing algorithm respectively (either Random Forest or LASSO). Our models yielded the best prediction performance for between-person variations in contentedness ($R^2 = 0.1$, $r = 0.34$), arousal ($R^2 = 0.09$, $r = 0.32$), and sadness ($R^2 = 0.04$, $r = 0.24$). Also, for within-person fluctuations, predictions were significantly better than chance for contentedness ($R^2 = 0.06$, $r = 0.26$), arousal ($R^2 = 0.05$, $r = 0.22$), and sadness ($R^2 = 0.02$, $r = 0.13$). However, overall, predictions were better for between-person differences than for within-person fluctuations.

Moreover, evaluation prediction performance of models trained only on voice acoustics or word embedding respectively revealed that predictions were mostly driven by the information coming from speech content as represented in the word embeddings. Prediction models trained on voice acoustics alone were not significantly better than chance. However, on average, predictions of between-person differences in contentedness ($R^2 = 0.01$, $r = 0.14$) and arousal ($R^2 = 0$, $r = 0.12$) as well as arousal fluctuations ($R^2 = 0$, $r = 0.12$) were slightly better than the baseline models' predictions.

Prediction models trained on word embeddings were significantly predictive of all affective states: Between-person differences in arousal ($R^2 = 0.09$, $r = 0.31$), contentedness ($R^2 = 0.1$, $r = 0.33$), and sadness ($R^2 = 0.06$, $r = 0.23$) as well as within-person fluctuations of arousal ($R^2 = 0.05$, $r = 0.22$), contentedness ($R^2 = 0.06$, $r = 0.26$), and sadness ($R^2 = 0.02$, $r = 0.13$).

For voice acoustics, as in study 1.1, the Random Forest algorithm performed slightly better than the LASSO algorithm, suggesting non-linear relationships between voice cues and affect experience. On the contrary, for word embeddings, LASSO models performed better than Random Forest models, indicating linear predictor-outcome relationships between speech content as captured with word embeddings and momentary affect experience.

Finally, while predictions of speaker age were not better than chance ($R^2 =$ -0.19, $r = 0.08$) for any of the feature sets, likely due to the low age variance in the data set, gender predictions yielded very good prediction results (prediction accuracy = 95.76%). These findings suggest that voice acoustics and speech content from the collected semi-structured speech samples in study 1.2 contain valuable information about speaker demographics.

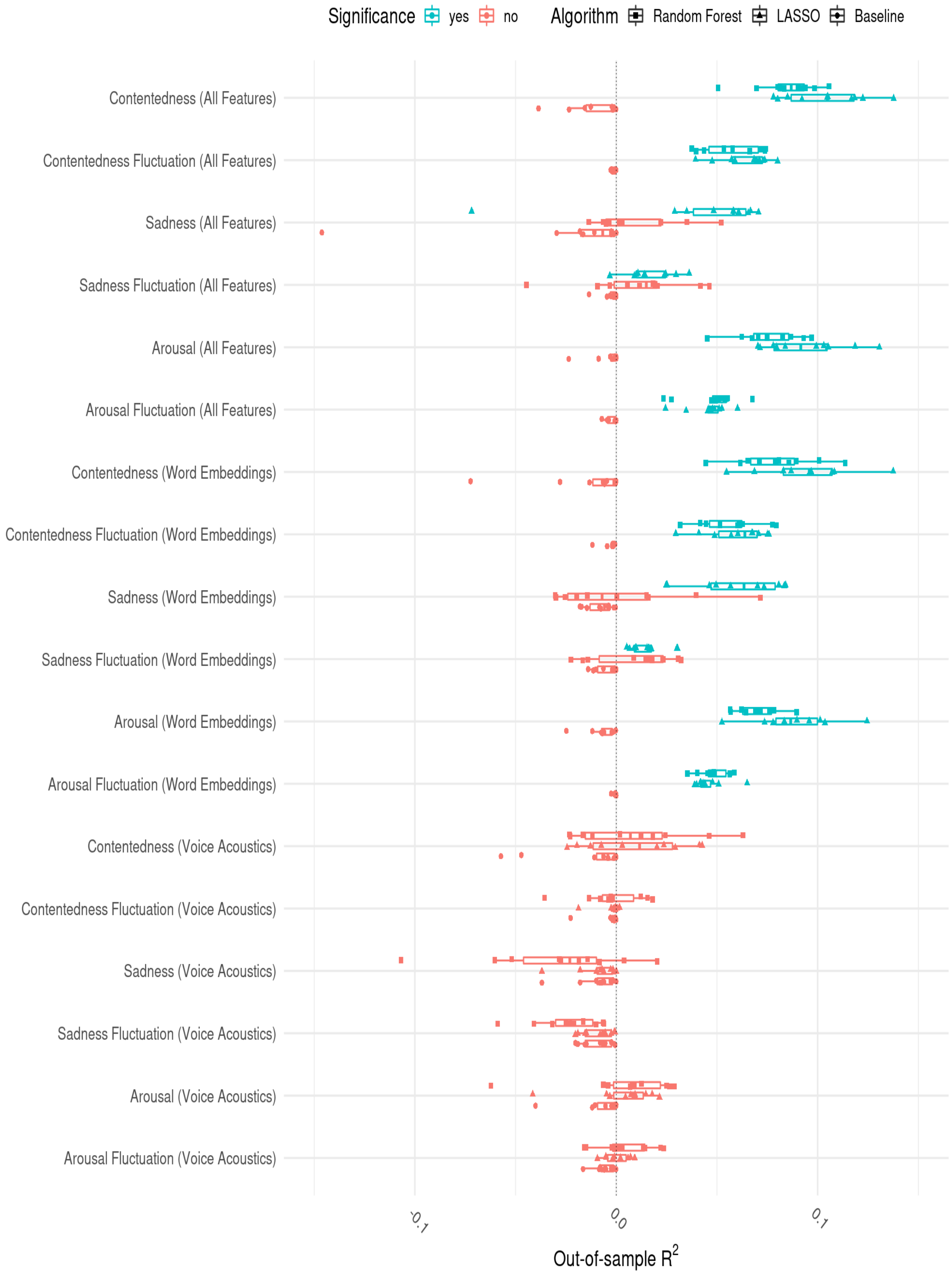


Figure 2.4: Box and whisker plot of prediction performance measures from 10-fold cross-validation for prediction models in different predictions tasks for each feature (sub) set.

Figure 2.5 shows the five most important features in the Random Forest model (based on permutation feature importance) for the prediction of arousal from acoustic voice features. We refrained from reporting feature importance scores for word embeddings as they are not as clearly interpretable as voice features. Overall, in line with findings from study 1.1, features related to loudness of the voice were most important from arousal predictions. Even though features describing spectral flux (i.e., how quickly the power spectrum of the voice signal changes) showed high correlations with self-reported affect experience (see figures 2.11 and 2.12 in the chapter's appendix), other voice features related to the voice spectrum (mean of the mel-frequency cepstral coefficient and bandwidth of the first formant) made it into the top five most important features in the Random Forest model. Again, in line with results from study 1.1, these findings indicate that a louder voice that has a quickly changing broad spectrum is indicative of heightened experienced arousal.

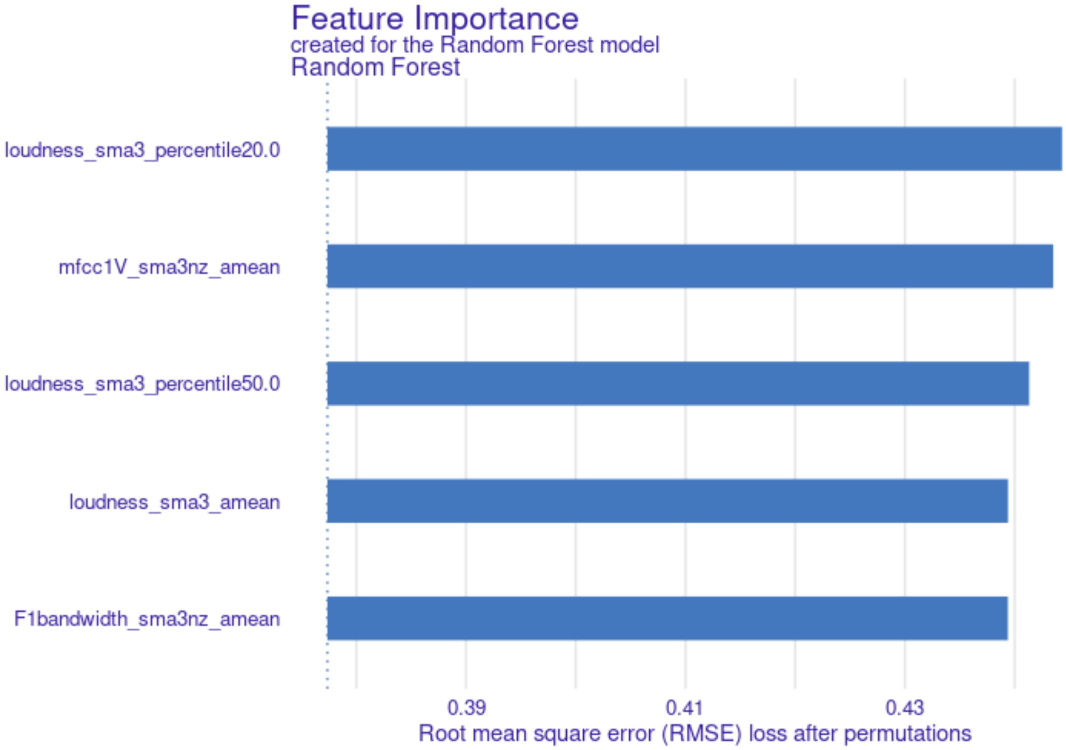


Figure 2.5: Permutation feature importance for the five most predictive features in the Random Forest model for the prediction of arousal. Permutation feature importance represents the decrease in the model's prediction performance (as measured by RMSE) after permuting a single variable.

Content Effects on Affect Predictions from Voice Acoustics

In order to exploratively investigate the effect of the emotional valence of the spoken content on affect predictions from voice cues, we used the sentiment score (M =0.02, SD = 0.29) within the interval of [-1; 1] that had been assigned to each speech transcript by the Google text-to-speech API. Here, in line with the analysis in study 1.1, we analyzed the absolute prediction errors in the prediction of between-person differences in arousal, contentedness, and sadness from voice cues using a Random Forest algorithm. Figure 2.6 depicts the speech sample's sentiment score on the x-axis and the absolute prediction error on the y-axis. Result indicate that content sentiment did not have a clear effect on affect predictions from voice cues. Absolute differences in prediction error with varying sentiment were small overall and the predictive performance of the models was generally limited. Possibly, strong negative and positive content sentiment could have reduced the prediction error for arousal and contentedness from voice cues.

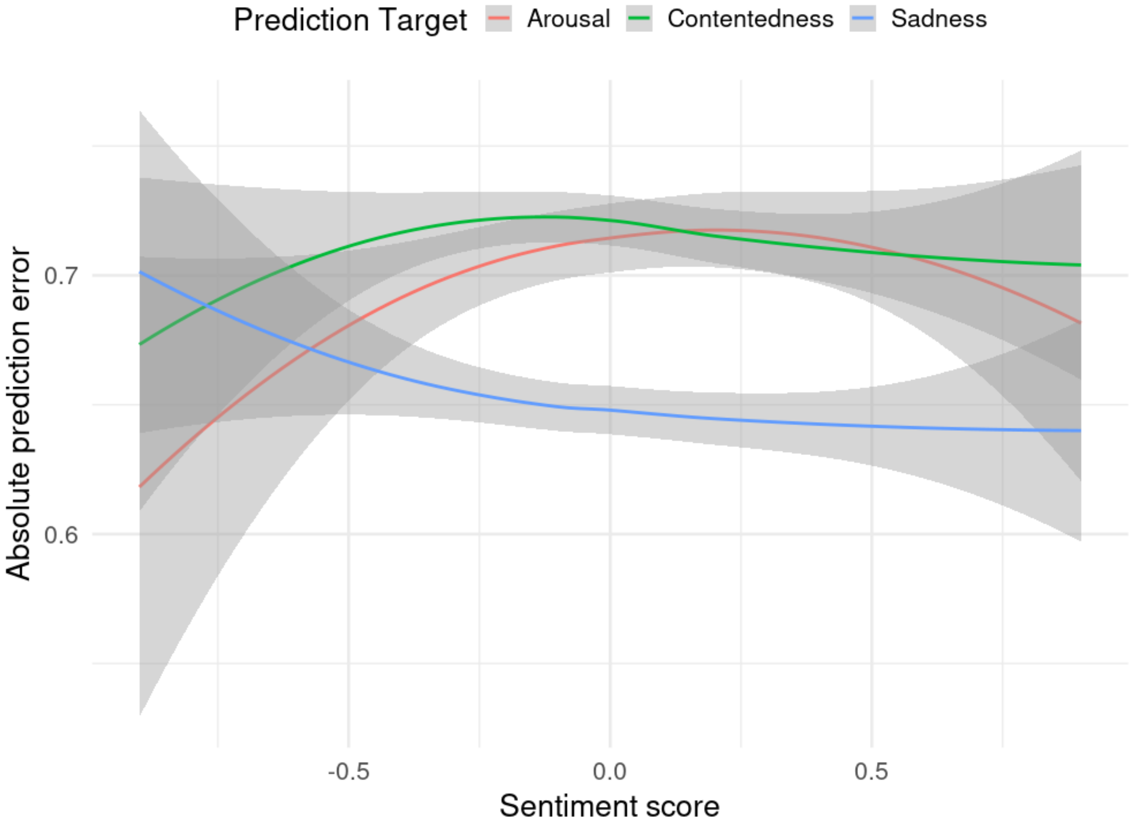


Figure 2.6: Sentiment score of speech content plotted against the absolute prediction error from Random Forest acoustic predictions

2.5 Discussion

In the present work, we extracted acoustic voice parameters and state-of-the-art word embeddings from speech samples collected using smartphones to predict between-person difference and within-person fluctuations in subjective affect experience. While voice acoustics provided limited predictive information of affective arousal across both studies, speech content had been shown to be predictive of arousal as well valence (sadness and contentedness). Overall, predictions were better when participants could talk freely (versus reading out loud predefined emotional sentences). In our models, we identified loudness and features related to fluctuations of the voice spectrum to be particularly predictive of affective arousal. Finally, experimental (study 1.1) and explorative (study 1.2) findings suggest that emotional speech content did not affect predictions from voice acoustics (i.e., what someone talks about does not influence how well affect can be predicted from voice cues).

2.5.1 Recognizing Affect Experience from Speech Cues

Our results indicate that speech samples, and particularly their content, allow for the automatic prediction of subjective momentary affect experience. However, our machine learning models achieve a lower prediction performance as reported in prior work on automatic predictions of affect *expression* (Schuller, 2018). Still, our reported performance is similar to studies predicting subjective affect *experience* from speech samples collected in the wild (Carlier et al., 2022; J. Sun et al., 2020; Weidman et al., 2020). This observation is in line with prior research suggesting that real-life emotions are more difficult to algorithmically recognize than acted or elicited emotions (Vogt et al., 2008). Also, there are only few instances of extreme affect experiences in our data sets compared to the data used in prior studies on acted or labelled emotions. As a result, we rather predicted *mood* in this work, which is, by definition, less intense than emotions (Scherer, 2003) and, consequently, more challenging to recognize.

Furthermore, across the two studies, arousal predictions from voice acoustics were better than those of emotional valence, highlighting prior work showing that the latter is more challenging to automatically infer due to its individual nature (Sridhar & Busso, 2022). Moreover, in line with prior work (J. Sun et al., 2020; Weidman et al., 2020), overall predictions of between-person differences in subjective affect experience were superior to those of within-person fluctuations.

Further, as done in prior research on voice-affect predictions (Weidman et al., 2020), we also compared the prediction performance of machine learning models trained on the much larger Compare2016 (6737 features) acoustic feature set (Schuller et al., 2016) in contrast to the economic eGeMAPS (88 features) feature set we had used. Just as in prior research, the larger voice feature set did not yield better affect predictions (Weidman et al., 2020). This finding suggests that an economic acoustic feature set is sufficient for affect detection from voice. Moreover, the small features set is less computationally expensive and would allow for online or on-device pre-processing in a scientific or applied setting.

Generally, our findings challenge the transferability of the optimistic prediction results from prior research work on the recognition of affect *expression* (e.g., enacted speech) to the recognition of subjective affect *experience* in everyday speech, particularly from acoustic voice cues. Thereby, our findings also question the proclaimed performance of commercial affect recognition algorithms deployed in daily life that have been mostly trained on enacted or labelled affect *expression*. Consequently, current expectations regarding the performance of emotion-detecting AI services, especially the ones that are focused on voice cues, applied to everyday speech might be overoptimistic. More research is needed to determine how well algorithms can pick up on subjective affect experience from day-to-day speech.

In future research, smartphones could play a prominent role in collecting and analyzing speech data and corresponding in-situ self-reports on subjective affect experience for affect inferences. Hereby, starting from our work, smartphones could be used as a mobile experimental lab to study different aspects of affect recognition from speech, for example by experimentally varying the content as done in study 1.1 (Miller, 2012).

2.5.2 The Context of Speech Production Matters

Our results indicate that the context of speech production (i.e., reading out predefined emotional sentences versus prompted free speech) had an impact on affect predictions. While the predefined sentences in study 1.1 allowed us to control for the emotionality of the content of participants' voice records, they were unable to express themselves freely, which could have impaired predictions from voice acoustics compared to study 1.2 where participants could talk freely. As a result, researchers and practitioners should consider the context in which speech had been produced in and keep in mind that findings and trained models might be specific to the given production context and do not necessarily generalize well to other production contexts.

2.5.3 The Role of Speech Content

In study 1.2, state-of-the-art word embeddings showed a superior affect prediction performance compared to voice acoustics suggesting that speech content could contain more affective information than speech form. This finding is in line with prior research that found speech content to be more predictive than voice acoustics when predicting momentary subjective experience of happiness (J. Sun et al., 2020; Weidman et al., 2020). As a consequence, even though prior research has suggested that voice acoustics could be more relevant for human affect inferences than speech content (Ben-David et al., 2016; Lin et al., 2020), future research and AI applications should consider both channels - content and form - simultaneously.

By experimentally varying the emotional valence of the spoken content in study 1.1 and exploratively investigating the effect of word sentiment on voice predictions in study 1.2, our findings suggest that the content what participants talked about did not have a substantial impact on affect predictions from voice cues. This insight could imply that it does not matter what people talk about when algorithmically inferring affect experience from voice cues. However, one has to keep than in mind that affect predictions from voice cues were overall not very strong, particularly in study 1.1. More research is needed to disentangle speech content and form in automatic affect recognition.

2.5.4 Arousal-Linked Voice Variations

Interpretable machine learning methods and descriptive correlations suggest that voice cues related to loudness and spectral features are associated with and predictive of affective arousal. Specifically, a louder voice with a quickly changing broad spectrum was found to be indicative of heightened experienced arousal. These arousal-linked voice patterns are in line with findings from prior research (Hildebrand et al., 2020; Weninger et al., 2013).

2.6 Limitations and Future Directions

In this section, we discuss the two specific limitations of this study. General limitations related to the data collection method and the measurement of self-reported affect experience are discussed in the general discussion of the present dissertation (see section 4.3).

First, we used slightly different operationalizations of affect experience and sample compositions in the two studies that might affect their comparability: In study 1.1, we assessed valence and arousal on a six-point Likert scale. In study 1.2, we used two items to assess affective valence (contentedness and sadness) and arousal on five-/four-point Likert scales. As a consequence, findings might not be directly comparable. Further, while study 1.1 drew on a representative German sample, study 1.2 was based on a student convenience sample from the United States with the respective limitations, such as potential constraints in generalizability of findings (Müller, Chen, Peters, Chaintreau, & Matz, 2021). Future studies should investigate multiple target emotions in diverse international samples from different cultural contexts in non-western countries.

Second, and most importantly, in contrast to prior work using passive speech sensing (e.g., via the EAR), our participants had to actively log their speech in the present work. This artificial setting might have had an effect on results. Moreover, the findings of this study might be subject to the specific instructions that had been given for the audio records: In study 1.1, participants were instructed to read out predefined sentences and, in study 1.2, participants were prompted to talk about the situation they were in as well as their current thoughts and feelings in a semi-structured fashion. While affect-linked acoustic voice cues in the two studies are similar and are possibly transferable to new voice data, word embeddings are specific to the given task in study 1.2. In this manner, future work should employ multiple different speech tasks for affect predictions and investigate how well predictions generalize from one to another.

Moreover, specifically to study 1.1, another related limitation lies in our privacy-preserving on-device data pre-processing approach. By applying on-device feature extraction, we had no opportunity to check in detail if participants truly complied with study instructions and had recorded their voice while reading out the predefined sentences accordingly (beyond the data-driven quality checks we had applied). Further, our approach did not allow to control for records' background noise (e.g., when participants were outside next to a road) or how they held their smartphone during the voice record. Since checking single raw audio files manually would be out of scope, future research could investigate additional data-driven approaches to check speech data quality directly on the device. Finally, in future work, smartphones could be used to log and immediately pre-process participants' everyday speech by using pre-trained language models to extract content features (e.g., specific topics or word embeddings) directly on the device, too. Thereby, no raw speech data would have to be transferred to a server and valuable information of language's content could be also used for privacy-respectful affect recognition.

2.7 Conclusion

In this work, we investigated if machine learning algorithms can recognize subjective affect experience from speech samples collected in the wild using smartphones. Extracted acoustic voice parameters provided limited predictive information of affective arousal across both studies, while speech content as reflected in state-of-the-art word embeddings had been shown to be predictive of arousal as well valence (sadness and contentedness). Overall, voice predictions were better when participants could talk freely (versus reading out loud predefined emotional sentences). Also, speech content showed superior prediction performance compared to voice acoustics. Further, experimental and explorative findings suggest that emotional speech content did not affect predictions from voice acoustics (i.e., what someone talked about did not affect how well emotions could be predicted from voice cues). Our findings challenge the transferability of the optimistic prediction results from prior research work and commercial emotion-detection AI algorithms on the recognition of affect *expression* (e.g., enacted and labelled speech) to the recognition of subjective affect *experience* in everyday speech. Finally, we discussed resulting implications for the algorithmic monitoring of affect experience.

2.8 Author Contribution

In addition to myself, Florian Bemmann (F.B.), Markus Buehner (M.B.), Gabriella Harari (G.H.), Zachariah Marrero (Z.M.), Ramona Schoedel (R.S.), and Clemens Stachl (C.S.) contributed to this study. M.B., G.H., R.S., and C.S. acted as supervisors. F.B. created the logging software to collect the data for study 1.1. R.S. managed data collection of study 1.1. Z.M. assisted with preprocessing the raw data for study 1.2.

2.9 Acknowledgements

We thank Peter Ehrich and Dominik Heinrich for their support with the technical implementation of the on-device acoustic feature extraction for study 1.1. We thank the Leibniz Institute for Psychology (ZPID) for the support with data collection for study 1.1 and Sumer Vaid for the technical support in setting up the computational analysis environment for study 1.2.

