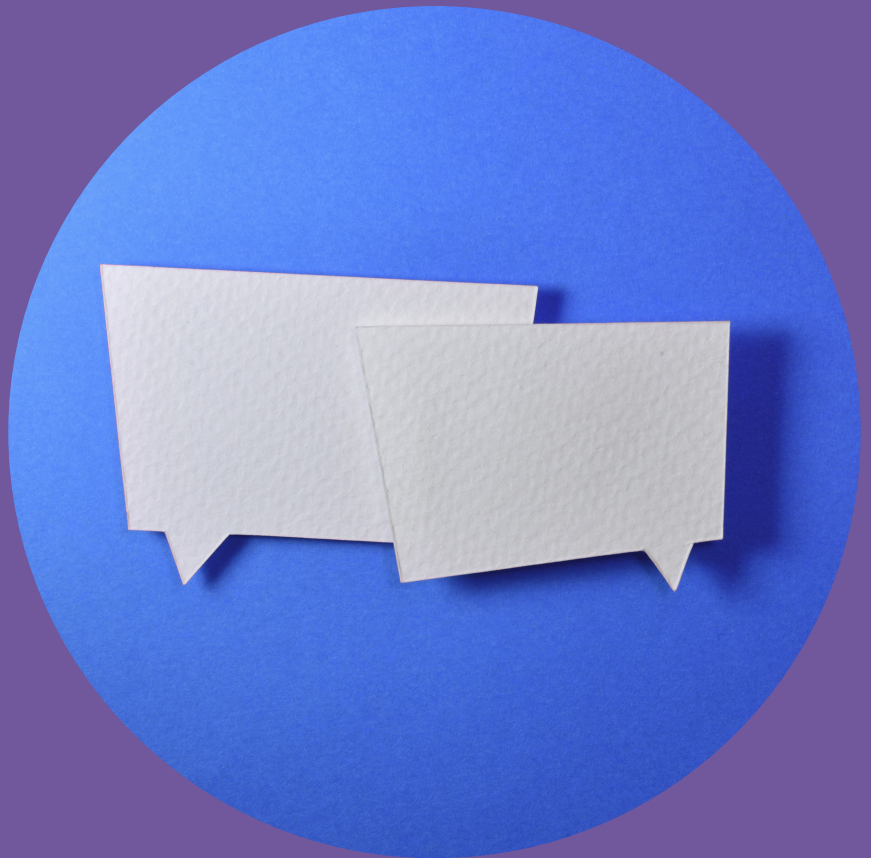


Investigating interpersonal accuracy in design and music performance.

Contextual influences in mutual understanding

Álvaro M. Chang-Arana



Investigating interpersonal accuracy in design and music performance.

Contextual influences in mutual understanding

Álvaro M. Chang-Arana

A doctoral thesis completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Science, at a public examination held at the R003/F239a Auditorio of the school on 27 March 2023 at 12.

Aalto University
School of Science
Department of Neuroscience and Biomedical Engineering

Supervising professor

Professor Iiro Jääskeläinen, Aalto University, Finland.

Thesis advisors

Professor Katja Hölttä-Otto, The University of Melbourne, Australia.

Professor Mikko Sams, Aalto University, Finland.

Preliminary examiners

Professor Jari Hietanen, Tampere University, Finland.

Doctor Jari Jussila, Häme University of Applied Sciences, Design Factory, Finland.

Opponent

Doctor Jari Jussila, Häme University of Applied Sciences, Design Factory, Finland.

Aalto University publication series

DOCTORAL THESES 22/2023

© 2023 Álvaro M. Chang-Arana

ISBN 978-952-64-1156-9 (printed)

ISBN 978-952-64-1157-6 (pdf)

ISSN 1799-4934 (printed)

ISSN 1799-4942 (pdf)

<http://urn.fi/URN:ISBN:978-952-64-1157-6>

Images: Cover photo by Miguel Á. Padriñán from Pexels:

<https://www.pexels.com/photo/two-white-message-balloons-1111368/> Edited by Charo Arana Ruiz

Unigrafia Oy

Helsinki 2023

Finland



Author

Álvaro M. Chang-Arana

Name of the doctoral thesis

Investigating interpersonal accuracy in design and music performance.