2.10 Appendix

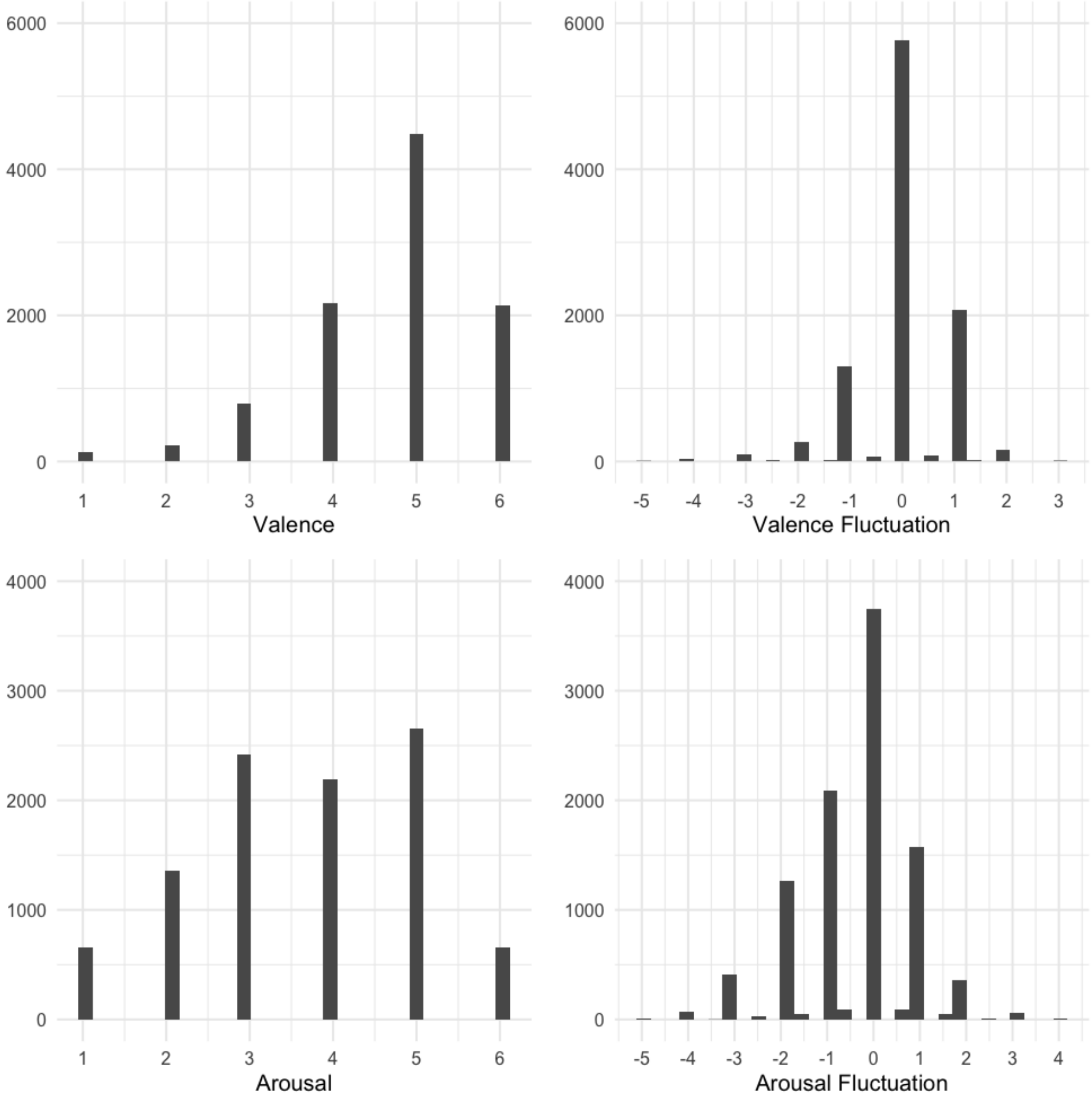


Figure 2.7: Distribution of affect measures in the data set from Germany that had been used in study 1.1.

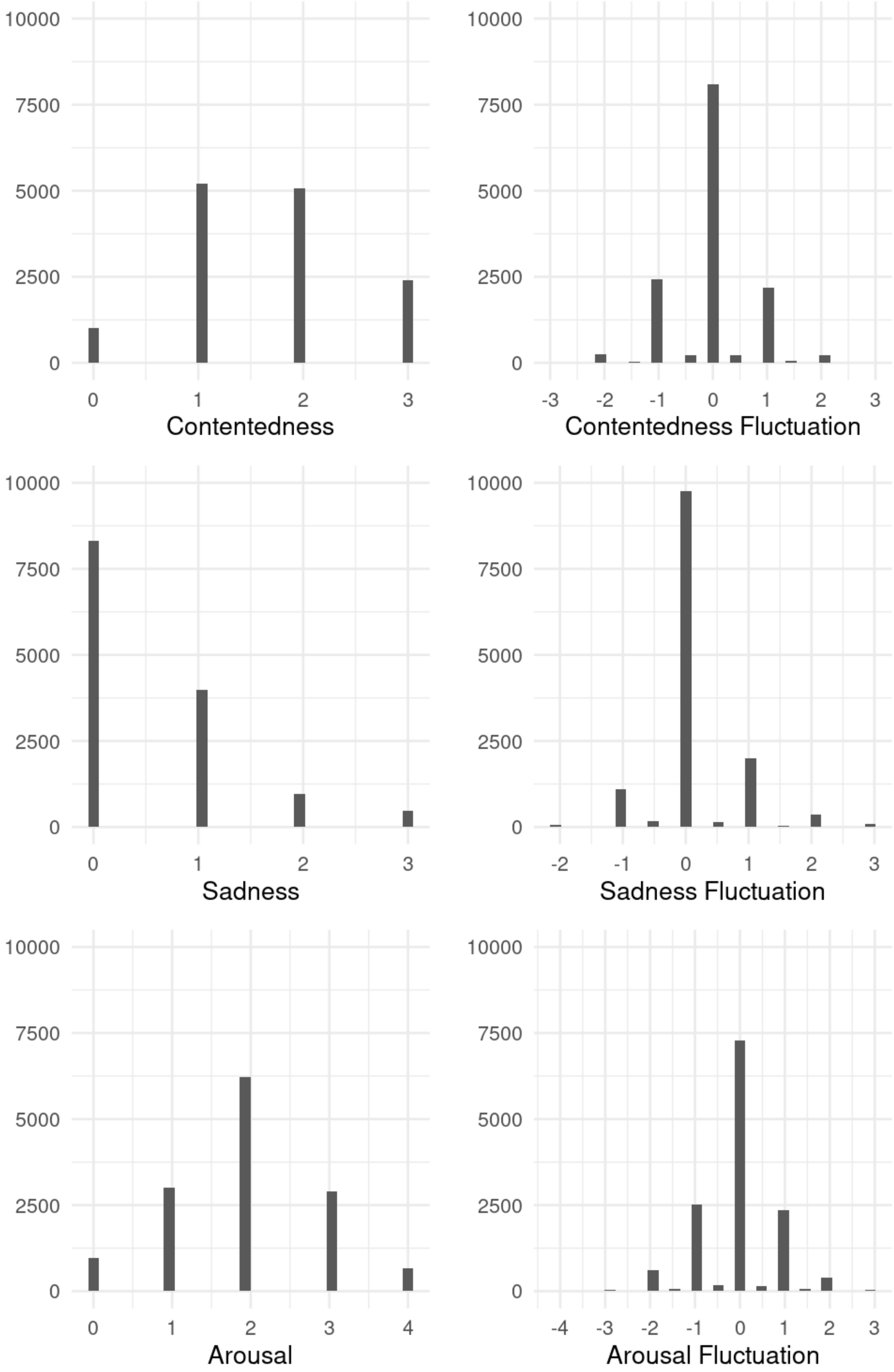


Figure 2.8: Distribution of affect measures in the data set from the United States that had been used in study 1.2.

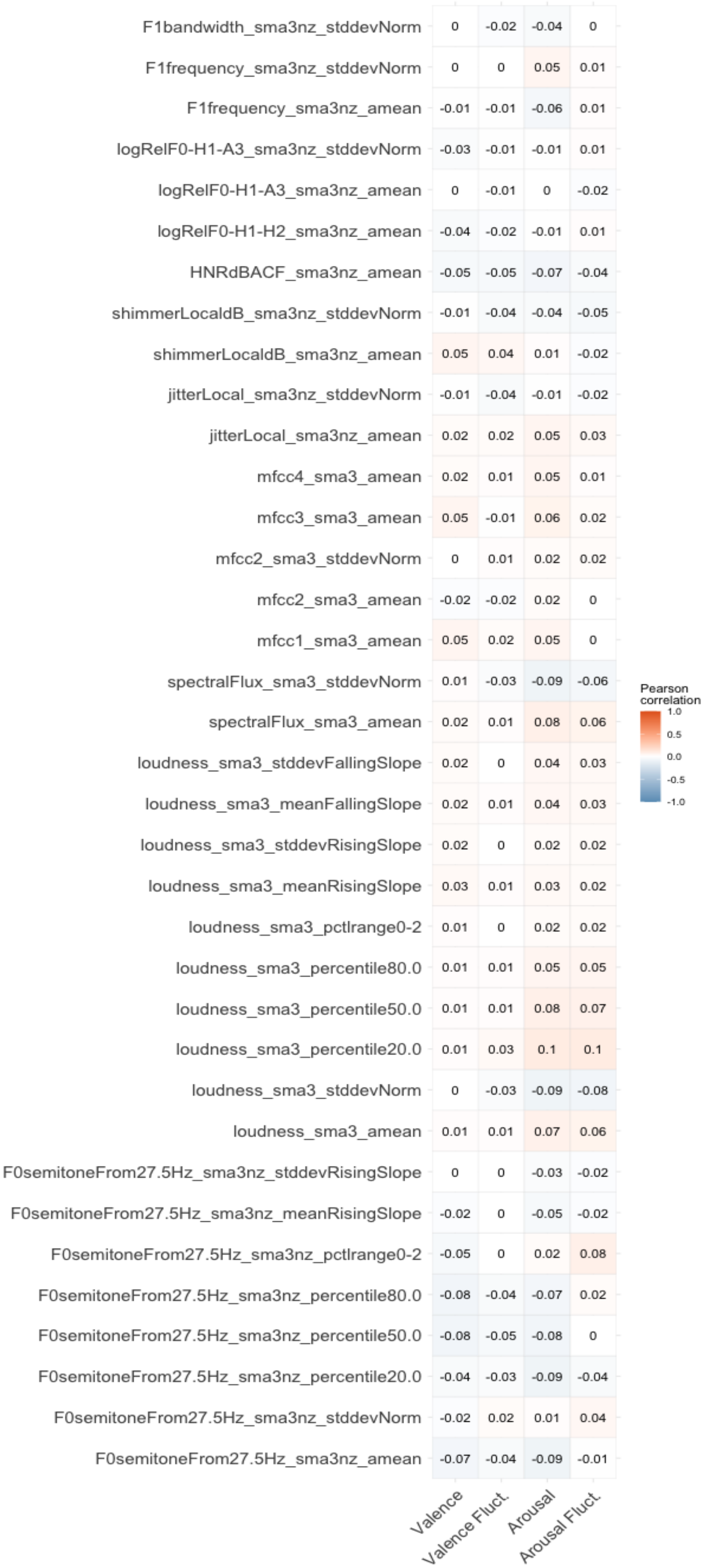
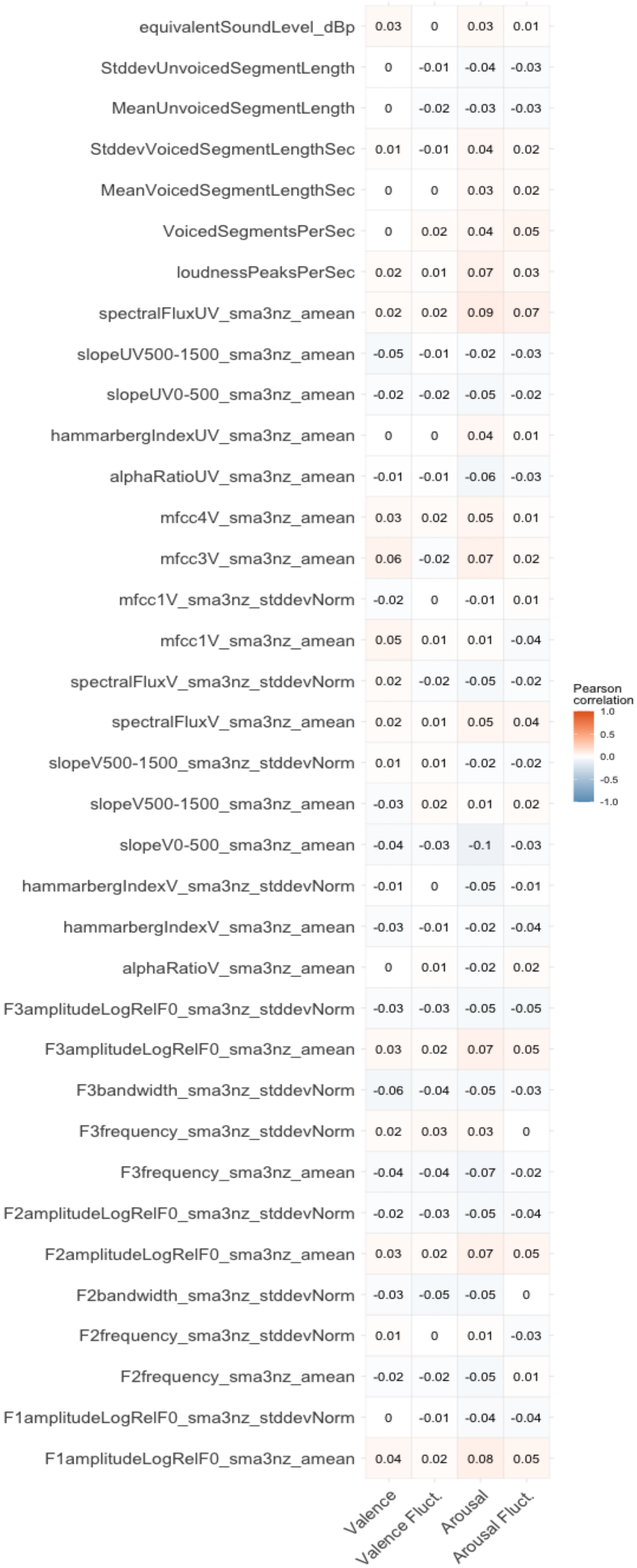
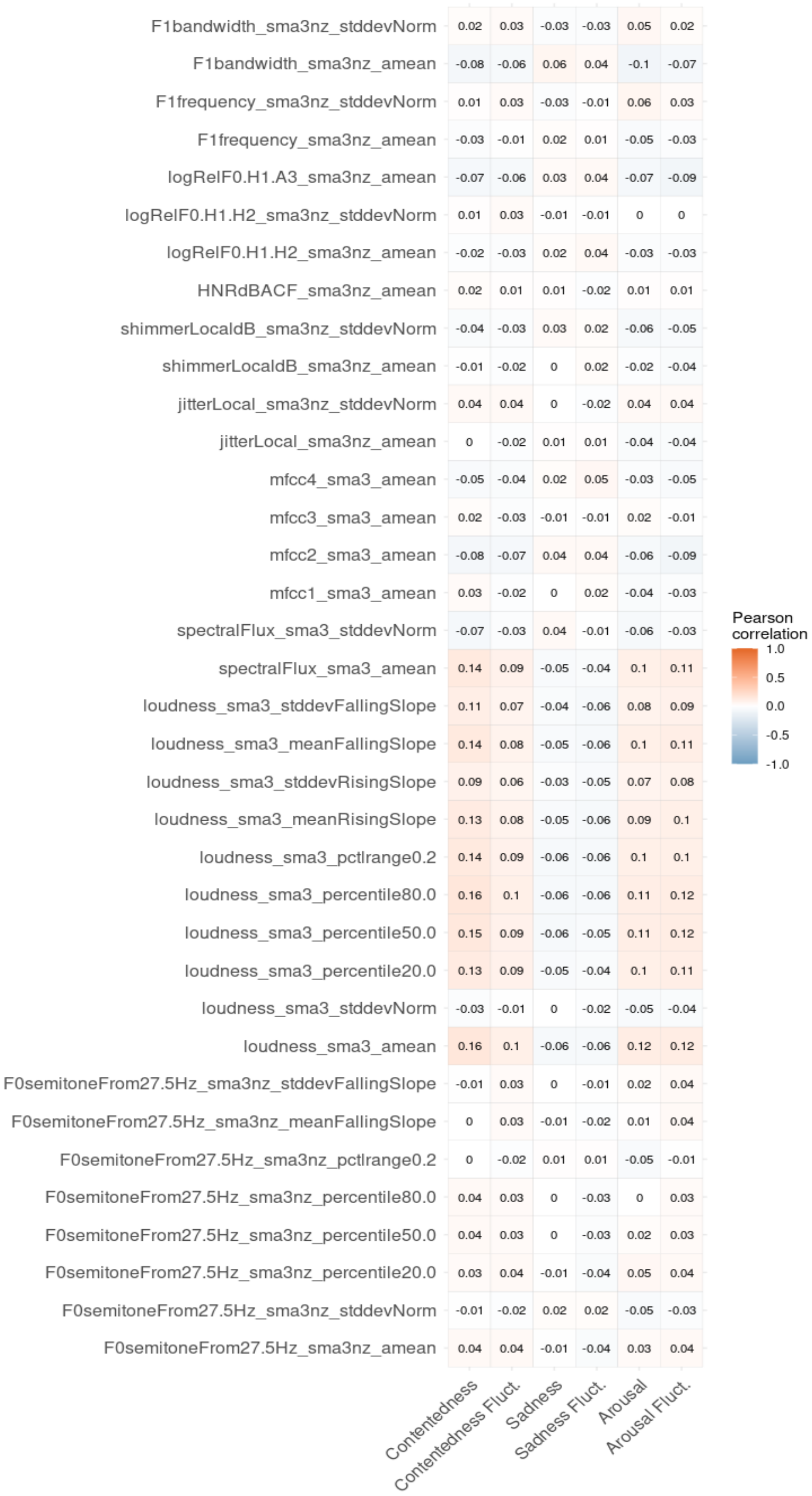


Figure 2.9: Pearson correlations of voice features with momentary affect experience in study 1.1. Those voice features are displayed for which the 95% confidence interval for the correlation coefficient does not contain zero for any of the outcomes.



	Valence	Valence Fluct.	Arousal	Arousal Fluct.
equivalentSoundLevel_dBp	0.03	0	0.03	0.01
StddevUnvoicedSegmentLength	0	-0.01	-0.04	-0.03
MeanUnvoicedSegmentLength	0	-0.02	-0.03	-0.03
StddevVoicedSegmentLengthSec	0.01	-0.01	0.04	0.02
MeanVoicedSegmentLengthSec	0	0	0.03	0.02
VoicedSegmentsPerSec	0	0.02	0.04	0.05
loudnessPeaksPerSec	0.02	0.01	0.07	0.03
spectralFluxUV_sma3nz_amean	0.02	0.02	0.09	0.07
slopeUV500-1500_sma3nz_amean	-0.05	-0.01	-0.02	-0.03
slopeUV0-500_sma3nz_amean	-0.02	-0.02	-0.05	-0.02
hammarbergIndexUV_sma3nz_amean	0	0	0.04	0.01
alphaRatioUV_sma3nz_amean	-0.01	-0.01	-0.06	-0.03
mfcc4V_sma3nz_amean	0.03	0.02	0.05	0.01
mfcc3V_sma3nz_amean	0.06	-0.02	0.07	0.02
mfcc1V_sma3nz_stddevNorm	-0.02	0	-0.01	0.01
mfcc1V_sma3nz_amean	0.05	0.01	0.01	-0.04
spectralFluxV_sma3nz_stddevNorm	0.02	-0.02	-0.05	-0.02
spectralFluxV_sma3nz_amean	0.02	0.01	0.05	0.04
slopeV500-1500_sma3nz_stddevNorm	0.01	0.01	-0.02	-0.02
slopeV500-1500_sma3nz_amean	-0.03	0.02	0.01	0.02
slopeV0-500_sma3nz_amean	-0.04	-0.03	-0.1	-0.03
hammarbergIndexV_sma3nz_stddevNorm	-0.01	0	-0.05	-0.01
hammarbergIndexV_sma3nz_amean	-0.03	-0.01	-0.02	-0.04
alphaRatioV_sma3nz_amean	0	0.01	-0.02	0.02
F3amplitudeLogRelF0_sma3nz_stddevNorm	-0.03	-0.03	-0.05	-0.05
F3amplitudeLogRelF0_sma3nz_amean	0.03	0.02	0.07	0.05
F3bandwidth_sma3nz_stddevNorm	-0.06	-0.04	-0.05	-0.03
F3frequency_sma3nz_stddevNorm	0.02	0.03	0.03	0
F3frequency_sma3nz_amean	-0.04	-0.04	-0.07	-0.02
F2amplitudeLogRelF0_sma3nz_stddevNorm	-0.02	-0.03	-0.05	-0.04
F2amplitudeLogRelF0_sma3nz_amean	0.03	0.02	0.07	0.05
F2bandwidth_sma3nz_stddevNorm	-0.03	-0.05	-0.05	0
F2frequency_sma3nz_stddevNorm	0.01	0	0.01	-0.03
F2frequency_sma3nz_amean	-0.02	-0.02	-0.05	0.01
F1amplitudeLogRelF0_sma3nz_stddevNorm	0	-0.01	-0.04	-0.04
F1amplitudeLogRelF0_sma3nz_amean	0.04	0.02	0.08	0.05

Figure 2.10: Pearson correlations of voice features with momentary affect experience in study 1.1. Those voice features are displayed for which the 95% confidence interval for the correlation coefficient does not contain zero for any of the outcomes.



	Contentedness	Contentedness Fluct.	Sadness	Sadness Fluct.	Arousal	Arousal Fluct.
F1bandwidth_sma3nz_stddevNorm	0.02	0.03	-0.03	-0.03	0.05	0.02
F1bandwidth_sma3nz_amean	-0.08	-0.06	0.06	0.04	-0.1	-0.07
F1frequency_sma3nz_stddevNorm	0.01	0.03	-0.03	-0.01	0.06	0.03
F1frequency_sma3nz_amean	-0.03	-0.01	0.02	0.01	-0.05	-0.03
logRelF0.H1.A3_sma3nz_amean	-0.07	-0.06	0.03	0.04	-0.07	-0.09
logRelF0.H1.H2_sma3nz_stddevNorm	0.01	0.03	-0.01	-0.01	0	0
logRelF0.H1.H2_sma3nz_amean	-0.02	-0.03	0.02	0.04	-0.03	-0.03
HNRdBACF_sma3nz_amean	0.02	0.01	0.01	-0.02	0.01	0.01
shimmerLocaldB_sma3nz_stddevNorm	-0.04	-0.03	0.03	0.02	-0.06	-0.05
shimmerLocaldB_sma3nz_amean	-0.01	-0.02	0	0.02	-0.02	-0.04
jitterLocal_sma3nz_stddevNorm	0.04	0.04	0	-0.02	0.04	0.04
jitterLocal_sma3nz_amean	0	-0.02	0.01	0.01	-0.04	-0.04
mfcc4_sma3_amean	-0.05	-0.04	0.02	0.05	-0.03	-0.05
mfcc3_sma3_amean	0.02	-0.03	-0.01	-0.01	0.02	-0.01
mfcc2_sma3_amean	-0.08	-0.07	0.04	0.04	-0.06	-0.09
mfcc1_sma3_amean	0.03	-0.02	0	0.02	-0.04	-0.03
spectralFlux_sma3_stddevNorm	-0.07	-0.03	0.04	-0.01	-0.06	-0.03
spectralFlux_sma3_amean	0.14	0.09	-0.05	-0.04	0.1	0.11
loudness_sma3_stddevFallingSlope	0.11	0.07	-0.04	-0.06	0.08	0.09
loudness_sma3_meanFallingSlope	0.14	0.08	-0.05	-0.06	0.1	0.11
loudness_sma3_stddevRisingSlope	0.09	0.06	-0.03	-0.05	0.07	0.08
loudness_sma3_meanRisingSlope	0.13	0.08	-0.05	-0.06	0.09	0.1
loudness_sma3_pctlrange0.2	0.14	0.09	-0.06	-0.06	0.1	0.1
loudness_sma3_percentile80.0	0.16	0.1	-0.06	-0.06	0.11	0.12
loudness_sma3_percentile50.0	0.15	0.09	-0.06	-0.05	0.11	0.12
loudness_sma3_percentile20.0	0.13	0.09	-0.05	-0.04	0.1	0.11
loudness_sma3_stddevNorm	-0.03	-0.01	0	-0.02	-0.05	-0.04
loudness_sma3_amean	0.16	0.1	-0.06	-0.06	0.12	0.12
F0semitoneFrom27.5Hz_sma3nz_stddevFallingSlope	-0.01	0.03	0	-0.01	0.02	0.04
F0semitoneFrom27.5Hz_sma3nz_meanFallingSlope	0	0.03	-0.01	-0.02	0.01	0.04
F0semitoneFrom27.5Hz_sma3nz_pctlrange0.2	0	-0.02	0.01	0.01	-0.05	-0.01
F0semitoneFrom27.5Hz_sma3nz_percentile80.0	0.04	0.03	0	-0.03	0	0.03
F0semitoneFrom27.5Hz_sma3nz_percentile50.0	0.04	0.03	0	-0.03	0.02	0.03
F0semitoneFrom27.5Hz_sma3nz_percentile20.0	0.03	0.04	-0.01	-0.04	0.05	0.04
F0semitoneFrom27.5Hz_sma3nz_stddevNorm	-0.01	-0.02	0.02	0.02	-0.05	-0.03
F0semitoneFrom27.5Hz_sma3nz_amean	0.04	0.04	-0.01	-0.04	0.03	0.04

Figure 2.11: Pearson correlations of voice features with momentary affect experience in study 1.2. Those voice features are displayed for which the 95% confidence interval for the correlation coefficient does not contain zero for any of the outcomes.

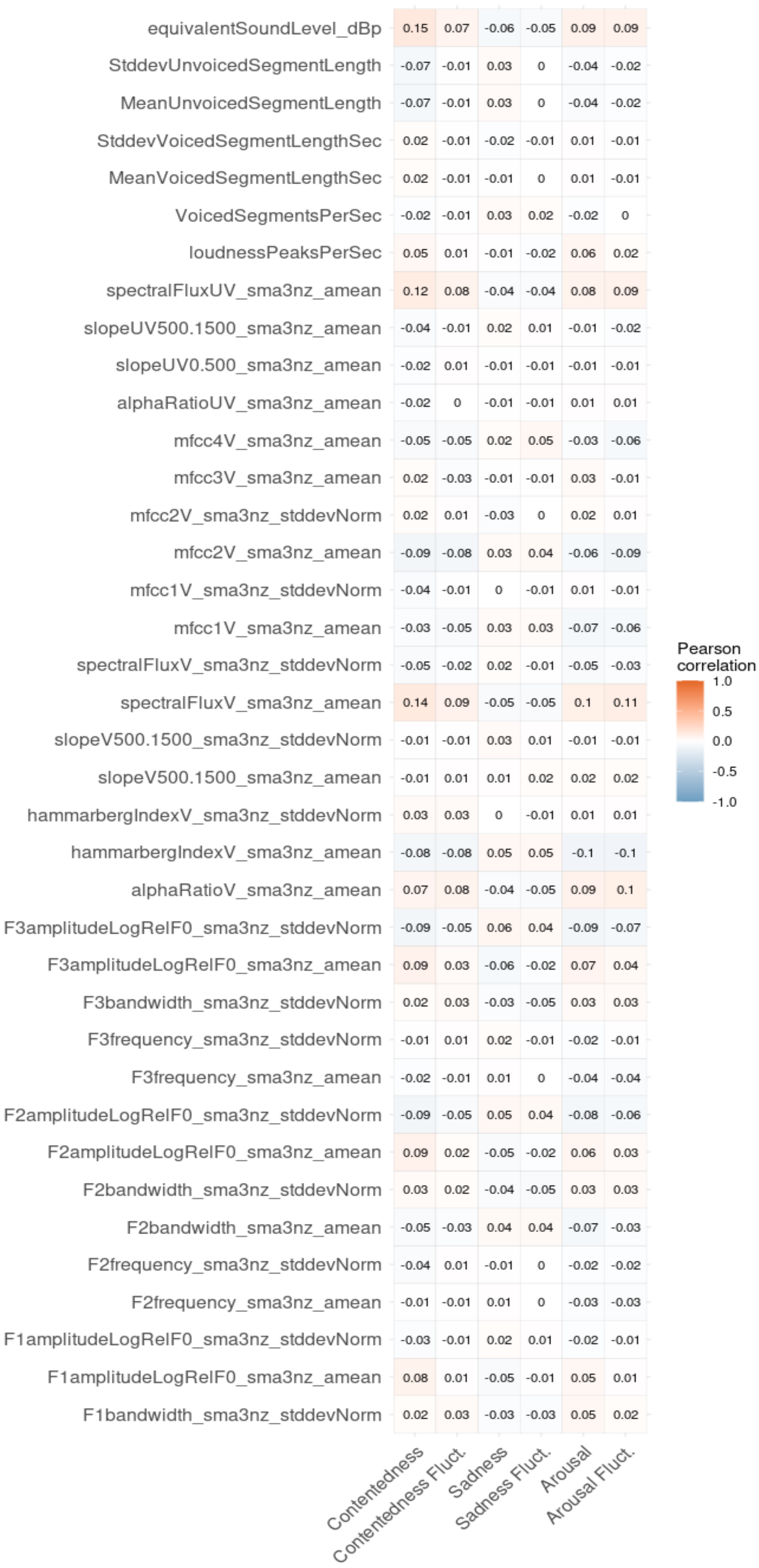


Figure 2.12: Pearson correlations of voice features with momentary affect experience in study 1.2. Those voice features are displayed for which the 95% confidence interval for the correlation coefficient does not contain zero for any of the outcomes.

Chapter 3

Affect Experience in Language Patterns Across Contexts

3.1 Abstract

Research suggests that digital traces of written language, such as social media posts, offer a unique window into our emotional lives. However, those studies rest on the critical assumption that people's language use, for instance of emotion words (e.g., "happy" and "sad"), is a reflection of their subjective affect experience in a given moment. In this work, we test this assumption by investigating (in-sample) associations of self-reported affect experience collected using the experience sampling method and everyday language characteristics logged with smartphones. Moreover, we investigate if those language features allow for the (out-of-sample) prediction of between-person differences and within-person fluctuations in affect experience. In those analyses, we also distinguish between private (e.g., messaging on WhatsApp) and public communication contexts (e.g., posting on Facebook) as well different time aggregations (trait, weekly, daily, momentary). From a data set of more than 10 million typed words from 486 participants, we extract features regarding typing dynamics, word use based on word dictionaries, and emoji and emoticon use. We identify distinct affect-linked language variations across communication contexts and time frames. Predictions from machine learning algorithms, however, are not significantly better than chance. Finally, we discuss implications for the assessment of subjective affect experience using occurrence-counts, such as word dictionaries, and opportunities leveraging smartphones to collect language data in the wild.

3.2 Introduction

Prior research studies suggest that the words we use in textual language reveal how we feel (Tov et al., 2013; Vine, Boyd, & Pennebaker, 2020). As digital footprints in the form of text, such as on blog posts, social media posts, or instant messages, have become ubiquitous, new opportunities to investigate affect-linked language have emerged. In this manner, Facebook posts have been found to be predictive of emotions (Eichstaedt & Weidman, 2020; Preoţiuc-Pietro et al., 2016). Moreover, researchers predicted affective disorders, for example depression, from digital text data, such as Facebook posts (Eichstaedt et al., 2018), tweets (De Choudhury, Counts, & Horvitz, 2013), and text messages (Tony Liu et al., 2021). This research field and corresponding commercial tools leveraging artificial intelligence (AI), for example in mental health care, are grounded on the critical assumption that the language one uses (i.e., the words) reflects how they feel in a given moment. However, recent works question the associations of and predictions from word use and subjective affect experience (Kelley, Mhaonaigh, Burke, Whelan, & Gillan, 2022; Kross et al., 2019; J. Sun et al., 2020). Rather, language use could be linked to one's *expression* of affect, which has some overlaps with one's subjective affect *experience*, but is not necessarily congruent (Kross et al., 2019). Moreover, language-based prediction models often do not generalize well to other communication contexts (e.g., from text messages to social media posts) due to each channel's specific linguistic peculiarities (Tingting Liu et al., 2022). The present work leverages a novel language logging approach using off-the-shelf smartphones to investigate language-linked between-person differences and within-person fluctuations in affect experience across different time frames and communication contexts.

3.2.1 Affective Text

For the investigation of linguistic associations of stable *trait* affect that is considered a personality characteristic, researchers would collect text data over a period of time, for example diary entries (Tov et al., 2013) or text messages (Massachi et al., 2020), and correlate the use of linguistic features, such as specific word categories (e.g., positive emotion words), with self-reported scores from questionnaires. Alternatively, they would analyze language traces of affective disorders, for example depression, that are also assumed to be relatively stable in a given time window (Eichstaedt et al., 2018; Tackman et al., 2018).

On the contrary, the investigation of the association of language use and fluctuating affective *states* (e.g., short-termed emotions and longer moods) is more challenging since researchers usually cannot ask people how they are feeling in the exact moment when they produce a piece of text. Therefore, many prior studies associated language use with tragic events that were expected to elicit strong emotions, like September 11 (Back, Küfner, & Egloff, 2010; Cohn, Mehl, & Pennebaker, 2004) or the death of George Floyd (Eichstaedt, Sherman, et al., 2021). Alternatively, researchers would hire raters to assign affective labels to collected text data, such as Facebook posts (Eichstaedt & Weidman, 2020; Preoţiuc-Pietro et al., 2016).

These kind of language studies often rely on word dictionaries that either assign a numeric sentiment score to each single word, for example using the popular Valence Aware Dictionary for Sentiment Reasoning (VADER) dictionary (Hutto & Gilbert, 2014), or that count the frequency of words from a set of given word categories (e.g., positive emotion words). The most widespread psycholinguistic dictionary is the Linguistic Inquiry and Word Count (LIWC) (Pennebaker, Booth, Boyd, & Francis, 2015). LIWC has multiple emotional word categories that contain words which are assumed to be indicative of positive or negative affect: The *positive emotion* word category includes words like “happy” and “joy” while the *negative emotion* category is comprised of the sub-categories *anxiety* (e.g., “worried), *anger* (e.g.,”angry”), and *sadness* (e.g., “tear”).

The assumption that the use of words from specific emotion word categories reflects how one feels in a given moment represents the foundation of many prior research studies. Sun and colleagues provide a comprehensive overview of research findings on the associations of positive and negative emotion LIWC dictionaries with affect in past research (J. Sun et al., 2020). However, recent studies question if emotion words truly reflect momentary subjective affect experience (Kross et al., 2019; J. Sun et al., 2020).

Apart from emotion words that serve as *direct* linguistic markers (e.g., “I’m feeling *happy*”) in the form of positive and negative emotion words, researchers have also found evidence for *indirect* linguistic markers of affect, such as *social processes* and the use of *function words* (e.g., first person singular). Beyond serving as a linguistic measure these indirect linguistic markers of affect can also create insights into the processes of affect experience. For example, socializing is known to improve positive affect compared to being alone (Lucas, Le, & Dyrenforth, 2008). Therefore, writing about social processes (e.g., “I am meeting *friends*”) might serve as an indirect marker of peoples’ positive affect that is associated with spending time with others. In a similar fashion, the use of 1st person singular that could indicate a strong focus on oneself has been liked to negative affect and depression (Tackman et al., 2018).

Modern digital communication is not only comprised of plain text, but often also features emoji and their predecessors, the emoticons. These can be used to emotionally enrich text (Riordan, 2017). In order to assess the specific emotional meaning of emoji, prior research had used raters that would assess a given set of emoji regarding their affective valence and arousal (Kralj Novak, Smailović, Sluban, & Mozetič, 2015; Krekhov, Emmerich, Fuchs, & Krueger, 2022; Kutsuzawa, Umemura, Eto, & Kobayashi, 2022). For instance, the "😘" emoji has been universally rated as expressing positive affect. However, it remains unclear if people employing this particular emoji are also always subjectively experiencing positive affect in the moment when using it.

Not only the textual content, such as the employed words and emoji, of digital text, but also the way how it had been created, analogue to the form of the voice in speech, can contain valuable affective information. These typing dynamics, even though not visible to the communication partner, can be logged though a device's (e.g., personal computer or smartphone) keyboard. The logged data can then be used to investigate affect-linked typing dynamics. In this manner, prior research suggests, for instance, that particularly typing speed is related to affective states (Ghosh et al., 2017; Ghosh, Hiware, Ganguly, Mitra, & De, 2019).

3.2.2 Predicting Affect from Text

More recently, researchers have been employing machine learning algorithms that are able to handle large and multi-dimensional data coming from digital textual footprints to predict affect. Using this approach, scientists have, for example, predicted affective states from Facebook posts (Eichstaedt & Weidman, 2020; Preoţiuc-Pietro et al., 2016) and typing dynamics (Ghosh et al., 2017; Mandi, Ghosh, De, & Mitra, 2022). In a similar fashion, studies found affective disorders to be predictable from social media text (De Choudhury et al., 2013; Eichstaedt et al., 2018), text messages (Tony Liu et al., 2021), and even solely from typing dynamics (Bennett, Ross, Baek, Kim, & Leow, 2022b, 2022a; Cao et al., 2017; Mastoras et al., 2019).

To make accurate inferences about one's affect experience from collected language data, particularly for fluctuating affect states, researchers face three major challenges. First, in order to predict an author's affective state from text, one needs to assess the author's affect in the exact moment of text production. Since scientists cannot not ask people what affect they had experienced in the moment they had written a piece of text in the past, for example a social media post, researchers usually use raters to assign emotion labels to existing text data (Kross et al., 2019). As a consequence, those studies are concerned with the associations with and recognition of *affect expression* from text rather than *subjective affect experience*. This is due to the fundamental difference of affect expression and affect experience (Kross et al., 2019), meaning how people subjectively feel versus the cues they express, for example through their language use, and that are interpreted by others (see section 1.3.3 for more details on the conceptual difference between affect experience and affect expression). In this manner, recent studies leveraging the experience sampling method (ESM) to assess in-situ self-reports question the association of emotional word use and subjective experienced affect. For instance, Kross and colleagues did not find any correlations of emotional word use in Facebook posts and self-reported subjective affect experience (Kross et al., 2019). In the same manner, transcribed speech samples collected with the Electronically Activated Recorder (EAR) and experience-sampled self-reports of the subjective experience of happiness did not yield significant associations with emotional word use (J. Sun et al., 2020).

A second challenge that scientists face when investigating affect-linked language is to collect sufficient textual data that corresponds with affect experience data (see section 1.3.3). Text that had been created in experimental settings is usually rather short and lacks ecologic validity. However, sufficient text data is needed for accurate predictions of psychological characteristics (Eichstaedt, Kern, et al., 2021). However, in studies leveraging social media posts, there is only data available when people have posted online. Consequently, there are many gaps in the data stream when people have not posted. As a result, granular affect predictions from social media text are not feasible and only possible for larger time frames, for example weeks (Eichstaedt & Weidman, 2020). A continuous tracking (e.g., on the hourly level) of affect experience, however, requires more granular data that social media text data cannot provide for most users. More granular data would also allow researchers to not only investigate between-person differences in affect (i.e., is this person sad?), but also assess within-person fluctuations (i.e., does this person feel more sad than they usually do?) that can also be of high theoretical and practical relevance (J. Sun et al., 2020).

Third, recent research suggests that each communication channel has its specific linguistic peculiarities that can have an influence on affect-linked language variations and the algorithmic recognition of affect from language traces (Jaidka, Guntuku, & Ungar, 2018; Tingting Liu et al., 2022). Possibly, these language variations are partially caused by differences in emotional self-disclosure across communication channels (Bazarova, Taft, Choi, & Cosley, 2013). However, most prior research has been based on language from only one single communication channel, such as one social media platform, due to the challenges associated with gathering textual data from the same user across platforms (e.g., Twitter posts and WhatsApp messages).

Novel data collection methods leveraging off-the-shelf smartphones allow researchers to overcome the aforementioned challenges. First, using smartphone-based experience sampling, subjective affective states can be assessed in-situ when they are experienced using self-reports. Thereby, there is no need for data labeling. Second, since smartphones have become our main point of communication, they allow researchers to log the majority of text one produces on a given day. Third, smartphones offer a promising opportunity to passively log textual data across communication channels (public and private contexts) by directly logging what is typed on the smartphone's keyboard or by taking screenshots (Bemmann & Buschek, 2020; Brinberg et al., 2021).

Based on recent developments in smartphone-based language logging methods, this work investigates how affect experience is revealed through language patterns in different time frames and communication contexts. Specifically, we investigate (in-sample) associations of self-reported affect with language features logged with smartphones in everyday life and if these features allow for the (out-of-sample) prediction of between-person differences and within-person fluctuations in affect experience. Further, we analyze which language features are most strongly associated with affect experience. Thereby, we aim to advance affect recognition from language cues and inform psycholinguistic theory of affect.

3.3 Method

3.3.1 Data Set

Data collection for this work was part of a large six-month panel study (from May until November 2020) based on the *PhoneStudy* research app at Ludwig-Maximilian-Universität München (Schoedel & Oldemeier, 2020). Data collection was approved by the responsible IRB board. The study comprised two two-week experience sampling phases (July 27, 2020, to August 9, 2020; September 21, 2020, to October 4, 2020) during which participants received two to four short questionnaires per day. Here, self-reported valence and arousal were assessed in two separate items on six-point Likert scales among other psychological properties as part of an experience sampling procedure. Further, trait affect had been assessed using the German version of the Positive and Negative Affect Schedule (PANAS) ($alpha_{\mathrm{PA}} = .92$, $alpha_{\mathrm{NA}} = .89$) at the beginning of the study period (Breyer & Bluemke, 2016).

To capture the text participants had typed on their phones, the PhoneStudy app also included a keyboard logging module that had been adapted from the *ResearchIME* research app (Bemmann & Buschek, 2020). Hereby, all typed words were categorized according to the German "SentiWS" sentiment dictionary (Remus, Quasthoff, & Heyer, 2010) and the latest German version of the Linguistic Inquiry and Word Count (Meier et al., 2019) directly on the device. After the respective word categories had been logged, the raw text was deleted to protect participants' privacy. Additionally, emoji and emoticons were logged in the clear. Also, for each text input, the respective app, a time stamp, and the input prompt text (e.g., "What's on your mind?" on Facebook) were logged.

We aggregated logged text data over four different time frames that correspond to assessed self-reports: trait, weekly, daily, and momentary. Regarding trait affect, we aggregated all text produced during the study period and matched it with the assessed scores for positive and negative trait affect. For weekly affect, we aggregated valence and arousal scores over one week (Monday through Sunday) for participants who had filled out at least one experience sampling questionnaire per day for that given week and coupled them with all logged text that had been produced in that week. With regard to daily affect, we computed the median valence and arousal for those days where participants had filled out at least two experience sampling instances and paired them with all text produced on that particular day. For momentary affect, we aggregated all text typed in a three-hour time window around (90 minutes before and after) a single experience sampling instances as done in prior research (J. Sun et al., 2020). Furthermore, for the same three-hour time window, we computed the fluctuation of assessed momentary affect in valence and arousal from one's (median) affect baseline (for participants with at least five experience sampling days) across all experience sampling instances. For example, if a participant had a valence baseline of "3" and reports a "6" in a particular moment, this fluctuation score of "+3" indicated that this person had been a lot more happy than usual. The baseline score for valence showed a correlation of 0.44 with positive trait affect and -0.31 with negative trait affect.

To distinguish affect-linked language variations across communication channels, we used the logged information on the app that had been used to produce the respective text (e.g., Facebook, Instagram, or WhatsApp) and the input prompt text (e.g., "What's on your mind?" on Facebook). Here, we manually categorized all input prompts across all logged apps to determine the context in which the respective text had been written. For example, if a participant had produced text in the Facebook app and the logged input prompt text was "What's on your mind?", the composed text would be assigned to public communication. On the contrary, if she had used the search bar (i.e., the prompt was "Search Facebook") the produced text would neither count as private nor public communication. Thereby, we were able to generally detect and differentiate between private and public communication on a granular level. We conducted the public versus private communication analyses for trait affect (i.e., all text produced during the entire study period) because a sufficient number of participants had created enough text data to be statistically analyzed.

In our language analyses regarding trait affect (all text and private versus public communication) and weekly affect experience, we included participants who had written more than 500 words in the respective time frame. This threshold had been also applied in similar previous work on text messages (Tony Liu et al., 2021). For daily and momentary affect, we used a threshold of 100 words as used in related prior research (Vine et al., 2020).

643 participants produced any text in the entire study period. 445 of those had filled out the PANAS questionnaire and wrote at least 500 words of text on their smartphone keyboards. Of those, 287 met the minimum word threshold of 500 words in private communication and 127 in public communication contexts. Thus, they were included in the analyses with regard to trait affect. For weekly affect, we included 169 participants, who had filled out one experience sampling questionnaire each day in a given week (Monday through Sunday) and wrote more than 500 words of text on their smartphones. With regard to daily affect, 289 participants had completed at least two experience sampling assessments and 100 words of written text for a given day. Finally, we had 755 experience sampling instances with at least 100 words of corresponding text from 183 participants. Overall, participants' self-reported positive trait affect was balanced ($M = 3.13$, $SD = 0.74$) and negative trait affect ($M = 1.88$, $SD = 0.7$) was low. Moreover, self-reported momentary valence was positive ($M = 4.54$, $SD = 1.1$) and overall arousal was slightly geared towards activity ($M = 3.9$, $SD = 1.28$). Table 3.1 provides an overview of descriptive statistics of the final samples across communication contexts and different time windows. The distribution of affect measures is shown in Figure 3.9 in the chapter's appendix in section 3.10.

Table 3.1: Overview of data (sub) sets used in study 2

Time frame	Trait	Trait	Trait	Week	Day	Moment
Communication context	All	Private	Public	All	All	All
$N_{\text{instances}}$	445	287	127	222	2318	755
$N_{\text{participants}}$	445	287	127	169	289	183
M(age)	41.66	41.88	41.64	38.32	39.61	37.55
%women	43.96	42.21	55.08	57.21	53.72	54.61
M(Positive affect)	3.13	3.10	3.13			
M(Negative affect)	1.88	1.88	1.87			
M(Valence)				4.85	4.65	4.54
M(Arousal)				4.06	4.04	3.90
M(Typing sessions)	2166.74	698.75	208.23	177.24	32.31	15.02
M(Words)	19564.45	8696.42	2357.81	1744.83	332.77	220.37

3.3.2 Keyboard Language Analyses

Users reveal affective information in language typed on smartphones through the words they use, emoji and emoticon use, and typing dynamics. Therefore, we extracted three groups of features to comprehensively characterize participants' language logs regarding language content and form. First, we used word dictionaries (LIWC and the German SentiWS sentiment lexicon) to categorize written words directly on the device. Second, we logged participants' use of emoji and emoticons and computed respective metrics. Third, we analyzed their writing form by extracting features regarding participants' typing dynamics. In this section, features from each feature grouped are described in detail. As mentioned in the previous section, we aggregated language features across four different time frames: trait, weekly, daily, momentary. For each time frame, we aggregated all text that had been produced during all *typing sessions* (i.e., when the smartphone keyboard was opened, and text had been produced) and computed the respective features. For example, for daily affect experience, all text from a given participant for that day was aggregated and the daily sentiment score was computed. For our descriptive analyses, we estimated the magnitude of the associations of extracted features with affect experience (e.g., daily valence) using pairwise Pearson correlation.

Word Dictionaries

We used the latest German word dictionaries from the well-established Linguistic Inquiry and Word Count (LIWC) (Meier et al., 2019; Pennebaker et al., 2015). LIWC dictionaries, such as *1st person singular* or *past focus*, have been theoretically derived and are clustered in a hierarchical structure: For example, the *sadness* category is a subcategory of *negative emotion*, which in turn is a subcategory of *affective processes*. Here, we computed the share of words from a given dictionary category (e.g., positive emotion words) from all written words for a given time window. Since LIWC has the same word categories across languages, we can compare word category scores for our German text data with prior studies that analyzed, for instance, English text data.

Moreover, we used the German SentiWS sentiment dictionary to obtain additional sentiment scores for logged words (Remus et al., 2010). SentiWS contains 1650 negative and 1818 positive words with respective word forms and corresponding sentiment weights within the interval of [-1; 1]. Using those scores, we computed the median, standard deviation, minimum, and maximum word sentiment across typing sessions for a given time window.

Emoji and Emoticons

With regard to participants' overall use of emoji and emoticons, we calculated the mean number of emoji and emoticons, the emoji- and emoticon-to-word ratios, and the number of unique emoji and emoticons across typing sessions for a given time frame. Additionally, we enriched logged emoji with a sentiment score within the interval of [-1; 1] from an emoji sentiment data base (Kralj Novak et al., 2015). In the same manner as for word sentiment, we computed the median, standard deviation, minimum, and maximum emoji sentiment of typing sessions in a given time frame.

Like in prior work (Koch, Romero, et al., 2022), we kept specific emoji and emoticons that had been used by at least 5% of participants in the whole sample of 643 participants to keep the focus on common emoji and emoticons. For those frequently used emoji and emoticons, we then counted how often each participant had used the respective specific one in a given time window and normalized their frequency use by dividing the count by the total number of emoji/ emoticons used by the respective participant in the time window.

Typing Dynamics

We computed a range of features that describe the typing dynamics of how logged text had been produced. Specifically, we calculated the median and standard deviation of the duration per typing session and per single word. Also, we computed the median and standard deviation of the number of words and characters that have been written per typing session. Finally, we calculated the median and standard deviation of the number of words that had been removed per typing session. Additionally, we utilized logged information from app use and input prompt texts to compute the share of text from all produced text that had been typed in different contexts (e.g., in private communication) and actions (e.g., commenting). Hereby, for instance, we were able to determine what share of all text that had been produced by a participant in a given time window had been typed in social media apps.

3.3.3 Predictive Modelling

We trained two supervised machine learning algorithms on the extracted features for the prediction of self-reported affect experience. Specifically, we compared the predictive performance of linear regularized regression models (LASSO) (Tibshirani, 1996) with those of a non-linear tree-based Random Forest (Breiman, 2001; Wright & Ziegler, 2017), and a baseline model. The baseline model would predict the respective mean value for the target variable (e.g., daily valence) of the respective training set for all cases in a test set. Additionally, we included the prediction of participants' age and gender from all collected text data as a benchmark. Models were evaluated using a ten-fold ten times repeated cross-validation scheme (Bischl et al., 2012). For those models, where there were multiple instances per participant (e.g., multiple experience sampling days per single participant), we blocked participants in the resampling procedure to ensure that each participant is either included in the training or test set.