Publisher School of Science

Unit Department of Neuroscience and Biomedical Engineering

Series Aalto University publication series DOCTORAL THESES 22/2023

Field of research Social psychology

Manuscript submitted 18 October 2022

Date of the defence 27 March 2023

Permission for public defence granted (date) 1 February 2023

Language English

☐ **Monograph**

☒ **Article thesis**

☐ **Essay thesis**

Abstract

We constantly attempt to know what someone else may be thinking or feeling, what kind of personality they have, what their beliefs are, etc. Despite how common this "mind reading" process is to us we are surprisingly inaccurate when inferring someone else's mental states. The correct understanding of others' mental states – interpersonal accuracy – is key for successful social interactions and its scientific study demands a complex balance between controlled experimental and naturalistic conditions. Design and music performance are two contexts yet unexplored through the lenses of interpersonal accuracy.

In design, being interpersonally accurate towards users is deemed important for design outcomes. User-understanding is broadly referred to as empathy in design. Yet, empathy is not clearly defined. We expose this problem and suggest some conceptual clarity. In music performance, interpersonal accuracy allows us to better understand the complex communication between musician-listener. A musician experiences emotions while performing, but it is not known whether listeners can detect these accurately. We adapt two interpersonal accuracy methodologies, empathic accuracy, and emotional recognition accuracy. Empathic accuracy allows measuring the similarity between remembered and inferred mental contents of interacting dyads. Emotional recognition accuracy allows to measure the accurate judgment of someone's non-verbal emotional expressions.

Through adapting empathic accuracy in design cases, it was observed that designers obtained approximately 50% accuracy, and showed higher accuracy when inferring design-related mental contents than mental contents irrelevant to design. In comparison to previous empathic accuracy literature, designers obtained higher empathic accuracy scores. We attribute these to contextual cues such as awareness of the conversation topic, and the demonstration of concrete objects. Although the causal link between designers' empathic accuracy and design outcomes remains unestablished. Through adapting emotional recognition accuracy into music performance, it was observed that listeners perceived lower anxiety than that reported by the musician across experimental conditions. Furthermore, the listener's emotional recognition accuracy is a complex skill affected by variables such as multimodal perception, and the listener's musical background. Altogether, inaccuracy was observed across the context of design and music. Interpersonal accuracy can also be affected by multimodal perception and the perceiver's background. We conclude suggesting some ideas to improve interpersonal accuracy.

Keywords Interpersonal accuracy, empathy, design, music performance, empathic accuracy, emotional recognition accuracy.

ISBN (printed) 978-952-64-1156-9

ISBN (pdf) 978-952-64-1157-6

ISSN (printed) 1799-4934

ISSN (pdf) 1799-4942

Location of publisher Helsinki

Location of printing Helsinki **Year** 2023

Pages 180

urn <http://urn.fi/URN:ISBN:978-952-64-1157-6>

A Lorena, Alicia y Mario.
Gracias por ser los pilares de todo lo que fui, soy y seré.

Acknowledgments

Since [Franz] Liszt wanted to give his pupils the best possible education, he always tried to introduce them to distinguished visitors who happened to pass through Weimar. In this way, these young men were frequently brought into touch with poets, painters, dramatists, scientists, and politicians, as well as musicians, whom they would never have met under normal circumstances. No other teacher of the day would have done that. It was all part of Liszt's well-known view that piano playing must never be confused with finger-dexterity, that piano playing involved the whole person, and to improve the one you must improve the other. (Walker, 1994, p. 189).

This view was first introduced to me by my piano teacher, Pablo Sabat. I was on my early 20's, if not younger, and fascinated by the idea that everything I could experience would nurture my musicianship. Little I knew that it would also prompt my development as an interdisciplinary scientist. It was even less imaginable that I would end up becoming a music psychologist and migrating to Finland to complete my higher education.