We evaluated the predictive performance of trained models based on the coefficient of determination (R^2) and Spearman's rank correlation (r). Before predictive modeling, we excluded features with more than 90% missing values and constant or near-constant features (i.e., with less than 5% variance). Moreover, we median-imputed extreme outlines (more than mean +/- 4 times *SD*) in the cross-validation procedure (Schoedel et al., 2022). To determine whether a model was predictive beyond chance ($alpha =$ 0.05), we carried out variance-corrected (one-sided) t-tests comparing the R^2 measures of all prediction models with those of the baseline models (Nadeau & Bengio, 2003). We adjusted for multiple comparison ($n = 30$) via Holm correction. All data processing and statistical analyses in this work were performed with the statistical software R version 4.1.1 (R Core Team, 2021). For machine learning, we used the *mlr3* framework (Lang et al., 2019). Specifically, we used the *glmnet* (Friedman et al., 2010) and *ranger* (Wright & Ziegler, 2017) packages to fit prediction models. We pre-registered our analyses before accessing the data as a transparent account of our work (Koch, Eichstaedt, & Stachl, 2022).

3.4 Results

3.4.1 Language Variations with Subjective Affect Experience

We found a range of affect-linked language variations across communication contexts and time frames in the collected keyboard language data. In the following section, we report on all pairwise Pearson correlations of language features with affect experience where the respective 95% confidence interval did not contain zero.

Typing Dynamics

We found distinct associations that are considered at least small (i.e., $r > .1$). of typing dynamics with trait affect experience only in private communication in the present data: Longer typing sessions (i.e., longer duration) (r = -0.15) and more text per typing session (r = -0.14) were negatively associated with negative trait affect. The variation (*SD*) in the text length per typing session measured in words (r= -0.13) and characters (r = -0.14) had a negative correlation with positive trait affect. Also, low typing speech as measured by the median typing duration per word was correlated with negative trait affect (r = 0.16). Where private messages had been produced also correlated with trait affect: Private messages composed in social media apps (e.g., Facebook) were positively associated with negative trait affect (r = 0.13) while the correlation with negative trait affect was negative for text produced in communication apps (e.g., WhatsApp) (r = -0.14).

With regard to weekly affect, it mattered most where text had been produced: We found a negative correlation of the share of text produced for data input (e.g., filling out online forms) and weekly valence (r = -0.13). Also, weekly arousal was positively correlated with producing text on internet apps (e.g., browser apps) (r = 0.21) and searching online (r = 0.16). On the contrary, the variation (*SD*) in the character count per typing session correlated negatively with weekly arousal (r = -0.13). For daily affect, we only discovered a small correlation of daily valence and the share of text that had been produced inputting data (r = -0.15). Finally, for momentary affect experience, there was a small negative correlation of the share of total text produced by posting online (e.g., on Facebook) with the fluctuations from one's valence baseline (r = -0.11) and of valence with the typing duration per word (r = -0.1).

Word Dictionaries

Different language patterns related to trait affect emerged from the logged word dictionaries in our data set (see Figure 3.1). Moreover, Figure 3.2 depicts how the use of theoretically relevant (emotion) dictionaries differs with communication contexts for trait affect.

With regard to *direct* linguistic markers of emotion, we found that median word sentiment was positively associated with positive trait affect in all logged text (r = 0.16) with respective correlation coefficients being smaller for private communication (r = 0.09) and even negative for public communication (r = -0.09). Negative emotion words were negatively correlated with positive trait affect (r = -0.1) and positively with negative trait affect (r = 0.18) across all text inputs and in the sub-contexts. The same correlational patterns held true for the anger and anxiety dictionaries that are sub-dictionaries of negative emotions words. Positive emotion words did not show any relevant correlations with trait affect in any communication context in our data.

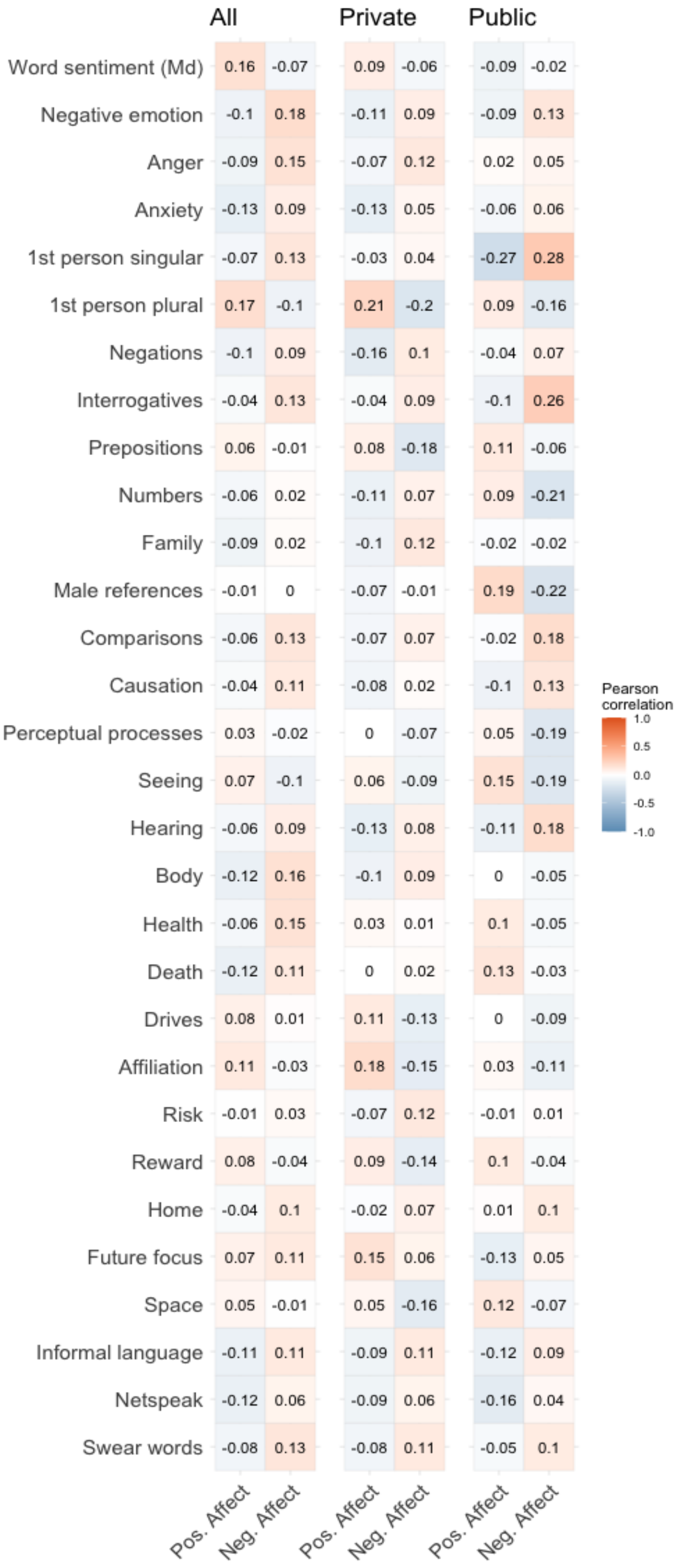


Figure 3.1: Correlations of dictionaries with trait affect for all logged text and public and private communication separately. We considered coefficients where the 95% confidence interval did not contain zero.

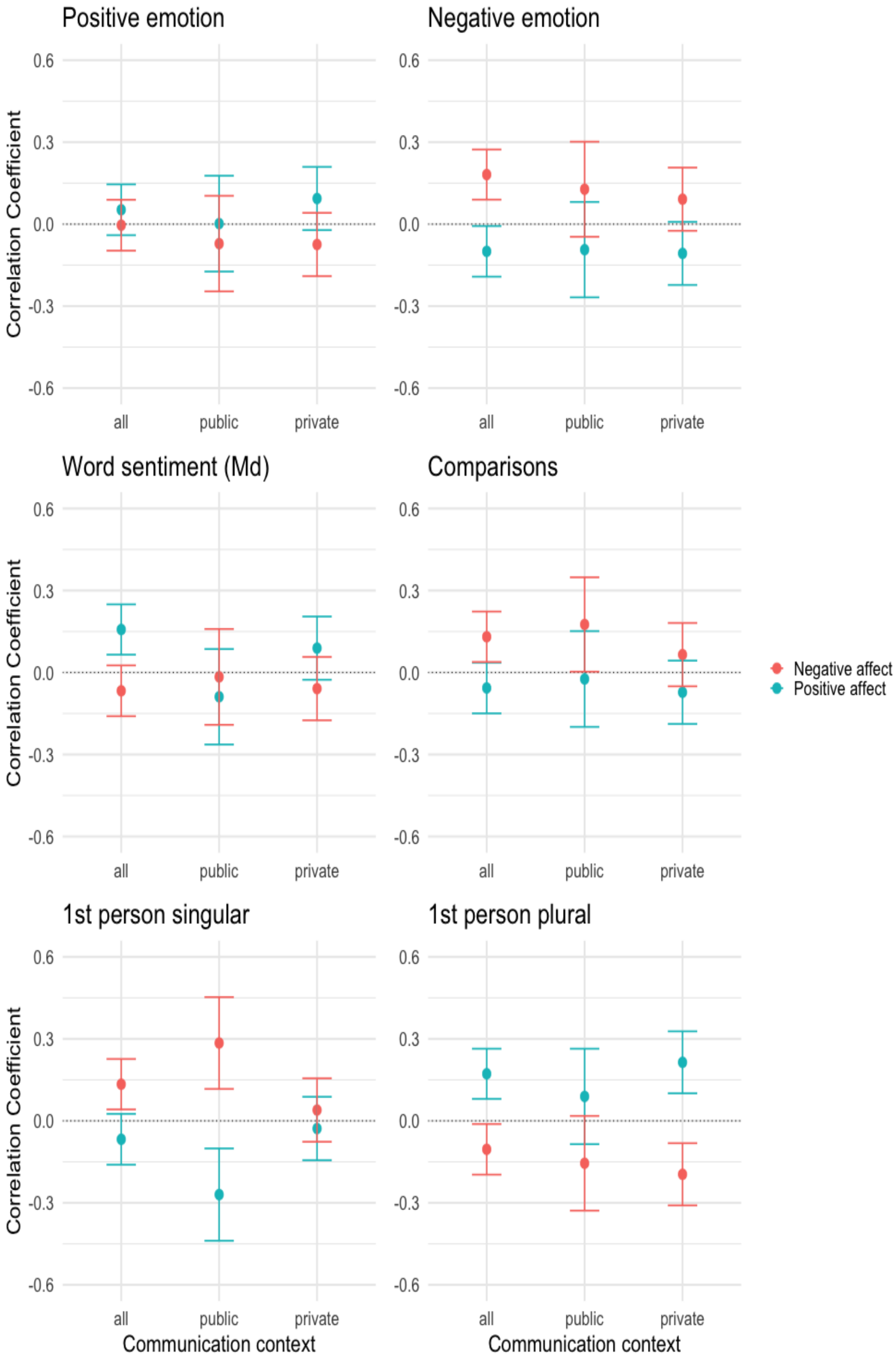


Figure 3.2: Selected correlations of dictionaries with trait affect for all logged text and public and private communication separately.

We found a range of distinct correlations of *indirect* linguistic markers of emotion with trait affect. First, the overall use of first person singular (e.g., “I”, “me”) was negatively correlated with positive trait affect (r = -0.07) and positively with negative trait affect (r = 0.13). Here, correlation coefficients were much bigger for public communication (r_{PA} = -0.27, r_{NA} = 0.28) than for private communication (r_{PA} = -0.03, r_{NA} = 0.04). On the contrary, the use of first person plural (e.g., “we”) was positively correlated with positive trait affect (r = 0.17) and negatively with negative trait affect (r = 0.17). These associations were stronger for private communication (r_{PA} = 0.21, r_{NA} = -0.2) than for public communication (r_{PA} = 0.09, r_{NA} = -0.16).

Moreover, using comparisons (e.g., “similar”) was associated positively with negative trait affect overall (r = 0.13) and particularly in public communication (r = 0.18). Also, the use of interrogatives (e.g., “why”) had a positive correlation with negative trait affect overall (r = 0.13), again especially in public communication.

In the same manner, two sub-categories of social processes, showed distinct correlational patterns: Male references had a positive correlation with positive trait affect (r = 0.19) and a negative one with negative trait affect (r = -0.22) in public communication. Family-related words correlated positively with negative trait affect (r = -0.1) and negatively with positive trait affect (r = -0.1) in private communication. Also, words related to the drive for affiliation (e.g., “friend”) had a positive correlation with positive trait affect overall (r = 0.11), particularly in private communication.

Words related to perceptual processes, in particular regarding seeing and hearing, were related to trait affect with the strongest effects in public communication: Here, using words from the seeing dictionary was positively correlated with positive trait affect (r = 0.15) and negatively correlated with negative trait affect (r = -0.19) in contrast to words from the hearing dictionary that were positively correlated to negative trait (r = -0.11) affect and negatively to positive trait affect (r = -0.11).

Furthermore, two sub-categories of the biological processes category correlated with trait affect: Words related to body (e.g., “head”) (r_{PA} = -0.12, r_{NA} = 0.16) and health (e.g., “drug”) (r_{PA} = -0.06, r_{NA} = 0.15) correlated negatively with positive trait affect and positively with negative trait affect. Also, the use of words regarding death (r_{PA} = -0.12, r_{NA} = 0.11) showed the same correlation pattern. Remarkably, the health and death word categories had a positive correlation with positive trait affect in public communication.

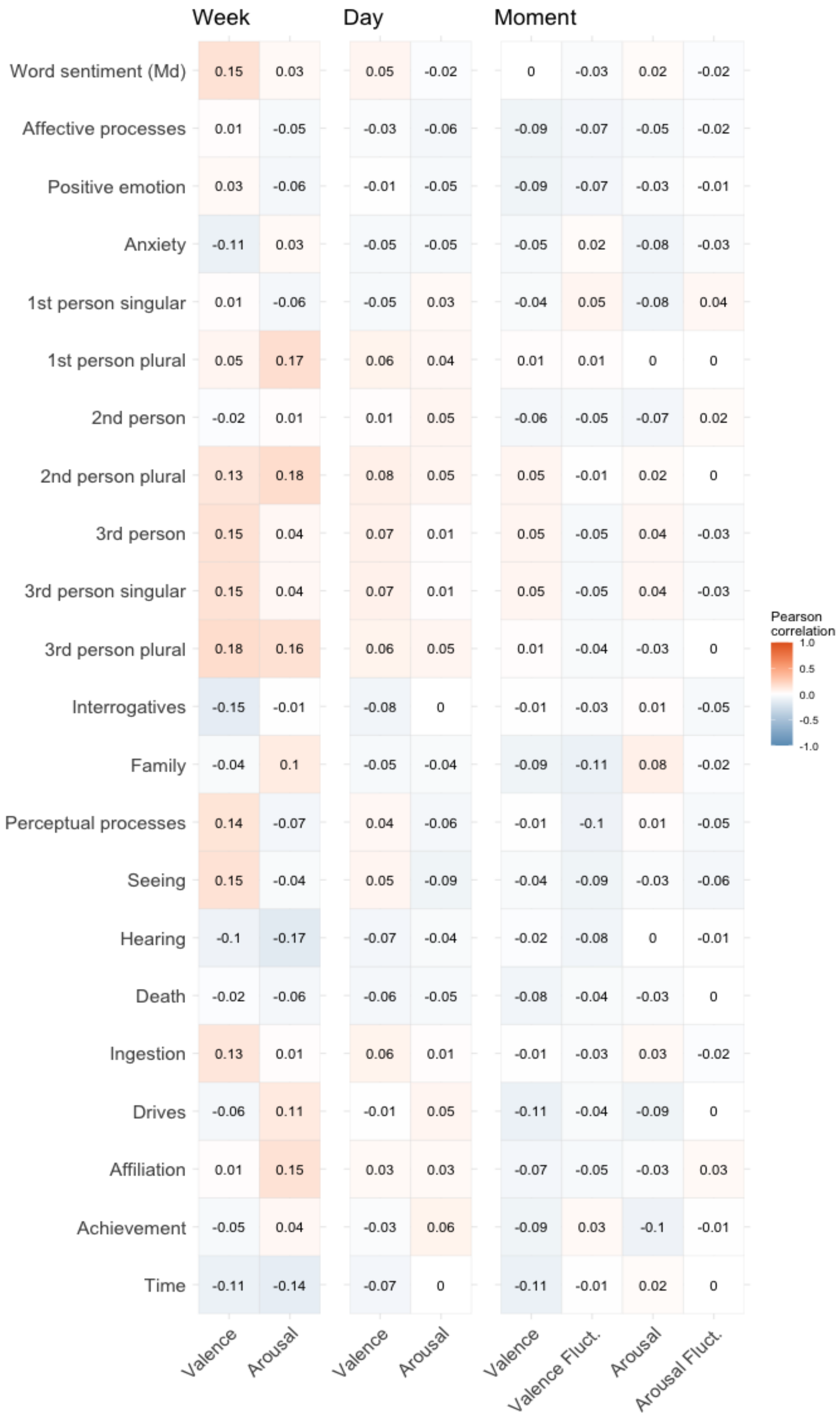
Finally, the overall use of informal language (r_{PA} = -0.11, r_{NA} = 0.11) and of its two sub-categories swear words and netspeak correlated positively with negative trait affect and negatively with positive trait affect across communication channels.

Weekly, Daily, and Momentary Affect Overall, correlations coefficients for weekly, daily, and momentary affect experience with word dictionaries were smaller than for trait affect. Moreover, with the time window getting smaller, the size of correlation coefficients also decreased. See Figure 3.3 for an overview of all dictionaries for which for any two time windows zero was not in the 95% confidence interval of the correlation coefficient. Moreover, Figure 3.4 depicts a detailed overview of theoretically relevant (emotion) dictionaries and their correlations with self-reported valence over time (weekly, daily, momentary).

Regarding *direct* linguistic markers of emotion, in the same fashion as for trait affect, the median word sentiment correlated positively with weekly valence (r = 0.15) and the use of anxiety-related words correlated negatively with weekly valence (r = -0.11). For daily and momentary affect experience, correlation coefficients of emotion dictionaries with valence or arousal got much smaller. Remarkably, positive emotion words even correlated slightly negatively with momentary valence and valence fluctuation.

Regarding *indirect* markers of emotion, function words in particular, showed a distinct correlation pattern. Weekly valence correlated positively with the use of 2nd person plural (r = 0.18), overall use of 3rd person (r = 0.15) and its sub-categories 3rd person singular (r = 0.15) and 3rd person plural (r = 0.18). In a similar fashion, weekly arousal correlated positively with 1st person plural (r = 0.17), 2nd person plural (r = 0.18), and 3rd person plural (r = 0.16).

In line with findings regarding trait affect, the use of interrogatives had a negative correlation with weekly valence (r = -0.15). The use of words related to family correlated positively with weekly arousal (r = 0.1) and negatively with momentary valence (r= -0.09) and valence fluctuation (r = -0.11). In the same fashion as for trait affect, we found perceptual processes (r = 0.14) and its sub-category seeing (r = 0.15) and hearing (r = -0.1) to be correlated with weekly valence. Further, hearing was correlated negatively with weekly arousal (r = -0.17). Moreover, the ingestion word category (e.g., "eating") had a positive correlation with weekly valence (r = 0.13). Further, words related to drives (r = 0.11), particularly the drive for affiliation (r = 0.15) had a positive correlation with weekly arousal. Moreover, momentary valence correlated negatively with drives (r = -0.11) and its sub-dictionary achievement (e.g., "victory") (r = -0.09). Finally, the use of words related to time correlated negatively with weekly valence (r = -0.11) and arousal (r = -0.14) as well as momentary valence (r = -0.11).



	Week		Day		Moment			
	Valence	Arousal	Valence	Arousal	Valence	Valence Fluct.	Arousal	Arousal Fluct.
Word sentiment (Md)	0.15	0.03	0.05	-0.02	0	-0.03	0.02	-0.02
Affective processes	0.01	-0.05	-0.03	-0.06	-0.09	-0.07	-0.05	-0.02
Positive emotion	0.03	-0.06	-0.01	-0.05	-0.09	-0.07	-0.03	-0.01
Anxiety	-0.11	0.03	-0.05	-0.05	-0.05	0.02	-0.08	-0.03
1st person singular	0.01	-0.06	-0.05	0.03	-0.04	0.05	-0.08	0.04
1st person plural	0.05	0.17	0.06	0.04	0.01	0.01	0	0
2nd person	-0.02	0.01	0.01	0.05	-0.06	-0.05	-0.07	0.02
2nd person plural	0.13	0.18	0.08	0.05	0.05	-0.01	0.02	0
3rd person	0.15	0.04	0.07	0.01	0.05	-0.05	0.04	-0.03
3rd person singular	0.15	0.04	0.07	0.01	0.05	-0.05	0.04	-0.03
3rd person plural	0.18	0.16	0.06	0.05	0.01	-0.04	-0.03	0
Interrogatives	-0.15	-0.01	-0.08	0	-0.01	-0.03	0.01	-0.05
Family	-0.04	0.1	-0.05	-0.04	-0.09	-0.11	0.08	-0.02
Perceptual processes	0.14	-0.07	0.04	-0.06	-0.01	-0.1	0.01	-0.05
Seeing	0.15	-0.04	0.05	-0.09	-0.04	-0.09	-0.03	-0.06
Hearing	-0.1	-0.17	-0.07	-0.04	-0.02	-0.08	0	-0.01
Death	-0.02	-0.06	-0.06	-0.05	-0.08	-0.04	-0.03	0
Ingestion	0.13	0.01	0.06	0.01	-0.01	-0.03	0.03	-0.02
Drives	-0.06	0.11	-0.01	0.05	-0.11	-0.04	-0.09	0
Affiliation	0.01	0.15	0.03	0.03	-0.07	-0.05	-0.03	0.03
Achievement	-0.05	0.04	-0.03	0.06	-0.09	0.03	-0.1	-0.01
Time	-0.11	-0.14	-0.07	0	-0.11	-0.01	0.02	0

Figure 3.3: Correlations of dictionaries with affect for logged text across weekly, daily, and momentary time frames. We considered coefficients where the 95% confidence interval did not contain zero for at least two time windows.

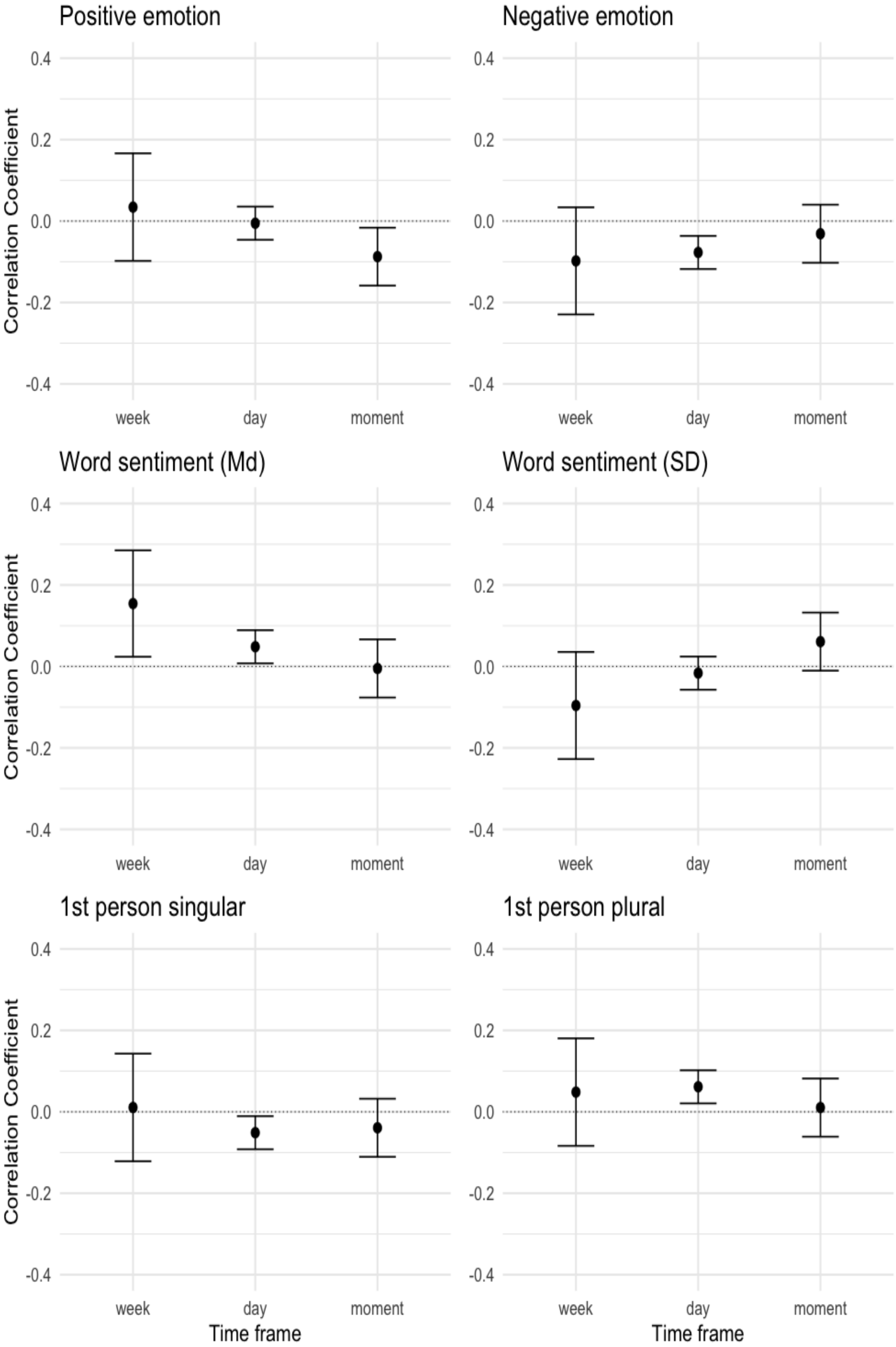


Figure 3.4: Selected correlations of dictionaries with affective valence for logged text across weekly, daily, and momentary time frames.

Emoji and Emoticons

This section presents discovered distinct patterns of general emoji and emoticon use across communication contexts and time frames that can be considered at least small (i.e., $r > .1$). For trait affect, the overall use of emoticons in particular was related to negative trait affect: Across all typed text, the number of unique emoticons used was negatively correlated with positive trait affect (r = -0.12 and positively with negative trait affect (r = 0.1). Also, in private communication, the total use of unique emoticons was positively associated with negative trait affect (r = 0.11) and negatively with positive trait affect (r = -0.19). In the same manner, the number of unique emoticons per typing session in private communication had a negative correlation with positive trait affect (r = -0.12). Furthermore, the standard deviation in emoji sentiment across typing sessions was negatively correlated with negative trait affect in private messages (r = -0.15). In public communication contexts, the maximum emoji sentiment across typing sessions was associated negatively with negative trait affect (r = -0.23).

With regard to weekly affect, we found a negative correlation of unique emoticons per typing session and weekly arousal (r = -0.17). For daily valence and arousal, no such patterns emerged from the data. Finally, there was a negative correlation of the total use of unique emoticons and momentary affective valence (r = -0.12).

While the overall use of emoticons was indicative of affect, the use of specific emoticons was not. On the contrary, preferences for specific emoji varied with affect experience. Figure 3.5 shows emoji that are associated with positive and negative trait affect across communication channels. Across communication contexts, the dog emoji (“🐶”) was correlated positively with positive trait affect and slightly negatively with negative trait affect. Also, “😇”, “🥰”, and “😉” had a positive correlation with positive trait affect and a negative one with negative trait affect. On the contrary, the “🤦‍♂️” emoji was correlated negatively with positive trait affect and positively with negative trait affect in all logged text and in private communication.

In the same manner as for word dictionaries, correlations of specific emoji were less conclusive for smaller time windows. Figure 3.6 depicts correlations of emoji with weekly and daily valence and arousal. Figure 3.7 shows emoji correlations with between-person differences and within-person fluctuations in momentary affect. Across time windows, “🌈” was negatively correlated with valence and arousal. Also, “💋” correlated negatively with daily and momentary valence and arousal. On the contrary, “😇” had a positive correlation with weekly and daily valence and a negative one with respective arousal.

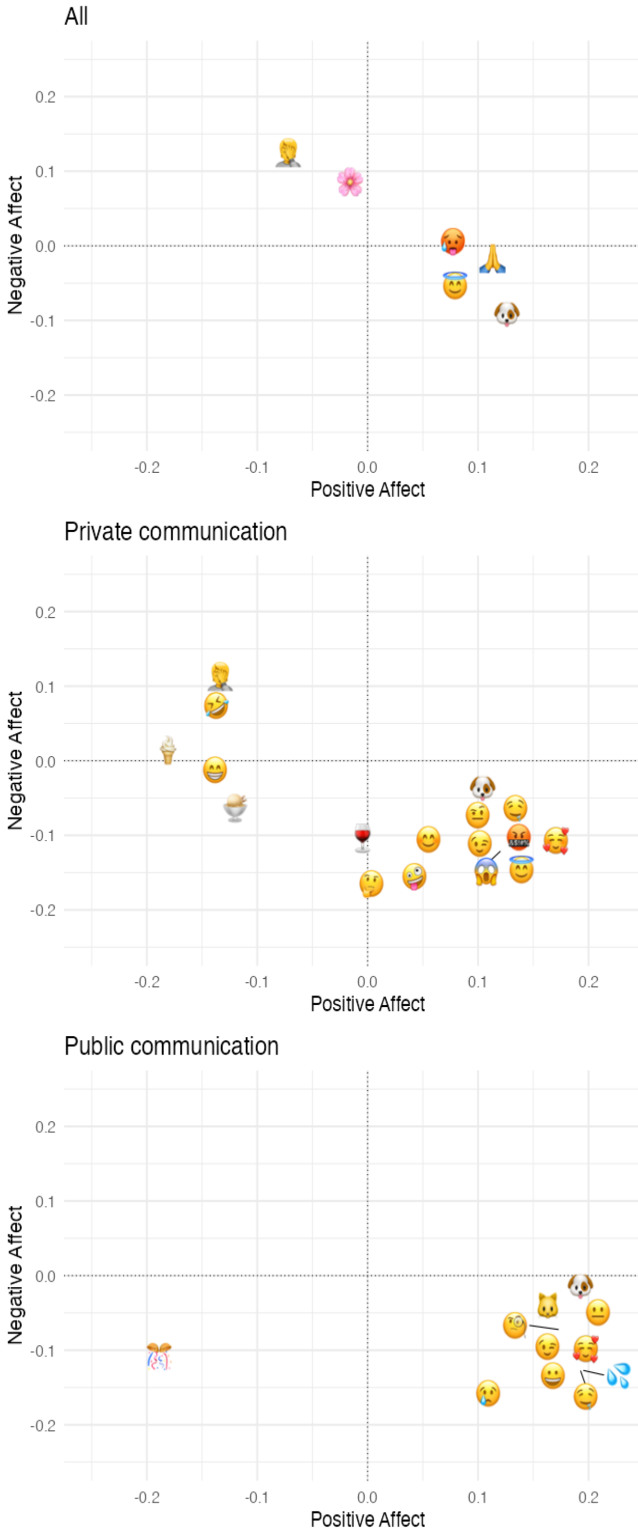


Figure 3.5: Correlations of specific emoji with trait affect on the dimensions of positive affect and negative affect for all produced text and private and public communication separately. Displayed are all emoji where the 95% confidence interval for the correlation coefficient did not contain zero.

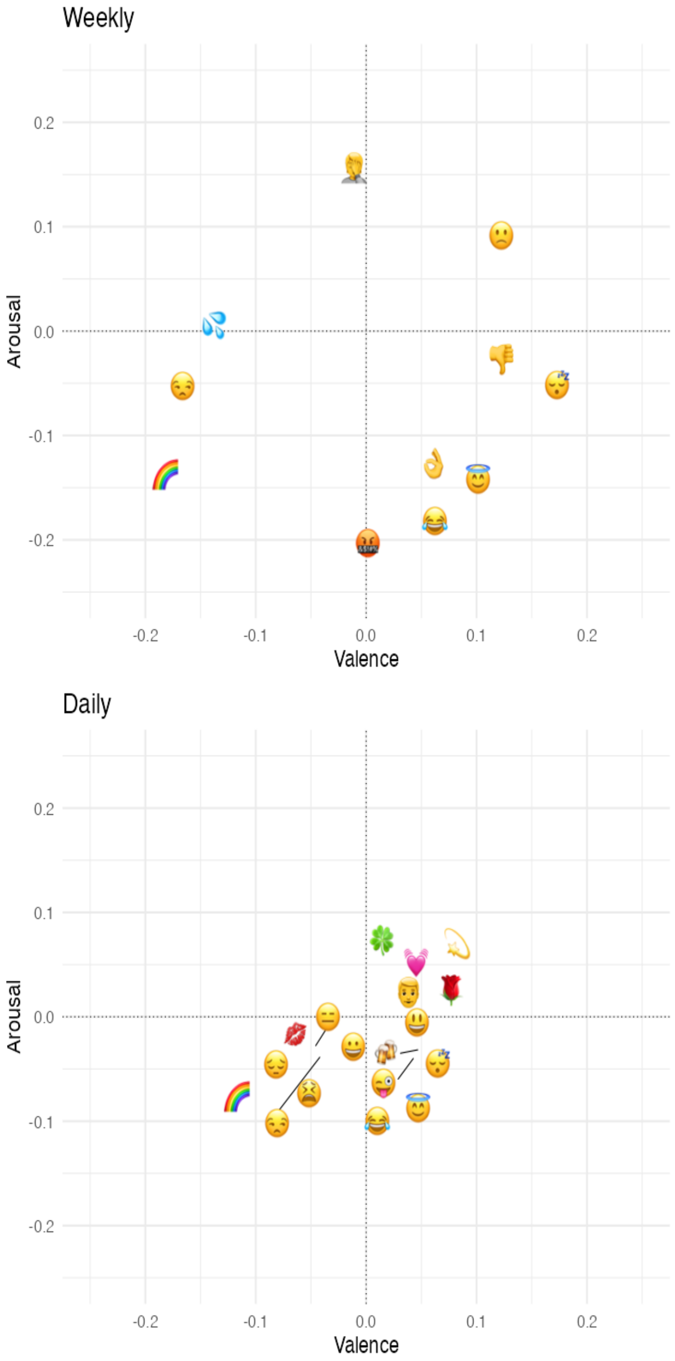


Figure 3.6: Correlations of specific emoji with affect experience on the dimensions of valence and arousal across weekly and daily time frames. Displayed are all emoji where the 95% confidence interval for the correlation coefficient did not contain zero.

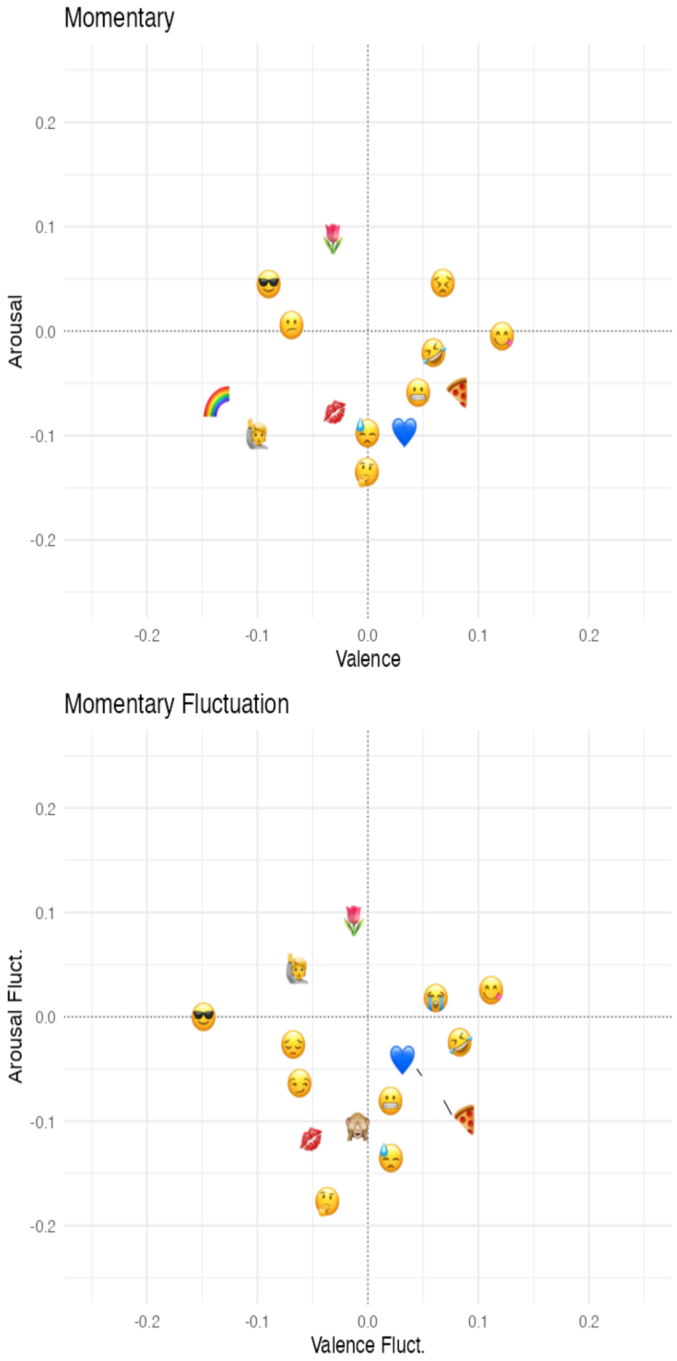


Figure 3.7: Correlations of specific emoji with affect experience on the dimensions of momentary valence and arousal as well as respective affect fluctuations. Displayed are all emoji where the 95% confidence interval for the correlation coefficient did not contain zero.

3.4.2 Predicting Subjective Affect Experience

Overall, none of the employed algorithms predicted affect experience in any communication context or time frame significantly better than chance. Figure 3.8 provides an overview of the performance of all learners across prediction tasks.

On average, Random Forest models performed slightly better than the LASSO algorithm across prediction tasks, indicating non-linear predictor-outcome relationships between keyboard language characteristics and affect experience across communication contexts and time frames. Consequently, we will report on the average performance of Random Forest models across cross-validation folds in this section.

With regard to communication contexts, even though not better than chance, predictions of trait affect were slightly better for private communication ($R^2{}_{\text{PA}}$ = -0.11, r_{PA} = 0.07, $R^2{}_{\text{NA}}$ = -0.11, r_{NA} = 0.06) than for public communication ($R^2{}_{\text{PA}}$ = -0.33, r_{PA} = 0.03, $R^2{}_{\text{NA}}$ = -0.3, r_{NA} = 0.04). Still, in comparison, predictions on all text data combined performed best ($R^2{}_{\text{PA}}$ = -0.05, r_{PA} = 0.07, $R^2{}_{\text{NA}}$ = -0.06, r_{NA} = 0.09).

With regard to the time frame, predictions were more accurate for bigger time windows, but still not better than chance. For instance, the prediction performance for weekly valence (R^2 = -0.06, r = 0.1) and arousal (R^2 = -0.06, r = 0.06) was, on average, better than for momentary valence (R^2 = -0.27, r = -0.02) and arousal (R^2 = -0.12, r = 0).

Finally, predictions of participants' age (R^2 = 0.46, r = 0.72) and gender (prediction accuracy = 73.95%) from all logged text suggest that typed smartphone language allows for inferences about user demographics.

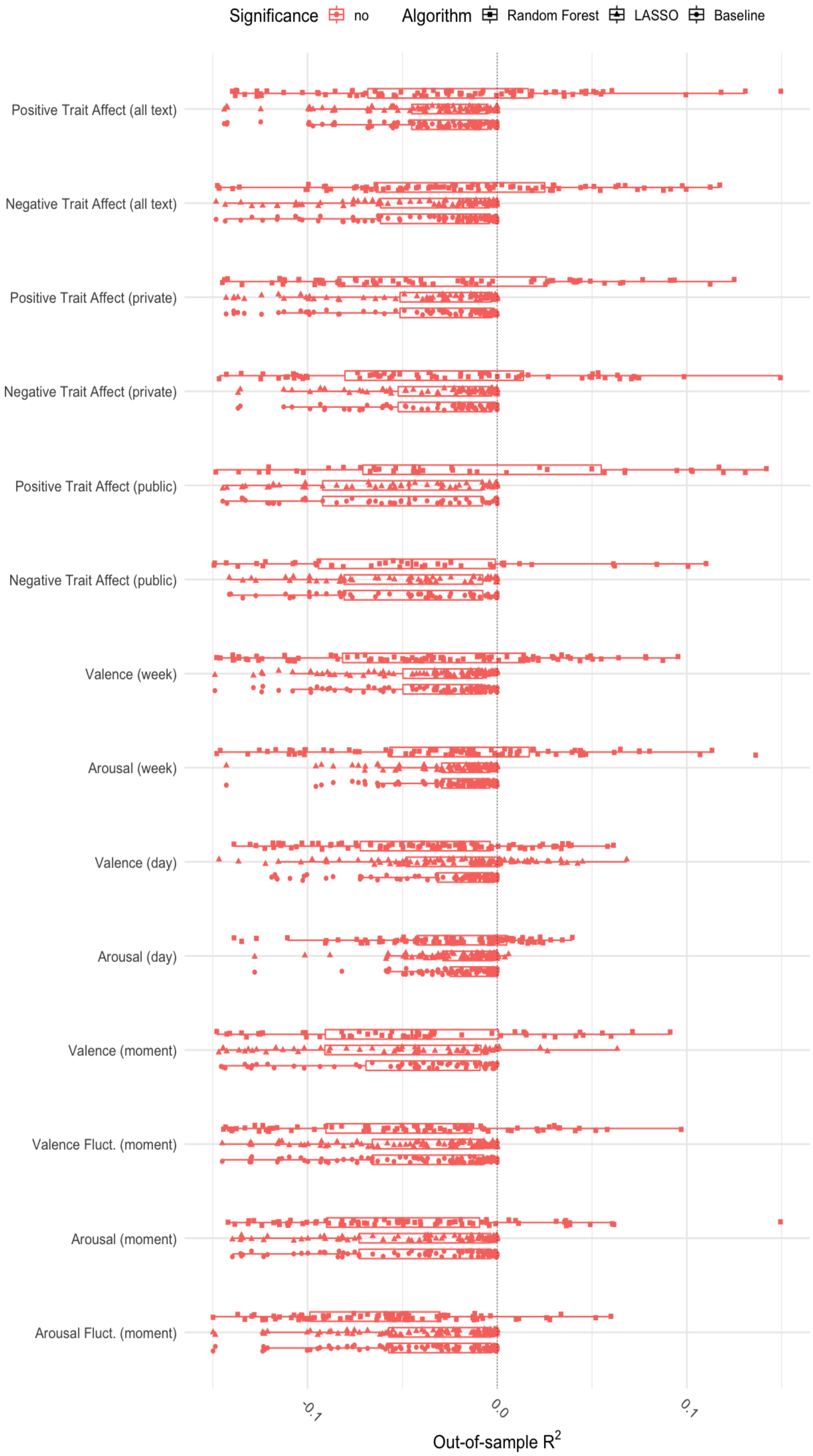


Figure 3.8: Prediction performance measures of prediction models from 10-fold ten times repeated cross-validation for affect predictions for each feature (sub) set and across communication contexts and time frames.

3.5 Discussion

The present study generated novel insights into affective language patterns across communication contexts and time frames by analyzing everyday textual language that had been logged using off-the-shelf smartphones. For our descriptive and predictive analyses, we leveraged typing dynamics, word (sentiment) dictionaries, and emoji and emoticon use. Beyond the distinctive (in-sample) associations of specific language features with affect experience we found, (out-of-sample) predictions of affect experience from the extracted language cues were, however, not significantly better than chance.

3.5.1 Affect Experience in Smartphone Language

In line with prior literature, we found typing speed (i.e., typing duration per word) to be associated with (negative) affect experience (Ghosh et al., 2017; Ghosh et al., 2019). Moreover, in our data, we found the context (i.e., social media apps versus communication apps) in which language had been produced to be indicative of affect experience.

Our findings with regard to *direct* linguistic markers of emotion, even though correlations were small, are in line with the theory that emotional word use corresponds to affect experience: We found a positive correlation of word sentiment with positive trait affect and a negative one with negative trait affect. Moreover, the use of negative emotion words showed expected negative correlations with positive trait affect and positive ones with negative trait affect across channels, while positive emotion words did not yield distinct correlations with trait affect. For smaller time frames, however, those correlations decreased in size or vanished entirely.

We also discovered a range of *indirect* linguistic markers of emotion in our data. For instance, in line with prior literature, we found the use of first person singular (e.g., “I” or “me”), particularly in public communication, to be positively associated with negative affect and negatively associated with positive affect (Tackman et al., 2018). On the contrary, the use of first person plural (e.g., “we”) showed a negative correlation with negative trait affect and a positive correlation with positive trait affect as found in prior work (J. Sun et al., 2020). Beyond these well-studied LIWC categories, additional noteworthy associations of indirect linguistic markers of emotion emerged in our data. For instance, the use of comparisons and interrogatives was associated with negative trait affect, particularly in public communication. Also, the use of hearing-related words was indicative of negative trait affect, while seeing-related words were associated with a positive trait affect, again particularly in public communication.

After prior studies had used raters to assign affective labels to emoji and emoticons (Kralj Novak et al., 2015; Krekhov et al., 2022; Kutsuzawa et al., 2022), the present work is, to our knowledge, the first to analyze the link of emoji and emoticon use with subjective affect experience. Our findings suggest that the overall use of emoticons is related to negative trait and state affect experience. Regarding emoji, no such overall pattern emerged, but some specific emoji, such as "🐶" and "🥰", were found to be related to positive trait affect, while others (e.g., "🤦‍♂️") correlated with negative trait affect. Again, as for word dictionaries, findings were less conclusive regarding associations of affect experience and emoji use for smaller time windows. Here, particularly the "🌈" emoji that had been used as a symbol of hope during the COVID-19 pandemic was consistently negatively correlated with weekly, daily, and momentary valence and arousal.

3.5.2 Predicting Affect Experience from Everyday Smartphone Language

Even though the reported non-significant prediction performance for affect experience in this work is inferior to prediction studies based on labelled text data (Eichstaedt & Weidman, 2020) it is in line with the limited prior work that used written smartphone language and affective self-reports from the wild (Carlier et al., 2022). As a result, our findings suggest that algorithmically recognizing subjective affect experience, especially in small time windows, from everyday smartphone language is difficult, particularly using the occurrence-counts, such as word dictionaries, as employed on the present study. Still, leveraging digital language traces logged with smartphones has many potential upsides compared to other text sources, such as social media, since text can be logged across communication channels and over time allowing to also model within-person fluctuations of affect. However, more research is needed before the promising applications, for instance in mood monitoring for mental health care, can be reliably deployed.

3.5.3 Communication Context Matters

In line with prior research on language differences across communication channels, our results illustrate the importance of considering the communication context in language analysis (Mehl, Robbins, & Holleran, 2013). Depending on the communication context (private messages versus public posting), some language-affect associations were different in our data. For example, I-talk (first person singular) was particularly highly correlated with negative trait affect in public communication, but not in private communication. On the contrary, we-talk (first person plural) correlated higher with positive trait affect in private communication than in the public one.

Also, predictions, even though not significantly above chance, were slightly better for text produced in private contexts than public ones. This could be due to participants engaging in selective emotional self-disclosure in public communication, such as social media (Qiu, Lin, Leung, & Tov, 2012), making inferences of affect experience more challenging. The slightly better predictions from private communication could also come from the difference in sample sizes that were available for private and public communication with the latter being smaller. Overall, as a consequence from our findings, researchers and practitioners should always consider the communication context the text data had been produced in and keep in mind that findings and trained models might be specific to the given communication context and do not necessarily generalize well to other communication contexts.

3.5.4 Time Context Matters

To our knowledge, this work is the first of its kind to investigate how affect-linked language patterns change over different time frames. Depending on the time window, some affect-language associations changed in the analyzed data. For instance, median word sentiment correlated positively with positive trait affect and weekly affective valence but had a close to zero correlation with momentary valence. Consequently, researchers and practitioners should always consider the time scale (e.g., one's overall trait affect or affect experience in a specific moment) which they want to investigate affect-linked language in. As our results have shown, the chosen time focus might have a strong impact on the findings.

Furthermore, for smaller selected time windows, for instance momentary instead of weekly affect, descriptive associations (i.e., correlation coefficients) and models' prediction performance were also weaker in our data. These findings illustrate that assessing affect experience, particularly on smaller time scales, using the current occurrence-count approaches, such as word dictionaries, is limited. Specifically, our results indicate that occurrence-based metrics (i.e., counting negative emotion words or single emoji) work to an extent for traits, but reach their limits for smaller time windows, possibly since they do not consider language context. For instance, if participants wrote “not happy” LIWC would count one negation and one positive emotion word. While trait associations and predictions over longer periods are not as affected by the lack of context consideration, state affect in small time frames is often highly context dependent and, consequently, requires linguistic context. Therefore, researchers should be aware of the limits of occurrence-counts, like word dictionaries, especially for small time frames and potentially use other methods as discussed in the next section.