My journey would not have been possible without the support of other people. I would like to first thank Professor Katja Hölttä-Otto. It was thanks to her bold decision to trust a music psychologist to work on a design project that I can write these lines. Katja, monet kiitokset, kun sinä esitit minulle uutta tutkimusaihetta ja haastoit minut tutkijana. I want to thank Professor Mikko Sams for reminding me to think about the wider picture and trusting I would be capable of concluding the process. Mikko, minä nautin todellisesti meidän keskusteluitamme tieteestä, musiikista, ja elämästä. Monet kiitokset näistä. I would also like to thank him for accepting being custos of my defense.

Professor Iiro Jääskeläinen, thank you for supporting me during the final steps of my doctoral studies. I would also like to thank Professor Jari Hietanen and Doctor Jari Jussila for reviewing my thesis and providing valuable feedback. Moreover, I would like to thank Doctor Jari Jussila for accepting being my opponent and engage in an interesting discussion. I am grateful to Doctor Enrico Glerean for giving me valuable methodological suggestions, as well as advising me on how to start writing this thesis. My gratitude also goes to my colleagues, co-authors, and participants of our studies. Thanks to your expertise, time, and support we were able to share exciting research with others.

This work is also a homage to the lecturers and teachers who inspired me since early in my academic career: Norma Reátegui, Orielle Montezuma, Lennia

Matos, Ricardo Braun, Pablo Zumaeta, Andrés Burga, and Pablo Sabat. Your examples keep influencing me deeply. You encouraged me to, literally, travel far in the pursuit of my curiosity.

I want to acknowledge the people who helped me find balance in my life during my doctoral studies. I want to thank Rebekka Angervo, Niklas Pokki, David Munk-Nielsen, Juhani Rantanen, and Joona Uusitalo for creating the space that allowed me to keep developing my musicianship and restoring my spirits. I am deeply grateful to great friends and colleagues who loved and supported me during these past years, helped me overcome challenging times, and shared with me wonderful memories: Jéssica Vínces, Víctor Talavera, Lidia Calderón, Graciela Valderrama, Anastasios Mavrolampados, Deniz Duman, Floris van der Marel, Marthe Dehli, Antti Surma-aho, Jie Li, Lauri Partanen, Emma Allingham, Anna Czepil, Megan Buchkowski, Laure Mostinckx, Carla Colomer, Itziar Villar, Joel Brassard, Mónica Bene, Luis Miguel Rojas-Berscia, and Oscar Cajo. I specially want to thank Rita Barrera, mi amiguis del mal. You have been my steady column during these past years. I feel grateful for your friendship and trust.

Lastly, my most heartfelt gratitude goes to my family, Lorena, Alicia, and Mario. I had the fortune of being raised in a space of unconditional love and support. You granted me the security, time, joy, and resources to explore my different interests. Your love will always be a stabilizing force in my life.

Lima, 8 February 2023
Álvaro M. Chang-Arana

Contents

Acknowledgments.....	ii
List of Abbreviations and Symbols	vi
List of Publications	vii
Author's Contribution.....	viii
1. Introduction.....	1
The indelible stamp of our social origin	1
Interpersonal accuracy	2
The context of empathy in design	4
The context of music performance perception	5
Defining the golden standards	6
1.1 Empathic accuracy.....	7
1.2 Emotion recognition accuracy	8
2. Objectives	10
3. Summary of studies.....	12
3.1 Study I – Under the umbrella: Components of empathy in psychology and design.....	12
3.2 Study II – Empathic accuracy in design: Exploring design outcomes through empathic performance and physiology	17
3.3 Study III – Reading the user's mind: Designers show high accuracy in inferring design-related thoughts and feelings	24
3.4 Study IV – Exploring the interpersonal level of music performance anxiety: Online listener's accuracy in detecting performer anxiety ..	28
4. Discussion	34
4.1 How to implement an IA approach to quantify the understanding between dyads interacting in design and music?.....	34
4.2 How accurate are perceivers when inferring the mental experiences of targets?.....	35
4.3 What contextual factors affect IA?	37
4.4 Future directions and limitations	38
4.5 Recommendations	39

4.5.1	Ask. Do not assume.	40
4.5.2	Exemplify and structure.	40
4.5.3	Context matters.	40
5.	General conclusions.....	41
6.	References	42

List of Abbreviations and Symbols

EA	Empathic accuracy
ERA	Emotion recognition accuracy
ECG	Electrocardiogram
EMG	Electromyography
GSR	Galvanic skin response
IA	Interpersonal accuracy
MPA	Music performance anxiety
Op.	Opus

List of Publications

This doctoral dissertation consists of a summary of the following publications:

Study I: Chang-Arana, Á. M., Surma-aho, A., Hölttä-Otto, K., & Sams, M. (2022). Under the umbrella: Components of empathy in psychology and design. *Design Science*, 8, e20.