3.6 Limitations and Outlook

Besides the general limitations related to the data collection method and the measurement of self-reported affect experience that are discussed in the general discussion of the present dissertation (see section 4.3), there are three specific limitations of this study.

First and foremost, to protect participants' privacy, we did not log the raw words participants had typed into their smartphones and, instead, categorized words based on common word dictionaries directly on their devices. As a consequence, we did not have the raw text data that we could apply advanced NLP methods to, like topic models or word embeddings. These techniques usually show superior predictive performance compared to dictionary approaches since, among other reasons, they are able to consider language's context (Eichstaedt, Kern, et al., 2021; Kjell, Sikström, Kjell, & Schwartz, 2022). Word dictionaries, such as LIWC and sentiment dictionaries, on the contrary, do not recognize language context (e.g., "I am *not* happy") and rely on comparably simple word count occurrences. Furthermore, dictionaries can only detect those words from their internal word lists. As a consequence, even though there is a "netspeak" category in LIWC that is supposed to capture online slang words, dictionaries cannot detect unknown words, such as newly emerging words or words with typos. In our study, on average, 50% of typed words had been recognized by LIWC. This dictionary detection rate is smaller than those reported in prior work using, for example, WhatsApp messages (Koch, Romero, et al., 2022) but similar to prior word based on logged smartphone language (Carlier et al., 2022). To overcome the limitations of dictionary approaches, future studies could use pre-trained language models and deploy them directly on participants' smartphones, which would allow researchers to extract context-aware language characteristics on the device without logging raw text data (Niu et al., 2020; T. Sun et al., 2022).

Second, even though smartphones have become our everyday companions, we only reported on participants' everyday language produced on their smartphones in this study. However, people often use and switch between multiple devices every day (Reeves et al., 2021). As a consequence, we were not able to capture participants' entirety of written language per day. Future work could expand the number of devices observed to include, for example personal computers, laptops, and tablets, using our language logging approach to capture all of participants' produced text.

Third, the findings of this study with regard to different time frames are bound to the specific modeling decision we made. For instance, we chose to analyze weekly, daily, and momentary affect experience. For momentary affect experience, we chose to aggregate all text produced up to 90 minutes before and after the experience sampling instance. Different modeling decision could lead to different findings: For example, if one were to aggregate all text for a given time frame *before* the experience sampling or *after*, or predicting language use from affect self-reports, results might look different (Kross et al., 2019). Therefore, future work could extend on our reported analyses in multiple ways by applying different time windows or running sensitivity analyses to find the optimal time window to capture momentary affect experience. Future studies could also experimentally induce desired affect or detect emotion onset from other sensing modalities (e.g., heart rate) and capture corresponding language using our language logging approach.

3.7 Conclusion

This study employed a novel data collection approach to analyze affective language patterns across communication contexts and time frames in everyday smartphone language. We found distinct in-sample associations of language use, such as I-talk, with affect experience that highlight the importance of distinguishing between time frames (i.e., trait, weekly, daily, momentary) and communication contexts (i.e., private versus public communication). Overall, however, prediction performance of employed machine learning models was not significantly better than chance. These findings emphasize the challenges of using occurrence-counts, such as word dictionaries, to infer subjective affect experience, particularly for small time windows. Finally, this work highlights the opportunities of employing smartphones to collect and analyze language data in everyday life.

3.8 Author Contribution

In addition to myself, Florian Bemmann (F.B.), Markus Buehner (M.B.), Johannes Eichstaedt (J.E.), Ramona Schoedel (R.S.), and Clemens Stachl (C.S.) contributed to this study. M.B. and C.S. acted as supervisors. F.B. created the logging software to collect the data. J.E. assisted with data modeling decisions. R.S. managed data collection.

3.9 Acknowledgements

We thank the Leibniz Institute for Psychology (ZPID) for their support with data collection and Petra Kralj Novak for providing us with the emoji sentiment data.

3.10 Appendix

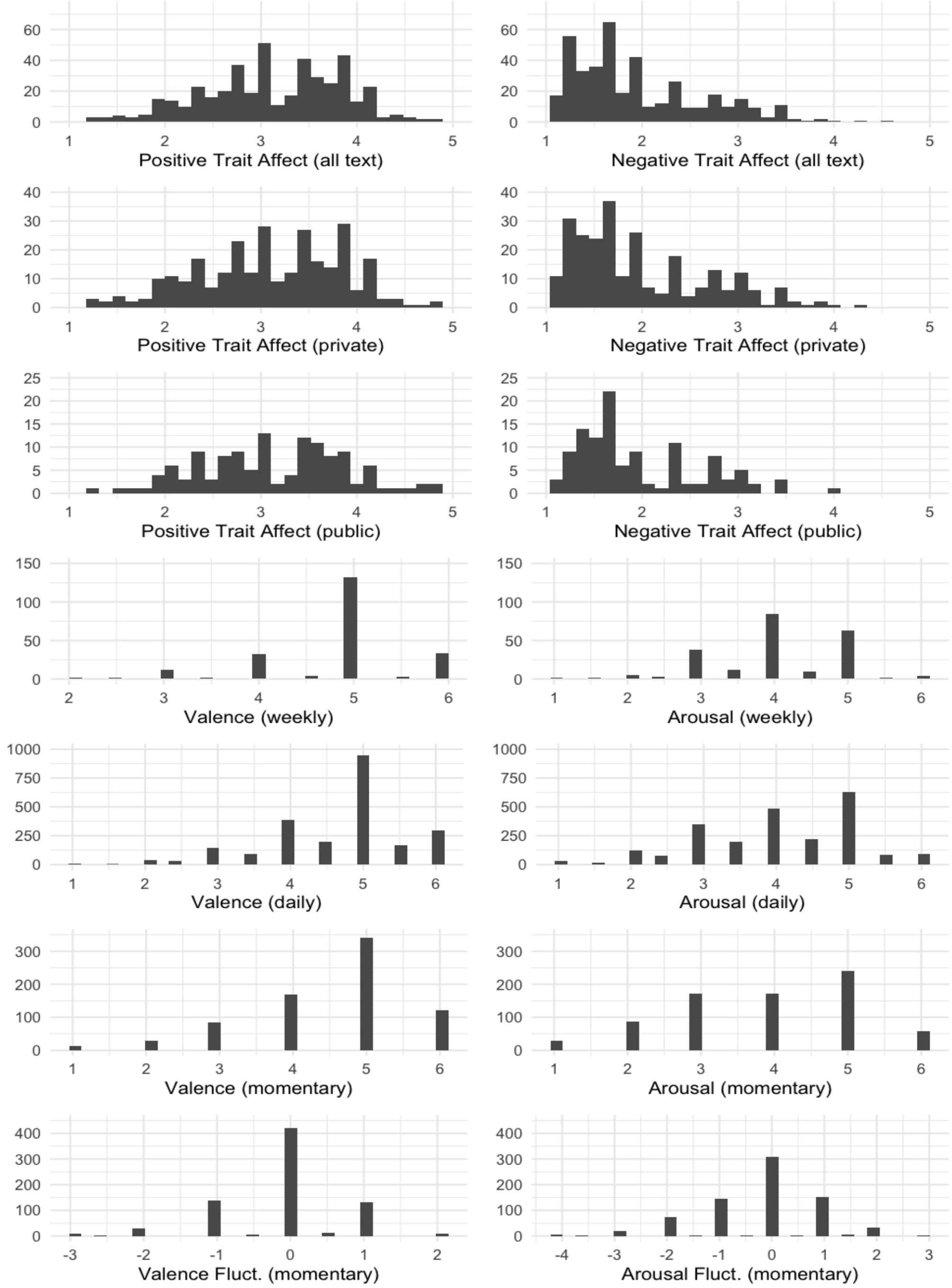


Figure 3.9: Distribution of affect measures across communication contexts and time frames in study 2.

Chapter 4

General Discussion

The present dissertation investigated between-person differences and within-person fluctuations in subjective affect experience in spoken and written language collected in everyday life with novel data collection methods using off-the-shelf smartphones.

Study 1 analyzed two data sets of speech samples from Germany and the United States to evaluate the predictive power of speech (voice cues and spoken content) for corresponding subjective momentary affect experience. While voice acoustics provided limited predictive information of affective arousal, speech content was predictive of affective arousal as well as valence (sadness and contentedness). Overall, predictions were better when participants could talk freely (versus reading out loud predefined emotional sentences). Finally, experimental and explorative findings suggest that emotional speech content had no effect on affect predictions from voice acoustics.

Study 2 investigated how affect experience was associated with and predictable from language characteristics in everyday language collected through the smartphone's keyboard across communication contexts and time frames. Distinctive (in-sample) associations of specific language features based on typing dynamics, word use from (sentiment) dictionaries, and emoji and emoticon use with affect experience were identified. Out-of-sample predictions of affect experience from extracted language cues were, however, not significantly better than chance.

Since specific findings from the two empirical studies have been discussed in the respective chapters, this general discussion focuses on the three overarching patterns in the overall findings across studies and their implications. Further, the overall contribution of the present dissertation is contextualized with regard to its limitations. Finally, future directions and challenges in affect recognition from natural language are discussed.

4.1 Overall Findings

4.1.1 Predicting Subjective Affect Experience is Hard

Overall, the prediction performance for the automated recognition of subjective affect experience reported in the two studies of this dissertation is lower than the prediction accuracies reported in prior work that are predicting affect expression using labelled written language samples (Eichstaedt & Weidman, 2020) or enacted or labelled speech samples (Shen, Changjun, & Chen, 2011). However, the reported prediction performance is comparable to prior studies predicting subjective self-reported affect experience (Carlier et al., 2022; J. Sun et al., 2020; Weidman et al., 2020). In line with prior work, this pattern suggests that the automatic recognition of one's subjective affect experience is more challenging than the prediction of labelled or enacted affective language (Vogt et al., 2008). Moreover, for trained prediction models in both empirical studies, prediction performance coefficients were smaller for within-person fluctuations than for between-person differences suggesting that it is more challenging to algorithmically detect the former than the latter. This insight is in line with prior research on the recognition of between-person differences and within-person fluctuations in affect experience from language (J. Sun et al., 2020; Weidman et al., 2020).

4.1.2 Content Trumps Form

While language form, such as voice cues and typing dynamics, have been shown to contain valuable information about affect experience in the present dissertation, content in spoken and written language yielded higher (in-sample) associations and (out-of-sample) predictions, particularly for affective valence. Specifically, in study 1.2, state-of-the-art word embeddings were more predictive for affect experience than voice acoustics. In study 2, word and emoji use showed higher correlations with affect experience than typing dynamics. This observation of the superiority of content over form in prediction models for subjective self-reported affect experience trained on language samples is in line with prior studies (J. Sun et al., 2020; Weidman et al., 2020).

4.1.3 The Context of Language Production Matters

Both studies of the present dissertation illustrate that the context in which spoken or written language had been produced influences how language characteristics are associated with and predictive for affect experience: Regarding speech, study 1 indicates that it matters for affect predictions from voice cues if the spoken content is predefined or one is able to talk freely. Moreover, results indicate that the emotional valence of the spoken content does not have an effect on the affect recognition from voice cues. Regarding written text, study 2 suggests that the time span over which language is aggregated (trait, weekly, daily, momentary) and the communication context (private versus public communication) influence how language is related to affect experience. This insight is in line with prior research showing that associations of language patterns with psychological phenomena differ across textual channels, such as different social media platforms and instant messaging services (Jaidka et al., 2018; Tingting Liu et al., 2022; Mehl et al., 2013).

4.2 Contributions of the Present Dissertation

4.2.1 Investigation of Natural Language Data Collected in the Wild

While prior works in this research area often relied on language data produced in artificial lab situations (Schuller, 2018; Vogt et al., 2008), the two studies that are part of the present dissertation are among the first to empirically investigate everyday language collected by leveraging novel mobile sensing methods over an extended period of time. Choosing this naturalistic approach yields more ecologically valid findings (Harari et al., 2016). More generally, the present dissertation illustrates that off-the-shelf smartphones enable the scientific collection of spoken and written language in everyday life. Thereby, researchers can collect and analyze traces of naturally occurring language footprints, that are often only accessible to tech corporations that have access to such data through, for example, voice assistants.

4.2.2 Exploration of Differences and Fluctuations in Affect Experience

In contrast to prior research that mainly used affect labels from raters or enacted emotions representing affect *expression* (Preoţiuc-Pietro et al., 2016), the present dissertation leveraged smartphone-based experience sampling to collect in-situ self-reports of affect *experience* in order to associate language with subjective affect and make predictions. Furthermore, due to the longitudinal nature of the collected data (i.e., multiple data points of subjective affect experience with corresponding language per participant), not only between-person differences, but also within-person fluctuations in affect experience have been examined in the present dissertation. Those within-person fluctuations in a person's emotional life over time can be very relevant for one's well-being (Eichstaedt & Weidman, 2020).

4.2.3 Analysis of Context Effects on Affect Recognition from Language

Prior research has shown that the context in which spoken or written language had been produced, such as in a public or a private communication context (Mehl et al., 2013), has an effect on prediction models that have been trained on these data (Jaidka et al., 2018; Tingting Liu et al., 2022). The present dissertation contributes to this research field by leveraging context information in the respective analyses. Specifically, Study 1 investigated the impact of fixed content versus being able to talk freely as well as the emotional valence of the content on affect predictions from voice cues. In study 2, context information on time aggregation and communication context (private versus public channels) has been used to analyze their effects on language footprints of affect experience in written language. These insights emphasize that language models for affect recognition need to take the context into account (i.e., context specific models) and have to be trained and tested accordingly.

4.2.4 Insights through Description, Prediction, and Explanation

While prior studies on affect inferences from language either focused solely on specific descriptive associations (Tackman et al., 2018) or advanced predictions (Shen et al., 2011), the present dissertation combines description, prediction, and explanation: Both empirical studies reported on descriptive in-sample associations of language features and affect experience. Further, affect experience was predicted using supervised machine learning algorithms in the two empirical studies. Finally, fitted predictive models were investigated using interpretable machine learning methods in study 1. By combining these approaches, the present dissertation aims at advancing affect theory by creating new insights into the language footprints of affect experience (Elhai & Montag, 2020; Harari et al., 2020; Mahmoodi, Leckelt, van Zalk, Geukes, & Back, 2017; Shrestha et al., 2021).

4.3 General Limitations

The general limitations of the overall findings of the two studies of this dissertation are twofold: First, the peculiarities of the data collection method and study samples that have implications for the generalizability of findings and, second, the affect self-reports that served as ground truth themselves. Those limitations that are specific to the single studies' results are discussed in the respective chapters (see section 2.6 and 3.6).

4.3.1 Data Collection

The analyzed data in the present dissertation is comprised of in-situ self-reports of affect experience from participants' everyday life in a non-clinical population. As a result, the data represents the "normal" everyday *moods* of regular people with only few cases of extreme positive or negative or very high or low aroused affect experience. As a result, the discovered language-affect associations and trained affect recognition models should be considered in this context. If the conducted analyses were to be replicated in a clinical sample that contains more cases of extreme affect experience, findings might be even more distinct. This does not necessarily mean, however, that findings and prediction models from the present dissertation generalize well to a clinical sample. Alternatively, future studies could aim to collect language samples when participants are known to experience strong emotions, for example based on their physiological signals (Hoemann et al., 2020).

Also, it should be noted that affective self-reports only from those situations when participants had their smartphone on them and felt comfortable to complete the experience sampling had been analyzed in the present dissertation. Moreover, study participants knew that their affect self-reports and corresponding language samples would be recorded and later analyzed. As a consequence, with regard to assessed subjective affect experience, participants might have only completed the experience sampling questionnaire in selected affective situations, for example not when they were experiencing extreme affective states, or they had not reported on extreme affect at all (Schoedel et al., 2022).

Also, participants might have not spoken or written as naturally as they would if they had not known that their data would be scientifically analyzed. Participants might have made the audio records for study 1 only in selected suitable situations, for example when they were alone in a quiet place. As a result, the resulting data set would also only contain participants' affect experience from being alone in quiet places. Further, regarding the keyboard language analyzed in study 2, participants might have altered their language use knowing that their data would be scientifically analyzed.

All data sets analyzed in the present dissertation have been collected in Western countries, specifically Germany and the United States. Prior research suggests that there are cultural differences in emotion experience (Lim, 2016) and that mood inference models from sensing data do not necessarily generalize to other countries and cultures (Meegahapola et al., 2023). Therefore, future research should make use of diverse samples from other cultural contexts and non-Western countries.

Furthermore, the German data set that is analyzed in both empirical studies only contained Android users, excluding those using iOS devices, because of the technical requirements of the logging software. However, past research suggests that the selection bias regarding participant demographics and personality for the German population is negligible (Götz, Stieger, & Reips, 2017; Keusch, Bähr, Haas, Kreuter, & Trappmann, 2020; Schoedel et al., 2022).

Finally, data collection of the German sample was conducted during the COVID-19 pandemic in 2020. The impact of the pandemic itself and the with corresponding legislative measures have been shown to have had an influence on people's day-to-day affect experience (Bäuerle et al., 2020; Rajkumar, 2020), their language use (Pisano et al., 2022; Romero, Mikiya, Nakatsuma, Fitz, & Koch, 2021), and their smartphone use in general (Jonnatan, Seaton, Rush, Li, & Hasan, 2022; Katsumata, Ichikohji, Nakano, Yamaguchi, & Ikuine, 2022). Therefore, future research may aim to replicate the findings of the present dissertation using samples that have been collected during "normal" times.

4.3.2 Affect Self-Reports as Ground Truth

The ground truth, i.e., the information that is assumed to be fundamentally true, used for the descriptive analyses as well as model training and evaluation in the present dissertation are self-reports on participants' subjective affect experience. However, these are prone to response biases (Gao, Rahaman, Shao, & Salim, 2021). These can introduce measurement error that can have a profound impact on the consecutive predictive modeling (Jacobucci & Grimm, 2020). Moreover, single items were employed to assess state affect experience on different dimensions (i.e., valence and arousal) in the empirical studies. This is an established approach to reduce participant burden by not having them fill out many lengthy experience sampling questionnaires that would also need to be compensated for. However, this single-item approach can introduce additional measurement error for affective responses (Dejonckheere et al., 2022). Future studies that are particularly interested in subjective affect experience should use multiple items assessing a broad range of affect experience in an intensive longitudinal design.

Beyond the psychometric challenges associated with using (single item) self-reports to assess momentary affect experience, there is a conceptual debate on how much of an underlying psychological construct, i.e., of affect experience, one can assess using questionnaires. In order to report one's affect experience most accurately through a survey item, one needs to have introspection to recognize it and have an adequate understanding to communicate it accordingly (Boyd, Pasca, & Lanning, 2020; Montag, Dagum, Hall, & Elhai, 2022). However, people can vary greatly in that regard (Critchley & Garfinkel, 2017). Possibly, in the future, algorithms can replace self-report questionnaires altogether by analyzing natural language directly since transformers (that use word embeddings as employed in study 1.2) have been reported to approach the upper limits in accuracy already (Kjell et al., 2022).

4.4 Future Directions and Challenges

Recognizing affect experience from naturally occurring language collected with smartphone holds many future promises for research and commercial applications in the form of multi-modal affect recognition, time-series, and idiographic models. However, to realize their potential the challenges of a precise conceptualization of affect, interdisciplinarity, ethics, and data privacy and data security need to be overcome.

4.4.1 Multi-Modal Affect Recognition

Emotional expressions are multimodal, dynamic patterns of behavior and language represents only one channel to express affect experience (Keltner, Sauter, Tracy, & Cowen, 2019). Therefore, affect recognition from smartphone data could be improved if multiple sensed or logged data streams beyond language use are merged. Future research could, for instance, combine language data with recorded app usage, movement patterns from GPS signals, and physical activity, possibly enriched with data from wearables' physiological sensors (Cai et al., 2020; Tzirakis, Chen, Zafeiriou, & Schuller, 2021; Yang et al., 2021).

4.4.2 Time-Series Models

Mobile sensing (language) data in combination with corresponding experience sampling assessments provides a granular continuous data stream over a given period of time. This longitudinal nature of the collected data allows to apply so-called "time-series models" that leverage temporal information from prior language and affect experience to forecast future affect experience (Busk et al., 2020; Suhara, Xu, & Pentland, 2017). Most work in this research field, like the present dissertation, has not yet employed those time-series models because they require an intensive longitudinal study design with many data points per single participant. Future work could, for example, use time-series modelling based on rich multi-modal sensing data with corresponding affect assessments to forecast mood and potentially symptoms of affective disorders.

4.4.3 Idiographic Models

Employed machine learning models in the present dissertation had been trained and tested on all participants using cross-validation. However, affect experience is highly subjective and individual depending on multiple factors, such as individual trait differences and previous life experiences. Further, how one expresses their affect experience through language can be also highly individual. To account for the subjective and individual nature of affect experience and expression, particularly for emotional valence, so-called idiographic (person-specific) models that are trained and evaluated on data from only one single person can be applied. As a results, they provide a person-specific model and corresponding predictions (Beck & Jackson, 2022; Piccirillo & Rodebaugh, 2019). For example, they would be able to capture if an individual uses specific emoji when in a bad mood. This modelling approach could be particularly promising to detect within-person fluctuations in affect experience since those had been shown to be hard to predict if prediction models had been trained and tested on all participants in the present dissertation. Additionally, if sufficient intensive longitudinal data is available, time-series models could also be used for single individuals. In a next step, researchers could then look for similarities across idiographic models to detect more general patterns related to affect experience (Beck & Jackson, 2022). Such idiographic models have already been shown to hold promising applications in psychological research and in applied settings, such as psychotherapy (Piccirillo & Rodebaugh, 2019).

4.4.4 A Precise Conceptualization of Affect

Prior studies on affect recognition from language vary greatly in how affect had been conceptualized and assessed: Some had been focused on discrete induced emotions (e.g., happiness), some on longer lasting moods based on the concept of core affect. As a consequence, a comparison of findings across studies is often challenging. Even though there is no universal agreement on the underlying conceptualization of affect (see chapter 1), researchers should use a clear and precise conceptualization and corresponding terminology of the kind of affect they are working on: For example, one should distinguish if one works on (intense, short termed) enacted emotions or everyday (low intensity, longer lasting) moods. This eases the comparison and aggregation of findings across studies and moves the research field forward.

4.4.5 Interdisciplinarity

To realize the potential of smartphone-based affect recognition from language the disciplines involved in this kind of research - from psychology, computer science, and data science to phonetics and linguistics - need to work together in interdisciplinary collaborations. Such projects require deep domain knowledge of affect and spoken and written language, programming skills to develop modalities to collect such language data using smartphones, data management and data processing skills, and statistical skills for algorithmic modeling (Miller, 2012; Seifert, Hofer, & Allemand, 2018). Also, an adequate technical infrastructure to collect, store, and analyze the data is required. These requirements can only be met in well-equipped interdisciplinary research groups. Therefore, researchers should receive training in the neighboring disciplines and incentives for collaboration though funding organizations, publication modalities, and academic career programs should be actively installed (Lazer et al., 2009, 2020; Seifert et al., 2018). The present dissertation has been produced in such a fruitful interdisciplinary collaboration that has made this kind of research even possible.

4.4.6 Ethics

With emotion recognition from language increasingly being used in research and practice, important ethical considerations need to be discussed. Gaining insights into one's personal emotional life can yield extremely sensitive data and may result in potential harms to individuals and society (Hernandez et al., 2021). Such emotion-detecting artificial intelligence (AI) tools can be used for good (e.g., improving one's affective well-being) or bad (e.g., exploiting moments of emotional vulnerability). Particularly, if idiographic models are realized, those may be even more sensitive as they would know the particularities of a single person. As a consequence, guidelines and mitigation mechanisms to detect and minimize those ethical risks must be installed (Andalibi & Buss, 2020; Hernandez et al., 2021; Mohammad, 2022; Stark & Hoey, 2021). For example, only those data should be collected that is necessary for research or a specific application. Also, the knowledge about individuals that arises from affect inferences from language should be handled and used responsibly. More generally, given that tech corporations, such as like Google, Facebook, or Amazon, either already use emotion recognition or have filed corresponding patents (Knight, 2016), research works, such as the present dissertation, should inform the public and policy makers about the possibilities that arise from affect inferences from digital language footprints (Andalibi & Buss, 2020).