Study II: Chang-Arana, Á. M., Piispanen, M., Himberg, T., Surma-aho, A., Alho, J., Sams, M., & Hölttä-Otto, K. (2020). Empathic accuracy in design: Exploring design outcomes through empathic performance and physiology. *Design Science*, 6, e16.

Study III: Chang-Arana, Á. M., Surma-aho, A., Li, J., Yang, M. C., & Hölttä-Otto, K. (2020). Reading the users' mind: Designers show high accuracy in inferring design-related thoughts and feelings. *ASME 2020 IDETC/CIE Conference*, 17th-19th August.

Study IV: Chang-Arana, Á. M., Mavrolampados, A., Thompson, M. R., Pokki, N., & Sams, M. (2022). Exploring the interpersonal level of music performance anxiety: Online listener's perception and accurate inference of anxiety. *Frontiers in Psychology*, 13:838041.

Author's Contribution

Publication 1: Under the umbrella: Components of empathy in psychology and design.

The candidate proposed the idea of writing this literature review, reviewed the literature and mapped empathy from psychology to design, and wrote the manuscript. Antti Surma-aho wrote valuable input and helped structure the manuscript's final form. Katja Hölttä-Otto and Mikko Sams actively provided feedback, contributed with content, and suggested further improvements.

Publication 2: Empathic accuracy in design: Exploring design outcomes through empathic performance and physiology

The candidate designed the study, gathered, and analyzed the participants' behavioral data, and wrote the manuscript. Matias Piispanen assisted in the data gathering process, created the code for processing and analyzing the physiological data, provided valuable input to the manuscript and gave feedback. Tommi Himberg assisted in the data collection planning by introducing us to different measuring equipment. He assisted Matias Piispanen in creating the code and analyzing data. Antti Surma-aho assisted in data collection and providing valuable input to the manuscript. Jussi Alho assisted in the data collection and processing planning, and assisted Matias Piispanen. Mikko Sams assisted in the initial planning and execution of the study. Katja Hölttä-Otto assisted in the initial planning of the study and supervised its development. All co-authors contributed to the preparation of the manuscript and provided valuable feedback.

Publication 3: Reading the users' mind: Designers show high accuracy in inferring design-related thoughts and feelings.

The candidate and Antti Surma-aho share the first authorship. We designed the study, collected, and analyzed the data, and wrote the manuscript. The candidate collected the data for the musicians' sub-study, Antti Surma-aho collected the data for the visual impairment sub-study. Jie Li collected the data for the driver sub-study. Maria C. Yang facilitated Antti Surma-aho the access to the visual impairment sub-study data. Katja Hölttä-Otto supervised the development of the study. All co-authors contributed to the preparation of the manuscript and provided valuable feedback.

Publication 4: Exploring the interpersonal level of music performance anxiety: Online listeners' perception and accurate inference of anxiety.

The candidate designed the study, performed Beethoven's piano sonata in F major op. 10 no. 2, gathered and analyzed the data, and wrote the manuscript. Anastasios Mavrolampados wrote the code for analyzing the music stimuli. Marc R. Thompson assisted in the initial planning of the study, and subsequent developments. Niklas Pokki assisted in the initial planning of the study, participated, and provided access to professional pianists who acted as judges in the study. Mikko Sams actively helped in shaping the paper and clarifying its content. All co-authors provided valuable feedback.

1. Introduction

The indelible stamp of our social origin¹

Humans are obsessed with reading other people's mind, or at least trying to. A young teenager who transfers schools asks themselves who is the popular kid; the popular kid at school wonders who is this new student and whether they would hit it off. An employee wonders how their request for a higher salary will be received by the supervisor; the supervisor's mind wonders out of the meeting to the newly opened position and ask themselves if the selected candidate can be trusted. One man finds another one attractive and wonders if it will be safe to ask him out or will risk being "outed" and exposed to physical damage; the other man thinks the other man across the hall is captivating but wonders too if he represents a threat. Every day, everywhere, we make innumerable inferences about other people regardless of whether they are correct or not.

We are profoundly social animals, and in our origins, we had existential pressures to develop interpersonal capacities. The problem-solving human brain was developed on an environment where the presence of other humans was prominent (Schaller, Park, & Kenrick, 2007). Coexistence pushed us to either adapt to benefit from social opportunities (e.g., securing more resources) or adapt to avoid social perils (e.g., being ostracized from the group) (Schaller et al., 2007). The prize of failing to adapt to these group challenges were probably life-threatening.

We have managed to create all kinds of fantastic technologies which have pushed us away from the fresh wounds and memories of our dangerous origins, creating an illusion of high control over our lives (Pigliucci, 2017). Yet, even if we can photograph blackholes, create nanorobots to fight diseases, and build magnificent concert halls, inside we are still that vulnerable human longing for connection. We thrive when having a sense of belonging and acknowledgment from a group (Ryan & Deci, 2017). Conversely, our well-being suffers enormously in loneliness (Cacioppo & Patrick, 2008) and when feeling excluded or rejected (Bernstein & Claypool, 2012; Gerber & Wheeler, 2009; MacDonald & Leary, 2005).

Therefore, we still have strong motivations to cooperate with others and to do so successfully. One necessary condition for a successful interaction is to interpret someone else's intentions *correctly*. Research suggests that the more

¹ Based on Darwin (1871/1981).

accurate we are at interpreting other's mental states, the better social outcomes we can obtain. Clinicians with high interpersonal accuracy obtain better diagnosis, treatments, and strong relationships with patients (Ruben, 2016). In the workplace, highly interpersonal individuals also tend to be better salespeople, negotiators and leaders (Schmid Mast & Hall, 2018). Leaders with high interpersonal accuracy tend to have more satisfied collaborators (Schmid Mast & Latu, 2016). There are other positive effects of high interpersonal accuracy on social interaction outcomes, such as more effective communications, establishing better rapport with an interaction partner, a better coordination of activities, more supportive behavior, and more resistance to stereotype biases when evaluating other people (Schmid Mast & Hall, 2018). For certain psychotherapy approaches, being interpersonally accurate is linked to positive changes in depressive symptoms and social adjustment (Crits-Christoph, Gibbons, Temes, Elkin, & Gallop, 2010).

Interpersonal accuracy

Interpersonal accuracy (IA) is the "accurate judgment about any verifiable characteristic of a person or about the group that a person belongs to" (Hall, Mast, & West, 2016, p. 5). An "accurate judgment" depends on the specific research context and terms under which it was operationalized: "Accuracy is an abstract construct that is always, and necessarily, instantiated in an operational definition" (Hall et al., 2016, p. 6). Consequently, the field of IA is an archipelago (Scriven, 1969) of research areas and methods, rather than a unified research tradition.

Yet, it is possible to identify three common elements in IA studies: 1) people from which we infer something, which can be referred to as "targets" or "encoders"; 2) the people that infer some characteristic from the targets, which can be referred to as "judges", "perceivers" or "decoders"; and 3) a "golden standard" criterium which allows to evaluate the accuracy of inferences (Hall et al., 2016). Typically, in an IA study a perceiver is asked to infer something about a target's behavior (in video, photographs, audio recordings, etc.). This "something" can be personality traits, emotions, thoughts, feelings, etc. The perceiver's inference is then compared to a golden standard or the operationalized criterium to quantify the perceiver's accuracy (Schmid Mast & Hall, 2018).