4.4.7 Data Privacy and Data Security

Beyond the aforementioned ethical considerations, ensuring data privacy and data security for research study participants and users of digital products leveraging automatic affect recognition from natural language is essential. Data privacy in this context means that one knows who collects, has access to, and can analyze these highly sensitive data. In this manner, study participants and users of digital products should be informed extensively about what data is collected for which purpose in order to obtain informed consent (Harari, 2020). Such an informed consent process had been followed in the data collections for the present dissertation. In today's world, however, with the rise of voice assistants and chat bots, digital language footprints are ubiquitous, and users are often not aware of what happens with their data. Also, policy making with regard to data privacy is only slowly catching up with technological progress in this field. Therefore, this dissertation also aims to raise awareness for how much affective information people might give away unknowingly through their language footprints. Users should be aware of what can be inferred from the data they give away to make an informed decision regarding their privacy. Once data had been collected, data security is concerned with the collected data being securely stored through technical and procedural measures (Harari, 2020). Data collection for the present dissertation included strict protocols including the secure storage protected against third party access and the deletion of sensitive data to ensure the utmost security of participants' data. Researchers and tech corporations could also use technical modalities to extract meaningful language characteristics (e.g., voice cues) directly on the device. Thereby, no raw data would have to be stored on a server. In the two studies of this dissertation, privacy-respectful smartphone logging approaches by extracting voice cues and text characteristics directly on the device had been employed. Moreover, researchers could leverage new technologies, such as blockchain technology, to store data securely (Jin, Zhang, Zhou, & Yu, 2019; Poonguzhali, Gayathri, Deebika, & Suriapriya, 2020). Ideally, affect predictions could also happen on the device employing pre-trained machine learning models (Dhar et al., 2021). Thereby, no data would have to be transferred onto external servers at all.

4.5 Conclusion

The present dissertation showcases the utility of smartphones to investigate subjective affect experience in natural language in everyday life. Hereby, caveats of prior research methods with regard to the conceptual difference of affect expression versus affect experience as well as collecting timely paired language and affect data can be overcome. By leveraging app-based experience sampling and on-device language data collection in everyday life, this work shows how characteristics of spoken and written language are associated with and predictive for subjective affect experience. Results suggest that recognizing subjective affect *experience* from spoken and written language data is more difficult than inferring affect *expression* as done in prior research. Moreover, the context of language production plays a major role for affect predictions. Using statistical methods in the areas of description, prediction, and explanation, this dissertation also analyzes specific affect-linked language characteristics. The promising applications and potential future directions of this technology come with multiple challenges with regard to the precise conceptualization of affect, interdisciplinarity of research groups, ethics, and data privacy and security. If these challenges can be overcome, natural language analysis based on data collected with smartphones represents a promising tool to monitor affective well-being and to advance the affective sciences.

References

Allport, G. W. (1942). The use of personal documents in psychological science. *Social Science Research Council Bulletin*, *49*.

Andalibi, N., & Buss, J. (2020). The Human in Emotion Recognition on Social Media: Attitudes, Outcomes, Risks. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). Honolulu HI USA: ACM. http://doi.org/10.1145/3313831.3376680

Back, M. D., Küfner, A. C. P., & Egloff, B. (2010). The Emotional Timeline of September 11, 2001. *Psychological Science*, *21*(10), 1417–1419. http://doi.org/10.1177/0956797610382124

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614–636. http://doi.org/10.1037/0022-3514.70.3.614

Bänziger, T., Hosoya, G., & Scherer, K. R. (2015). Path Models of Vocal Emotion Communication. *PLOS ONE*, *10*(9), e0136675. http://doi.org/10.1371/journal.pone.0136675

Bänziger, T., Mortillaro, M., & Scherer, K. R. (2012). Introducing the Geneva Multimodal expression corpus for experimental research on emotion perception. *Emotion (Washington, D.C.)*, *12*(5), 1161–1179. http://doi.org/10.1037/a0025827

Barrett, L. F. (2022, April). Facial Expressions Do Not Reveal Emotions. *Scientific American*. https://www.scientificamerican.com/article/darwin-was-wrong-your-facial-expressions-do-not-reveal-your-emotions/.

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. *Psychological Science in the Public Interest*, *20*(1), 1–68. http://doi.org/10.1177/1529100619832930

Barrett, L. F., & Bliss-Moreau, E. (2009). Chapter 4 Affect as a Psychological Primitive. In *Advances in Experimental Social Psychology* (Vol. 41, pp. 167–218). Academic Press. http://doi.org/10.1016/S0065-2601(08)00404-8

Batliner, A., Hacker, C., Steidl, S., Nöth, E., D'Arcy, S., Russell, M., & Wong, M. (2004). "You Stupid Tin Box" - Children Interacting with the AIBO Robot: A Cross-linguistic Emotional Speech Corpus. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*. Lisbon, Portugal: European Language Resources Association (ELRA).

Batliner, A., Schuller, B., Seppi, D., Steidl, S., Devillers, L., Vidrascu, L., … Amir, N. (2011). The Automatic Recognition of Emotions in Speech. In *Cognitive Technologies* (pp. 71–99). http://doi.org/10.1007/978-3-642-15184-2_6

Bäuerle, A., Teufel, M., Musche, V., Weismüller, B., Kohler, H., Hetkamp, M., … Skoda, E.-M. (2020). Increased generalized anxiety, depression and distress during the COVID-19 pandemic: A cross-sectional study in Germany. *Journal of Public Health*, *42*(4), 672–678. http://doi.org/10.1093/pubmed/fdaa106

Bazarova, N. N., & Choi, Y. H. (2014). Self-Disclosure in Social Media: Extending the Functional Approach to Disclosure Motivations and Characteristics on Social Network Sites. *Journal of Communication*, *64*(4), 635–657. http://doi.org/10.1111/jcom.12106

Bazarova, N. N., Taft, J. G., Choi, Y. H., & Cosley, D. (2013). Managing Impressions and Relationships on Facebook: Self-Presentational and Relational Concerns Revealed Through the Analysis of Language Style. *Journal of Language and Social Psychology*, *32*(2), 121–141. http://doi.org/10.1177/0261927X12456384

Beck, E. D., & Jackson, J. J. (2022). Personalized Prediction of Behaviors and Experiences: An Idiographic Person–Situation Test. *Psychological Science*, *33*(10), 1767–1782. http://doi.org/10.1177/09567976221093307

Bemmann, F., & Buschek, D. (2020). LanguageLogger: A Mobile Keyboard Application for Studying Language Use in Everyday Text Communication in the Wild. *Proceedings of the ACM on Human-Computer Interaction*, *4*(EICS), 1–24. http://doi.org/10.1145/3397872

Ben-David, B. M., Multani, N., Shakuf, V., Rudzicz, F., & van Lieshout, P. H. H. M. (2016). Prosody and Semantics Are Separate but Not Separable Channels in the Perception of Emotional Speech: Test for Rating of Emotions in Speech. *Journal of Speech, Language, and Hearing Research*, *59*(1), 72–89. http://doi.org/10.1044/2015_JSLHR-H-14-0323

Bennett, C. C., Ross, M. K., Baek, E., Kim, D., & Leow, A. D. (2022a). Predicting clinically relevant changes in bipolar disorder outside the clinic walls based on pervasive technology interactions via smartphone typing dynamics. *Pervasive and Mobile Computing*, *83*, 101598. http://doi.org/10.1016/j.pmcj.2022.101598

Bennett, C. C., Ross, M. K., Baek, E., Kim, D., & Leow, A. D. (2022b). Smartphone accelerometer data as a proxy for clinical data in modeling of bipolar disorder symptom trajectory. *Npj Digital Medicine*, *5*(1), 1–10. http://doi.org/10.1038/s41746-022-00741-3

Biecek, P. (2018). DALEX: Explainers for Complex Predictive Models in R. *Journal of Machine Learning Research*, *19*(84), 1–5.

Bischl, B., Mersmann, O., Trautmann, H., & Weihs, C. (2012). Resampling Methods for Meta-Model Validation with Recommendations for Evolutionary Computation. *Evolutionary Computation*, *20*(2), 249–275. http://doi.org/10.1162/EVCO_a_00069

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research* (pp. xv, 256). New York, NY, US: Guilford Press.

Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). The development and psychometric properties of LIWC-22. *Austin, TX: University of Texas at Austin*.

Boyd, R. L., Pasca, P., & Lanning, K. (2020). The Personality Panorama: Conceptualizing Personality through Big Behavioural Data. *European Journal of Personality*, *34*(5), 599–612. http://doi.org/10.1002/per.2254

Boyd, R. L., & Schwartz, H. (2021). Natural Language Analysis and the Psychology of Verbal Behavior: The Past, Present, and Future States of the Field. *Journal of Language and Social Psychology*, *40*. http://doi.org/10.1177/0261927X20967028

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Breyer, B., & Bluemke, M. (2016). Deutsche Version der Positive and Negative Affect Schedule PANAS (GESIS Panel). *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*. http://doi.org/10.6102/ZIS242

Brinberg, M., Ram, N., Yang, X., Cho, M.-J., Sundar, S. S., Robinson, T. N., & Reeves, B. (2021). The idiosyncrasies of everyday digital lives: Using the Human Screenome Project to study user behavior on smartphones. *Computers in Human Behavior*, *114*, 106570. http://doi.org/10.1016/j.chb.2020.106570

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Univ of California Press.

Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W. F., & Weiss, B. (2005). A database of German emotional speech. In *Interspeech* (Vol. 5, pp. 1517–1520).

Buschek, D., Bisinger, B., & Alt, F. (2018). ResearchIME: A Mobile Keyboard Application for Studying Free Typing Behaviour in the Wild. In *Proceedings of the*

2018 CHI Conference on Human Factors in Computing Systems - CHI '18 (pp. 1–14). Montreal QC, Canada: ACM Press. http://doi.org/10.1145/3173574.3173829

Busk, J., Faurholt-Jepsen, M., Frost, M., Bardram, J. E., Kessing, L. V., & Winther, O. (2020). Forecasting Mood in Bipolar Disorder From Smartphone Self-assessments: Hierarchical Bayesian Approach. *JMIR mHealth and uHealth*, *8*(4), e15028. http://doi.org/10.2196/15028

Cai, L., Boukhechba, M., Wu, C., Chow, P. I., Teachman, B. A., Barnes, L. E., & Gerber, M. S. (2020). State affect recognition using smartphone sensing data. In *Proceedings of the 2018 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies* (pp. 120–125). New York, NY, USA: Association for Computing Machinery. http://doi.org/10.1145/3278576.3284386

Cao, B., Zheng, L., Zhang, C., Yu, P. S., Piscitello, A., Zulueta, J., ... Leow, A. D. (2017). DeepMood: Modeling Mobile Phone Typing Dynamics for Mood Detection. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 747–755). Halifax NS Canada: ACM. http://doi.org/10.1145/3097983.3098086

Carlier, C., Niemeijer, K., Mestdagh, M., Bauwens, M., Vanbrabant, P., Geurts, L., ... Kuppens, P. (2022). In Search of State and Trait Emotion Markers in Mobile-Sensed Language: Field Study. *JMIR Mental Health*, *9*(2), e31724. http://doi.org/10.2196/31724

Chomsky, N. (1975). *Reflections on language.* Pantheon Books.

Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic Markers of Psychological Change Surrounding September 11, 2001. *Psychological Science*, *15*(10), 687–693. http://doi.org/10.1111/j.0956-7976.2004.00741.x

Conner, T. S., Tennen, H., Fleeson, W., & Barrett, L. F. (2009). Experience Sampling Methods: A Modern Idiographic Approach to Personality Research. *Social and Personality Psychology Compass*, *3*(3), 292–313. http://doi.org/10.1111/j.1751-9004.2009.00170.x

Critchley, H. D., & Garfinkel, S. N. (2017). Interoception and emotion. *Current Opinion in Psychology*, *17*, 7–14. http://doi.org/10.1016/j.copsyc.2017.04.020

Csikszentmihalyi, M., & Larson, R. (2014). Validity and Reliability of the Experience-Sampling Method. In M. Csikszentmihalyi (Ed.), *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi* (pp. 35–54). Dordrecht: Springer Netherlands. http://doi.org/10.1007/978-94-017-9088-8_3

Darwin, C. (1886). *The Expression of the Emotions in Man and Animals.* D. Appleton.

De Choudhury, M., Counts, S., & Horvitz, E. (2013). Social media as a measurement tool of depression in populations. In *Proceedings of the 5th Annual ACM Web Science Conference on - WebSci '13* (pp. 47–56). Paris, France: ACM Press. http://doi.org/10.1145/2464464.2464480

Defren, S., de Brito Castilho Wesseling, P., Allen, S., Shakuf, V., Ben-David, B., & Lachmann, T. (2018). Emotional Speech Perception: A set of semantically validated German neutral and emotionally affective sentences. In *9th International Conference on Speech Prosody 2018* (pp. 714–718). ISCA. http://doi.org/10.21437/SpeechProsody.2018-145

Dejonckheere, E., Demeyer, F., Geusens, B., Piot, M., Tuerlinckx, F., Verdonck, S., & Mestdagh, M. (2022). Assessing the reliability of single-item momentary affective measurements in experience sampling. *Psychological Assessment*, *34*(12), 1138–1154. http://doi.org/10.1037/pas0001178

Demetriou, C., Ozer, B. U., & Essau, C. A. (2015). Self-Report Questionnaires. In R. L. Cautin & S. O. Lilienfeld (Eds.), *The Encyclopedia of Clinical Psychology* (pp. 1–6). Hoboken, NJ, USA: John Wiley & Sons, Inc. http://doi.org/10.1002/9781118625392.wbecp507

Dhar, S., Guo, J., Liu, J. (Jason)., Tripathi, S., Kurup, U., & Shah, M. (2021). A Survey of On-Device Machine Learning: An Algorithms and Learning Theory Perspective. *ACM Transactions on Internet of Things*, *2*(3), 1–49. http://doi.org/10.1145/3450494

Dukes, D., Abrams, K., Adolphs, R., Ahmed, M. E., Beatty, A., Berridge, K. C., … Sander, D. (2021). The rise of affectivism. *Nature Human Behaviour*, 1–5. http://doi.org/10.1038/s41562-021-01130-8

Eichstaedt, J. C., Kern, M. L., Yaden, D. B., Schwartz, H. A., Giorgi, S., Park, G., … Ungar, L. H. (2021). Closed- and open-vocabulary approaches to text analysis: A review, quantitative comparison, and recommendations. *Psychological Methods*, *26*(4), 398–427. http://doi.org/10.1037/met0000349

Eichstaedt, J. C., Sherman, G. T., Giorgi, S., Roberts, S. O., Reynolds, M. E., Ungar, L. H., & Guntuku, S. C. (2021). The emotional and mental health impact of the murder of George Floyd on the US population. *Proceedings of the National Academy of Sciences*, *118*(39), e2109139118. http://doi.org/10.1073/pnas.2109139118

Eichstaedt, J. C., Smith, R. J., Merchant, R. M., Ungar, L. H., Crutchley, P., Preoţiuc-Pietro, D., … Schwartz, H. A. (2018). Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences*, *115*(44), 11203–11208. http://doi.org/10.1073/pnas.1802331115

Eichstaedt, J. C., & Weidman, A. C. (2020). Tracking Fluctuations in Psychological States Using Social Media Language: A Case Study of Weekly Emotion. *European Journal of Personality*, *34*(5), 845–858. http://doi.org/10.1002/per.2261

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3-4), 169–200. http://doi.org/10.1080/02699939208411068

Elhai, J. D., & Montag, C. (2020). The compatibility of theoretical frameworks with machine learning analyses in psychological research. *Current Opinion in Psychology*, *36*, 83–88. http://doi.org/10.1016/j.copsyc.2020.05.002

Eyben, F., Scherer, K. R., Schuller, B. W., Sundberg, J., Andre, E., Busso, C., ... Truong, K. P. (2016). The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing. *IEEE Transactions on Affective Computing*, *7*(2), 190–202. http://doi.org/10.1109/TAFFC.2015.2457417

Eyben, F., Weninger, F., Gross, F., & Schuller, B. (2013). Recent Developments in openSMILE, the Munich Open-source Multimedia Feature Extractor. In *Proceedings of the 21st ACM International Conference on Multimedia* (pp. 835–838). New York, NY, USA: ACM. http://doi.org/10.1145/2502081.2502224

Eyben, F., Wöllmer, M., & Schuller, B. (2010). Opensmile: The munich versatile and fast open-source audio feature extractor. In *Proceedings of the international conference on Multimedia - MM '10* (p. 1459). Firenze, Italy: ACM Press. http://doi.org/10.1145/1873951.1874246

Fairbanks, G., & Pronovost, W. (1939). An experimental study of the pitch characteristics of the voice during the expression of emotions. *Speech Monographs*, *6*, 87–104. http://doi.org/10.1080/03637753909374863

Faurholt-Jepsen, M., Busk, J., Frost, M., Vinberg, M., Christensen, E. M., Winther, O., ... Kessing, L. V. (2016). Voice analysis as an objective state marker in bipolar disorder. *Translational Psychiatry*, *6*(7), e856–e856. http://doi.org/10.1038/tp.2016.123

Ferreira, D., Kostakos, V., & Dey, A. K. (2015). AWARE: Mobile Context Instrumentation Framework. *Frontiers in ICT*, *2*, 6. http://doi.org/10.3389/fict.2015.00006

Freud, S., & Strachey, J. (1901). The psychopathology of everyday life. The Standard Edition of the complete psychological works of Sigmund Freud. *Trans. James Strachey*, *24*, 1953–74.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1.

Gao, N., Rahaman, M. S., Shao, W., & Salim, F. D. (2021, November). Investigating the Reliability of Self-report Data in the Wild: The Quest for Ground Truth.

arXiv. Retrieved from `https://arxiv.org/abs/arXiv:2107.00389`

Ghosh, S., Ganguly, N., Mitra, B., & De, P. (2017). Evaluating effectiveness of smartphone typing as an indicator of user emotion. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 146–151). http://doi.org/10.1109/ACII.2017.8273592

Ghosh, S., Hiware, K., Ganguly, N., Mitra, B., & De, P. (2019). Emotion detection from touch interactions during text entry on smartphones. *International Journal of Human-Computer Studies*, *130*, 47–57. http://doi.org/10.1016/j.ijhcs.2019.04.005

Götz, F. M., Stieger, S., & Reips, U.-D. (2017). Users of the main smartphone operating systems (iOS, Android) differ only little in personality. *PLOS ONE*, *12*(5), e0176921. http://doi.org/10.1371/journal.pone.0176921

Grimm, M., Kroschel, K., & Narayanan, S. (2008). The Vera am Mittag German audio-visual emotional speech database. In *2008 IEEE International Conference on Multimedia and Expo* (pp. 865–868). http://doi.org/10.1109/ICME.2008.4607572

Gross, J. J., & John, O. P. (1997). Revealing feelings: Facets of emotional expressivity in self-reports, peer ratings, and expressive behavior. *Journal of Personality and Social Psychology*, 434–447.

Gross, J. J., John, O. P., & Richards, J. M. (2000). The Dissociation of Emotion Expression from Emotion Experience: A Personality Perspective. *Personality and Social Psychology Bulletin*, *26*(6), 712–726. http://doi.org/10.1177/0146167200268006

Harari, G. M. (2020). A process-oriented approach to respecting privacy in the context of mobile phone tracking. *Current Opinion in Psychology*, *31*, 141–147. http://doi.org/10.1016/j.copsyc.2019.09.007

Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., & Gosling, S. D. (2016). Using Smartphones to Collect Behavioral Data in Psychological Science: Opportunities, Practical Considerations, and Challenges. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *11*(6), 838–854. http://doi.org/10.1177/1745691616650285

Harari, G. M., Vaid, S. S., Müller, S. R., Stachl, C., Marrero, Z., Schoedel, R., ... Gosling, S. D. (2020). Personality Sensing for Theory Development and Assessment in the Digital Age. *European Journal of Personality*, *34*(5), 649–669. http://doi.org/10.1002/per.2273

Hernandez, J., Lovejoy, J., McDuff, D., Suh, J., O'Brien, T., Sethumadhavan, A., ... Czerwinski, M. (2021). Guidelines for Assessing and Minimizing Risks of Emotion Recognition Applications. In *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 1–8).

http://doi.org/10.1109/ACII52823.2021.9597452

Hildebrand, C., Efthymiou, F., Busquet, F., Hampton, W. H., Hoffman, D. L., & Novak, T. P. (2020). Voice analytics in business research: Conceptual foundations, acoustic feature extraction, and applications. *Journal of Business Research*, *121*, 364–374. http://doi.org/10.1016/j.jbusres.2020.09.020

Hill, K. (2022). Microsoft Plans to Eliminate Face Analysis Tools in Push for "Responsible A.I." *The New York Times*.

Hoemann, K., Khan, Z., Feldman, M. J., Nielson, C., Devlin, M., Dy, J., … Quigley, K. S. (2020). Context-aware experience sampling reveals the scale of variation in affective experience. *Scientific Reports*, *10*, 12459. http://doi.org/10.1038/s41598-020-69180-y

Huang, Z., & Epps, J. (2018). Prediction of Emotion Change From Speech. *Frontiers in ICT*, *5*, 11. http://doi.org/10.3389/fict.2018.00011

Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), 216–225. http://doi.org/10.1609/icwsm.v8i1.14550

Iliev, R., Dehghani, M., & Sagi, E. (2015). Automated text analysis in psychology: Methods, applications, and future developments. *Language and Cognition*, *7*(02), 265–290. http://doi.org/10.1017/langcog.2014.30

Jackson, J. C., Watts, J., List, J.-M., Puryear, C., Drabble, R., & Lindquist, K. A. (2021). From Text to Thought: How Analyzing Language Can Advance Psychological Science. *Perspectives on Psychological Science*, 17456916211004899. http://doi.org/10.1177/17456916211004899

Jacobucci, R., & Grimm, K. J. (2020). Machine Learning and Psychological Research: The Unexplored Effect of Measurement. *Perspectives on Psychological Science*, *15*(3), 809–816. http://doi.org/10.1177/1745691620902467

Jaidka, K., Giorgi, S., Schwartz, H. A., Kern, M. L., Ungar, L. H., & Eichstaedt, J. C. (2020). Estimating geographic subjective well-being from Twitter: A comparison of dictionary and data-driven language methods. *Proceedings of the National Academy of Sciences*, 201906364. http://doi.org/10.1073/pnas.1906364117

Jaidka, K., Guntuku, S. C., & Ungar, L. H. (2018). Facebook versus Twitter: Cross-platform Differences in Self-Disclosure and Trait Prediction. In *Twelfth International AAAI Conference on Web and Social Media* (p. 10).

Jin, X.-L., Zhang, M., Zhou, Z., & Yu, X. (2019). Application of a Blockchain Platform to Manage and Secure Personal Genomic Data: A Case Study of

LifeCODE.ai in China. *Journal of Medical Internet Research*, *21*(9), e13587. http://doi.org/10.2196/13587

Jonnatan, L., Seaton, C. L., Rush, K. L., Li, E. P. H., & Hasan, K. (2022). Mobile Device Usage before and during the COVID-19 Pandemic among Rural and Urban Adults. *International Journal of Environmental Research and Public Health*, *19*(14), 8231. http://doi.org/10.3390/ijerph19148231

Kassam, K. S., & Mendes, W. B. (2013). The effects of measuring emotion: Physiological reactions to emotional situations depend on whether someone is asking. *PloS One*, *8*(7), e64959. http://doi.org/10.1371/journal.pone.0064959

Kathawalla, U.-K., Silverstein, P., & Syed, M. (2021). Easing Into Open Science: A Guide for Graduate Students and Their Advisors. *Collabra: Psychology*, *7*(1), 18684. http://doi.org/10.1525/collabra.18684

Katsumata, S., Ichikohji, T., Nakano, S., Yamaguchi, S., & Ikuine, F. (2022). Changes in the use of mobile devices during the crisis: Immediate response to the COVID-19 pandemic. *Computers in Human Behavior Reports*, *5*, 100168. http://doi.org/10.1016/j.chbr.2022.100168

Kelley, S., Mhaonaigh, C., Burke, L., Whelan, R., & Gillan, C. (2022). Machine learning of language use on Twitter reveals weak and non-specific predictions. *Npj Digital Medicine*, *5*, 35. http://doi.org/10.1038/s41746-022-00576-y

Keltner, D., Sauter, D., Tracy, J., & Cowen, A. (2019). Emotional Expression: Advances in Basic Emotion Theory. *Journal of Nonverbal Behavior*, *43*(2), 133–160. http://doi.org/10.1007/s10919-019-00293-3

Kennedy, B., Ashokkumar, A., Boyd, R. L., & Dehghani, M. (2021, February). Text Analysis for Psychology: Methods, Principles, and Practices. PsyArXiv. http://doi.org/10.31234/osf.io/h2b8t

Keusch, F., Bähr, S., Haas, G.-C., Kreuter, F., & Trappmann, M. (2020). Coverage Error in Data Collection Combining Mobile Surveys With Passive Measurement Using Apps: Data From a German National Survey. *Sociological Methods & Research*, 0049124120914924. http://doi.org/10.1177/0049124120914924

Kjell, O., Giorgi, S., & Schwartz, H. A. (2021, April). Text: An R-package for Analyzing and Visualizing Human Language Using Natural Language Processing and Deep Learning. PsyArXiv. http://doi.org/10.31234/osf.io/293kt

Kjell, O., Sikström, S., Kjell, K., & Schwartz, H. A. (2022). Natural language analyzed with AI-based transformers predict traditional subjective well-being measures approaching the theoretical upper limits in accuracy. *Scientific Reports*, *12*(1), 3918. http://doi.org/10.1038/s41598-022-07520-w

Knight, W. (2016). Amazon Working on Making Alexa Recognize Your Emotions. *MIT Technology Review*. https://www.technologyreview.com/s/601654/amazon-working-on-making-alexa-recognize-your-emotions/.