Literature consistently shows that we are rather inaccurate when inferring someone else's mental contents. This is true even when the perceiver thoroughly tries to imagine the target's mental contents, and regardless of the target being well-known to the perceiver (Damen, Pollman, & Grassow, 2021). Perceivers sharing similar experiences to the target's is associated with lower IA because, rather than asking, the perceiver is likely to use their own experiences as a frame of reference when processing the target's memories, likely biasing this perception and interfering with processing the new information from the target (Damen et al., 2021; Damen, van Amelsvoort, van der Wijst, Pollmann, & Krahmer, 2021; Kang, Cakar, Shumaker, Brook O'Donnell, & Falk, 2022).

Design is a discipline where correctly inferring a target's (i.e., user) mental contents is particularly crucial for future design outcomes. Designers are

encouraged to remember that their own experiences are different from those of the users, and these should not be confused (McDonagh, Woodcock, & Iqbal 2018). Yet, while designers are aware of the importance of correctly inferring a user's mental states for their practice, they disagree on what does it mean to understand a user. In lack of a better word, designers adopted the term "empathy", first originated in aesthetics, and further studied by psychologists over the past decades (Stueber, 2019). Although there is a general agreement in the field that empathy has a cognitive and an affective component (Kouprie & Sleeswijk Visser, 2009), it may adopt multiple meanings. Thus, design lost precision when addressing the question of how does correctly inferring a user's mental states influences design outcomes. IA can provide a frame for gaining specificity and quantifying the accuracy of designers when inferring a user's mental contents.

IA is influenced by individual characteristics of the perceivers. Individuals who perceive themselves as belonging to a low socioeconomical status (in contrast to individuals who perceive themselves as belonging to higher socioeconomical status) showed high IA when inferring different attributes from targets, such as their sexual orientation, and scored higher in tests measuring accurate inference of emotions from eye sights and recognizing feelings, attitudes and intentions from non-verbal behaviors, such as face, voice, gestures and body postures (Bänziger, Scherer, Hall, & Rosenthal, 2011; Bjornsdottir, Alaei, & Rule, 2017). Men who are violent to their female partners show lower accuracy when inferring their thoughts and feelings but are equally accurate than other men when inferring thoughts and feelings of female strangers. That is, violent men's low accuracy is particular to the context of intimate relationships (Clements, Holtzworth-Munroe, Schweinle, & Ickes, 2007). Relatedly, the accurate inference of thoughts and feelings relates to relationship satisfaction. This satisfaction is particularly stronger when couples are accurate inferring negative emotions than positive ones (Sened et al., 2017).

IA is also influenced by the perceiver's contextual factors. As reviewed by Schmid (2016), these factors may be motivational, cognitive states, affective states, and motor. Motivational factors refer to, for instance, the conscious effort perceivers make to try being accurate. This has shown to be somehow helpful in achieving IA. Cognitive states refer to different information processing styles that may intervene while in IA tasks. Studies suggest that perceivers will show high IA when they can allocate their full cognitive resources to the inference task. Affective states refer to the perceiver's emotional state when decoding a target. The perceiver's positive affect may be beneficial for non-verbal decoding accuracy, whereas negative affect has a contrary effect. Regarding motor factors, recognizing emotions does not require mimicking someone else's facial movements. Yet, mimicking body postures was associated with higher empathic accuracy, a type of IA metric (Fujiwara & Daibo, 2021).

Studies on music performance perception may enable to deepen into contextual variables influencing IA. The listener's perception of a performance is influenced by the sensorial channels through which the performance is communicated, as well as the musical background of the listener (Broughton & Stevens, 2009; Stanley et al., 2002; Wapnick et al., 2004). Conversely, the emotions experienced by a musician may affect their musical outputs and body movements,

both of which may influence the listener's perception (Dahl and Friberg, 2007; Vines et al., 2011; Van Zijl et al., 2014). What remains to be investigated is how the listener's perception differs from a musician's experiences while performing. IA can provide a frame to investigate variables influencing the complex communication occurring in musical contexts. Next, we will explain in further detail the state of the art in empathy in design and music performance research.

The context of empathy in design

In design, empathy is deemed as necessary for successful outcomes especially during the early stages of the design process (Brown 2009; Kramer, Agolino & Roschuni, 2016). Thus, empathy has become a central concept in design research and practice over the past two decades (Köppen & Meinel, 2015). Its complicated nature is often highlighted in the literature (Surma-aho & Hölttä-Otto, 2022).