Koch, T. K., Eichstaedt, J. C., & Stachl, C. (2022). Affect Experience in Everyday Language Logged with Smartphones. http://doi.org/10.23668/psycharchives.5399

Koch, T. K., Romero, P., & Stachl, C. (2022). Age and gender in language, emoji, and emoticon usage in instant messages. *Computers in Human Behavior*, *126*, 106990. http://doi.org/10.1016/j.chb.2021.106990

Koch, T. K., & Schoedel, R. (2021). Predicting Affective States from Acoustic Voice Cues Collected with Smartphones. http://doi.org/10.23668/psycharchives.4454

Kosinski, M., Matz, S., Gosling, S., Popov, V., & Stillwell, D. (2015). Facebook as a Research Tool for the Social Sciences. *The American Psychologist*, *70*, 543–556. http://doi.org/10.1037/a0039210

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proc Natl Acad Sci U S A*, *110*(15), 5802–5. http://doi.org/10.1073/pnas.1218772110

Kralj Novak, P., Smailović, J., Sluban, B., & Mozetič, I. (2015). Sentiment of Emojis. *PLOS ONE*, *10*(12), e0144296. http://doi.org/10.1371/journal.pone.0144296

Krekhov, A., Emmerich, K., Fuchs, J., & Krueger, J. H. (2022). Interpolating Happiness: Understanding the Intensity Gradations of Face Emojis Across Cultures. In *CHI Conference on Human Factors in Computing Systems* (pp. 1–17). New York, NY, USA: Association for Computing Machinery. http://doi.org/10.1145/3491102.3517661

Kross, E., Verduyn, P., Boyer, M., Drake, B., Gainsburg, I., Vickers, B., … Jonides, J. (2019). Does counting emotion words on online social networks provide a window into people's subjective experience of emotion? A case study on Facebook. *Emotion*, *19*(1), 97–107. http://doi.org/10.1037/emo0000416

Kuppens, P., Oravecz, Z., & Tuerlinckx, F. (2010). Feelings change: Accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology*, *99*(6), 1042–1060. http://doi.org/10.1037/a0020962

Kutsuzawa, G., Umemura, H., Eto, K., & Kobayashi, Y. (2022). Classification of 74 facial emoji's emotional states on the valence-arousal axes. *Scientific Reports*, *12*. http://doi.org/10.1038/s41598-021-04357-7

Lang, M., Binder, M., Richter, J., Schratz, P., Pfisterer, F., Coors, S., … Bischl, B. (2019). Mlr3: A modern object-oriented machine learning framework in R. *Journal of Open Source Software*, *4*(44), 1903. http://doi.org/10.21105/joss.01903

Lazarevic, L., Bjekic, J., Zivanovic, M., & Knezevic, G. (2020). Ambulatory assessment of language use: Evidence on the temporal stability of Electronically Activated Recorder and stream of consciousness data. *Behavior Research Methods*. http://doi.org/10.3758/s13428-020-01361-z

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., … Van Alstyne, M. (2009). Computational Social Science. *Science*, *323*(5915), 721–723. http://doi.org/10.1126/science.1167742

Lazer, D., Pentland, A., Watts, D. J., Aral, S., Athey, S., Contractor, N., … Wagner, C. (2020). Computational social science: Obstacles and opportunities. *Science*, *369*(6507), 1060–1062. http://doi.org/10.1126/science.aaz8170

Lim, N. (2016). Cultural differences in emotion: Differences in emotional arousal level between the East and the West. *Integrative Medicine Research*, *5*(2), 105–109. http://doi.org/10.1016/j.imr.2016.03.004

Lin, Y., Ding, H., & Zhang, Y. (2020). Prosody Dominates Over Semantics in Emotion Word Processing: Evidence From Cross-Channel and Cross-Modal Stroop Effects. *Journal of Speech, Language, and Hearing Research*, *63*(3), 896–912. http://doi.org/10.1044/2020_JSLHR-19-00258

Liu, Tingting, Giorgi, S., Tao, X., Bellew, D., Curtis, B., & Ungar, L. (2022, February). Cross-Platform Difference in Facebook and Text Messages Language Use: Illustrated by Depression Diagnosis. arXiv. http://doi.org/10.48550/arXiv.2202.01802

Liu, Tony, Meyerhoff, J., Eichstaedt, J. C., Karr, C. J., Kaiser, S. M., Kording, K. P., … Ungar, L. H. (2021). The relationship between text message sentiment and self-reported depression. *Journal of Affective Disorders*. http://doi.org/10.1016/j.jad.2021.12.048

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., … Stoyanov, V. (2019, July). RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv. http://doi.org/10.48550/arXiv.1907.11692

Lucas, R. E., Le, K., & Dyrenforth, P. S. (2008). Explaining the Extraversion/Positive Affect Relation: Sociability Cannot Account for Extraverts' Greater Happiness. *Journal of Personality*, *76*(3), 385–414. http://doi.org/10.1111/j.1467-6494.2008.00490.x

Mahmoodi, J., Leckelt, M., van Zalk, M., Geukes, K., & Back, M. (2017). Big Data approaches in social and behavioral science: Four key trade-offs and a call for integration. *Current Opinion in Behavioral Sciences*, *18*, 57–62. http://doi.org/10.1016/j.cobeha.2017.07.001

Mandell, J. (2020). Spotify Patents A Voice Assistant That Can Read Your Emotions. *Forbes*. https://www.forbes.com/sites/joshmandell/2020/03/12/spotify-patents-a-voice-assistant–that-can-read-your-emotions/.

Mandi, S., Ghosh, S., De, P., & Mitra, B. (2022). Emotion detection from smartphone keyboard interactions: Role of temporal vs spectral features. In *Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing* (pp. 677–680). Virtual Event: ACM. http://doi.org/10.1145/3477314.3507159

Marrero, Z. N. K., Gosling, S. D., Pennebaker, J. W., & Harari, G. M. (2022). Evaluating voice samples as a potential source of information about personality. *Acta Psychologica*, *230*, 103740. http://doi.org/10.1016/j.actpsy.2022.103740

Massachi, T., Fong, G., Mathur, V., Pendse, S. R., Hoefer, G., Fu, J. J., … Huang, J. (2020). Sochiatrist: Signals of Affect in Messaging Data. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW2), 1–25. http://doi.org/10.1145/3415182

Mastoras, R.-E., Iakovakis, D., Hadjidimitriou, S., Charisis, V., Kassie, S., Alsaadi, T., … Hadjileontiadis, L. J. (2019). Touchscreen typing pattern analysis for remote detection of the depressive tendency. *Scientific Reports*, *9*(1), 13414. http://doi.org/10.1038/s41598-019-50002-9

Matero, M., Hung, A., & Schwartz, H. A. (2022). Evaluating Contextual Embeddings and their Extraction Layers for Depression Assessment. In *Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis* (pp. 89–94). Dublin, Ireland: Association for Computational Linguistics. http://doi.org/10.18653/v1/2022.wassa-1.9

May, J. M. (2001). *Cicero on the ideal orator*. Oxford University Press, USA.

Meegahapola, L., Droz, W., Kun, P., de Götzen, A., Nutakki, C., Diwakar, S., … Gatica-Perez, D. (2023). Generalization and Personalization of Mobile Sensing-Based Mood Inference Models: An Analysis of College Students in Eight Countries. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *6*(4), 176:1–176:32. http://doi.org/10.1145/3569483

Mehl, M. R. (2017). The Electronically Activated Recorder (EAR): A Method for the Naturalistic Observation of Daily Social Behavior. *Current Directions in Psychological Science*, *26*(2), 184–190. http://doi.org/10.1177/0963721416680611

Mehl, M. R., Pennebaker, J. W., Crow, D. M., Dabbs, J., & Price, J. H. (2001). The Electronically Activated Recorder (EAR): A device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments, & Computers*, *33*(4), 517–523. http://doi.org/10.3758/BF03195410

Mehl, M. R., Robbins, M. L., & Holleran, S. E. (2013). How Taking a Word for a Word Can Be Problematic: Context-Dependent Linguistic Markers of Extraversion and Neuroticism. *Journal of Methods and Measurement in the Social Sciences*, *3*(2). http://doi.org/10.2458/v3i2.16477

Meier, T., Boyd, R. L., Pennebaker, J. W., Mehl, M. R., Martin, M., Wolf, M., & Horn, A. B. (2019). "LIWC auf Deutsch": The Development, Psychometrics, and Introduction of DE- LIWC2015. *PsyArXiv*. http://doi.org/10.31234/osf.io/uq8zt

Miller, G. (2012). The Smartphone Psychology Manifesto. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *7*(3), 221–237. http://doi.org/10.1177/1745691612441215

Milling, M., Pokorny, F., Bartl-Pokorny, K., & Schuller, B. (2022). Is Speech the New Blood? Recent Progress in AI-Based Disease Detection From Audio in a Nutshell. *Frontiers in Digital Health*, *4*, 886615. http://doi.org/10.3389/fdgth.2022.886615

Mohammad, S. M. (2022). Ethics Sheet for Automatic Emotion Recognition and Sentiment Analysis. *Computational Linguistics*, *48*(2), 239–278. http://doi.org/10.1162/coli_a_00433

Molnar, C. (2019). *Interpretable Machine Learning*.

Montag, C., Dagum, P., Hall, B. J., & Elhai, J. D. (2022). Do we still need psychological self-report questionnaires in the age of the Internet of Things? *Discover Psychology*, *2*(1), 1. http://doi.org/10.1007/s44202-021-00012-4

Muaremi, A., Gravenhorst, F., Grünerbl, A., Arnrich, B., & Tröster, G. (2014). *Assessing Bipolar Episodes Using Speech Cues Derived from Phone Calls. Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST* (Vol. 100). http://doi.org/10.1007/978-3-319-11564-1_11

Müller, S. R., Chen, X. L., Peters, H., Chaintreau, A., & Matz, S. C. (2021). Depression predictions from GPS-based mobility do not generalize well to large demographically heterogeneous samples. *Scientific Reports*, *11*(1), 14007. http://doi.org/10.1038/s41598-021-93087-x

Nadeau, C., & Bengio, Y. (2003). Inference for the Generalization Error. *Machine Learning*, *52*(3), 239–281. http://doi.org/10.1023/A:1024068626366

Niu, W., Kong, Z., Yuan, G., Jiang, W., Guan, J., Ding, C., … Wang, Y. (2020, October). Real-Time Execution of Large-scale Language Models on Mobile. arXiv. Retrieved from `https://arxiv.org/abs/arXiv:2009.06823`

Parthasarathy, S., Rozgic, V., Sun, M., & Wang, C. (2019). Improving Emotion Classification through Variational Inference of Latent Variables. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing*

(ICASSP) (pp. 7410–7414). http://doi.org/10.1109/ICASSP.2019.8682823

Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015). Linguistic Inquiry and Word Count: LIWC 2015 [Computer software]. Pennebaker Conglomerates. Inc.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, *77*(6), 1296–1312. http://doi.org/10.1037/0022-3514.77.6.1296

Petrizzo, D., & Popolo, P. S. (2021). Smartphone Use in Clinical Voice Recording and Acoustic Analysis: A Literature Review. *Journal of Voice*, *35*(3), 499.e23–499.e28. http://doi.org/10.1016/j.jvoice.2019.10.006

Picard, R. W. (2000). *Affective computing*. MIT press.

Piccirillo, M. L., & Rodebaugh, T. L. (2019). Foundations of idiographic methods in psychology and applications for psychotherapy. *Clinical Psychology Review*, *71*, 90–100. http://doi.org/10.1016/j.cpr.2019.01.002

Pisano, F., Manfredini, A., Brachi, D., Landi, L., Sorrentino, L., Bottone, M., … Marangolo, P. (2022). How Has COVID-19 Impacted Our Language Use? *International Journal of Environmental Research and Public Health*, *19*(21), 13836. http://doi.org/10.3390/ijerph192113836

Poonguzhali, N., Gayathri, S., Deebika, A., & Suriapriya, R. (2020). A Framework For Electronic Health Record Using Blockchain Technology. In *2020 International Conference on System, Computation, Automation and Networking (ICSCAN)* (pp. 1–5). http://doi.org/10.1109/ICSCAN49426.2020.9262369

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, *17*(3), 715–734. http://doi.org/10.1017/S0954579405050340

Preoţiuc-Pietro, D., Schwartz, H. A., Park, G., Eichstaedt, J. C., Kern, M., Ungar, L., & Shulman, E. (2016). Modelling Valence and Arousal in Facebook posts. In *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* (pp. 9–15). San Diego, California: Association for Computational Linguistics. http://doi.org/10.18653/v1/W16-0404

Qiu, L., Lin, H., Leung, A. K., & Tov, W. (2012). Putting Their Best Foot Forward: Emotional Disclosure on Facebook. *Cyberpsychology, Behavior, and Social Networking*, *15*(10), 569–572. http://doi.org/10.1089/cyber.2012.0200

R Core Team. (2021). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Rajkumar, R. P. (2020). COVID-19 and mental health: A review of the existing literature. *Asian Journal of Psychiatry*, *52*, 102066. http://doi.org/10.1016/j.ajp.2020.102066

Reeves, B., Ram, N., Robinson, T. N., Cummings, J. J., Giles, C. L., Pan, J., . . . Yeykelis, L. (2021). Screenomics: A Framework to Capture and Analyze Personal Life Experiences and the Ways that Technology Shapes Them. *Human–Computer Interaction*, *36*(2), 150–201. http://doi.org/10.1080/07370024.2019.1578652

Remus, R., Quasthoff, U., & Heyer, G. (2010). SentiWS - A Publicly Available German-language Resource for Sentiment Analysis. In *LREC*.

Riordan, M. A. (2017). Emojis as Tools for Emotion Work: Communicating Affect in Text Messages. *Journal of Language and Social Psychology*, *36*(5), 549–567. http://doi.org/10.1177/0261927X17704238

Romero, P., Mikiya, Y., Nakatsuma, T., Fitz, S., & Koch, T. K. (2021, July). Modelling Personality Change During Extreme Exogenous Conditions. PsyArXiv. http://doi.org/10.31234/osf.io/rtmjw

Russell, J. A. (1980). A Circumplex Model of Affect. *Journal of Personality and Social Psychology*, *39*, 1161–1178. http://doi.org/10.1037/h0077714

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, *76*, 805–819. http://doi.org/10.1037/0022-3514.76.5.805

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, *99*(2), 143–165. http://doi.org/10.1037/0033-2909.99.2.143

Scherer, K. R. (2000). Psychological models of emotion. *The Neuropsychology of Emotion*, *137*(3), 137–162.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, *40*(1-2), 227–256. http://doi.org/10.1016/S0167-6393(02)00084-5

Schoedel, R., Kunz, F., Bergmann, M., Bemmann, F., Bühner, M., & Sust, L. (2022, August). Snapshots of Daily Life: Situations Investigated Through the Lens of Smartphone Sensing. PsyArXiv. http://doi.org/10.31234/osf.io/f3htz

Schoedel, R., & Oldemeier, M. (2020). Basic Protocol: Smartphone Sensing Panel Study. http://doi.org/10.23668/psycharchives.2901

Schuller, B. (2018). Speech emotion recognition: Two decades in a nutshell, benchmarks, and ongoing trends. *Communications of the ACM*, *61*(5), 90–99. http://doi.org/10.1145/3129340

Schuller, B., Batliner, A., Steidl, S., & Seppi, D. (2011). Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge. *Speech Communication*, *53*(9), 1062–1087. http://doi.org/10.1016/j.specom.2011.01.011

Schuller, B., Steidl, S., Batliner, A., Hirschberg, J., Burgoon, J., Baird, A., … Evanini, K. (2016). *The INTERSPEECH 2016 Computational Paralinguistics Challenge: Deception, Sincerity and Native Language* (p. 2005). http://doi.org/10.21437/Interspeech.2016-129

Schwartz, R., & Pell, M. D. (2012). Emotional Speech Processing at the Intersection of Prosody and Semantics. *PLoS ONE*, *7*(10), e47279. http://doi.org/10.1371/journal.pone.0047279

Seifert, A., Hofer, M., & Allemand, M. (2018). Mobile Data Collection: Smart, but Not (Yet) Smart Enough. *Frontiers in Neuroscience*, *12*.

Shen, P., Changjun, Z., & Chen, X. (2011). Automatic Speech Emotion Recognition using Support Vector Machine. In *Proceedings of 2011 International Conference on Electronic & Mechanical Engineering and Information Technology* (Vol. 2, pp. 621–625). http://doi.org/10.1109/EMEIT.2011.6023178

Shrestha, Y. R., He, V. F., Puranam, P., & von Krogh, G. (2021). Algorithm Supported Induction for Building Theory: How Can We Use Prediction Models to Theorize? *Organization Science*, *32*(3), 856–880. http://doi.org/10.1287/orsc.2020.1382

Siegel, E. H., Sands, M. K., Van den Noortgate, W., Condon, P., Chang, Y., Dy, J., … Barrett, L. F. (2018). Emotion fingerprints or emotion populations? A meta-analytic investigation of autonomic features of emotion categories. *Psychological Bulletin*, *144*, 343–393. http://doi.org/10.1037/bul0000128

Skinner, E. R. (1935). A calibrated recording and analysis of the pitch, force and quality of vocal tones expressing happiness and sadness; and a determination of the pitch and force of the subjective concepts of ordinary, soft and loud tones. *Speech Monographs*, *2*, 81–137. http://doi.org/10.1080/03637753509374833

Sridhar, K., & Busso, C. (2022). Unsupervised Personalization of an Emotion Recognition System: The Unique Properties of the Externalization of Valence in Speech. *IEEE Transactions on Affective Computing*, *13*(4), 1959–1972. http://doi.org/10.1109/TAFFC.2022.3187336

Stark, L., & Hoey, J. (2021). The Ethics of Emotion in Artificial Intelligence Systems. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 782–793). Virtual Event Canada: ACM. http://doi.org/10.1145/3442188.3445939

Stone, P. J., Bales, R. F., Namenwirth, J. Z., & Ogilvie, D. M. (1962). The general inquirer: A computer system for content analysis and retrieval based on the sentence as a unit of information. *Behavioral Science*, *7*(4), 484–498. http://doi.org/10.1002/bs.3830070412

Suhara, Y., Xu, Y., & Pentland, A. 'Sandy'. (2017). DeepMood: Forecasting Depressed Mood Based on Self-Reported Histories via Recurrent Neural Networks. In *Proceedings of the 26th International Conference on World Wide Web* (pp. 715–724). Perth Australia: International World Wide Web Conferences Steering Committee. http://doi.org/10.1145/3038912.3052676

Sun, J., Schwartz, H. A., Son, Y., Kern, M. L., & Vazire, S. (2020). The language of well-being: Tracking fluctuations in emotion experience through everyday speech. *Journal of Personality and Social Psychology*, *118*(2), 364–387. http://doi.org/10.1037/pspp0000244

Sun, T., Allix, K., Kim, K., Zhou, X., Kim, D., Lo, D., . . . Klein, J. (2022, December). A Pre-Trained BERT Model for Android Applications. arXiv. Retrieved from https://arxiv.org/abs/arXiv:2212.05976

Tackman, A., Sbarra, D., Carey, A., Donnellan, M., Horn, A., Holtzman, N., . . . Mehl, M. R. (2018). Depression, Negative Emotionality, and Self-Referential Language: A Multi-Lab, Multi-Measure, and Multi-Language-Task Research Synthesis. *Journal of Personality and Social Psychology*, *116*. http://doi.org/10.1037/pspp0000187

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.

Tov, W., Ng, K. L., Lin, H., & Qiu, L. (2013). Detecting well-being via computerized content analysis of brief diary entries. *Psychological Assessment*, *25*(4), 1069–1078. http://doi.org/10.1037/a0033007

Tzirakis, P., Chen, J., Zafeiriou, S., & Schuller, B. (2021). End-to-end multimodal affect recognition in real-world environments. *Information Fusion*, *68*, 46–53. http://doi.org/10.1016/j.inffus.2020.10.011

Van Berkel, N., Ferreira, D., & Kostakos, V. (2017). The Experience Sampling Method on Mobile Devices. *ACM Computing Surveys*, *50*(6), 1–40. http://doi.org/10.1145/3123988

Verheijen, L., & Stoop, W. (2016). Collecting Facebook Posts and WhatsApp Chats. In *Text, Speech, and Dialogue* (pp. 249–258). Cham: Springer.

Vinciarelli, A., & Mohammadi, G. (2014). A Survey of Personality Computing. *IEEE Transactions on Affective Computing*, *5*(3), 273–291. http://doi.org/10.1109/TAFFC.2014.2330816

Vine, V., Boyd, R. L., & Pennebaker, J. W. (2020). Natural emotion vocabularies as windows on distress and well-being. *Nature Communications*, *11*(1), 4525.

http://doi.org/10.1038/s41467-020-18349-0

Vlahos, J. (2019). *Talk to Me: How Voice Computing Will Transform the Way We Live, Work, and Think*. Eamon Dolan Books.

Vogt, T., André, E., & Wagner, J. (2008). Automatic Recognition of Emotions from Speech: A Review of the Literature and Recommendations for Practical Realisation. In C. Peter & R. Beale (Eds.), *Affect and Emotion in Human-Computer Interaction* (Vol. 4868, pp. 75–91). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-540-85099-1_7

Wang, R., Chen, F., Chen, Z., Li, T., Harari, G. M., Tignor, S., … Campbell, A. T. (2014). StudentLife: Assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 3–14). Seattle Washington: ACM. http://doi.org/10.1145/2632048.2632054

Weidman, A. C., Sun, J., Vazire, S., Quoidbach, J., Ungar, L. H., & Dunn, E. W. (2020). (Not) hearing happiness: Predicting fluctuations in happy mood from acoustic cues using machine learning. *Emotion (Washington, D.C.)*, *20*(4), 642–658. http://doi.org/10.1037/emo0000571

Weninger, F., Eyben, F., Schuller, B. W., Mortillaro, M., & Scherer, K. R. (2013). On the Acoustics of Emotion in Audio: What Speech, Music, and Sound have in Common. *Frontiers in Psychology*, *4*. http://doi.org/10.3389/fpsyg.2013.00292

Wiggers, K. (2022, January). New startup shows how emotion-detecting AI is intrinsically problematic. *VentureBeat*.

Wilting, J., Krahmer, E. J., & Swerts, M. G. J. (2006). Real vs. Acted emotional speech. *Proceedings of the International Conference on Spoken Language Processing (Interspeech 2006)*.

Wright, M. N., & Ziegler, A. (2017). Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software*, *77*(1), 1–17. http://doi.org/10.18637/jss.v077.i01

Wright, M. N., Ziegler, A., & König, I. R. (2016). Do little interactions get lost in dark random forests? *BMC Bioinformatics*, *17*(1), 145. http://doi.org/10.1186/s12859-016-0995-8

Wu, C., Fritz, H., Bastami, S., Maestre, J. P., Thomaz, E., Julien, C., … Nagy, Z. (2021). Multi-modal data collection for measuring health, behavior, and living environment of large-scale participant cohorts. *GigaScience*, *10*(6). http://doi.org/10.1093/gigascience/giab044

Yang, K., Wang, C., Gu, Y., Sarsenbayeva, Z., Tag, B., Dingler, T., … Goncalves, J. (2021). Behavioral and Physiological Signals-Based Deep Multimodal Approach for Mobile Emotion Recognition. *IEEE Transactions on Affective Computing*, 1–1. http://doi.org/10.1109/TAFFC.2021.3100868

Yarkoni, T., & Westfall, J. (2017). Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122. http://doi.org/10.1177/1745691617693393

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(2), 301–320. http://doi.org/10.1111/j.1467-9868.2005.00503.x