Generally, it is understood as an approach to obtain better design outcomes through matching the mental contents and imagination of designers and users (Heylighen & Dong, 2019). It has also been defined as an emotional understanding of end-users achieved by immersing in the context and mental life of users (Battarbee et al., 2002), internalizing user's requirements (Battarbee, 2004), and a support for requirement definition and concept generation (Fila & Hess, 2014). Designers have also incorporated insights from psychology. This can be seen in design research which defines empathy as being composed of an affective and cognitive component (Kouprie & Sleeswijk Visser, 2009) or as perspective-taking (Köppen & Meinel, 2015).

The literature of empathy in design does not limit to conceptual discussions. It also includes different technological approaches aimed at increasing or tuning designers' empathy towards users. These include wearable simulators which purpose is to "engage more closely with user experiences" (Kullman 2016, p. 83) by restricting the designer's physical capabilities, creating experience simulators (Smeenk, Sturm & Eggen, 2017), role-playing (Medler & Magerko, 2010), personas defined as "abstractions of groups of real consumers who share common characteristics and needs" (Miaskiewicz & Kozar 2011, p. 418), storytelling (Madsen & Nielsen 2010), empathy maps (Both & Baggereor n.d.), etc. In addition, designers have developed multi-step frameworks which, if followed, should facilitate the empathic connection with users (see Kouprie & Sleeswijk Visser, 2009; and Smeenk, Sturm & Eggen, 2019a).

Designers acknowledge that there are limits to their understanding of users. Different demographics, social experiences, and cultural backgrounds put a limit to designers' empathic horizon (Costanza-Chock, 2020; McDonagh et al., 2018). With the acknowledgment of an empathic horizon, designers are also aware of ethical implications of viewing empathy as a pragmatic tool. Some designers advocate for not considering users as "simply the economic bottom line of designers" (Walther et al. 2017, p. 132), and highlighting that empathizing with users demand a set of technical skills but also ethical considerations (Devecchi & Guerrini 2017; Walther et al. 2017; Hess et al. 2021).

Empathy is central for designers and there is growing concern for clarifying empathy and pushing forward its theoretical and scientific development (Ball & Christensen, 2019; Cash, 2018; Surma-aho & Hölttä-Otto, 2022). Yet, it remains unclear how empathy leads to successful outcomes. One major challenge to answer this question is raised by the highly context-dependent nature of design. Designers work on real-life scenarios, where innumerable variables influence their interaction with users. Additionally, design cases can be drastically different one to another depending on the group of users and contexts designers happen to be exposed to. Taking this naturalistic context into account, how can scientists quantify the empathy between designers and users and further test its alleged casual properties in design outcomes?

The context of music performance perception

The interaction between a listener and a performer may be influenced by three different elements: multimodal perception, the listener's musical background, and the musician's emotional state (Kwan, 2016). Multimodal perception studies in music focus on how different sensorial information (e.g., audio-only, video-only, and audiovisual stimuli) affects the listener's perception of musical qualities. Studies suggest that audiovisual stimuli (and not audio-only or video-only) provide the most information to listeners about different musical qualities such as expressivity (Davidson, 1993; Vuoskoski et al., 2014), coordination (Chang et al., 2017), phrasing, intensity, fluency, professionalism (Nusseck & Wanderley, 2009; Waddell & Williamon, 2017), etc. Critically, listeners rely more on visual information to judge the quality of a performance, even if they consider sound quality as the most determining factor (Tsay, 2013, 2014). Furthermore, observing the performer's ancillary gestures (i.e., movements secondary to the direct production of sound) may convey expressive cues and intentions to listeners (Davidson, 2007; Thompson & Luck, 2012; Van Zijl et al., 2014; Wanderley et al., 2005).

A listener's perception of a performance is also influenced in different ways by their musical background. For instance, listeners with a musical background can detect differences in expressiveness in both audiovisual and audio-only stimuli, but listeners without a musical background can only detect those on the audiovisual stimuli (Davidson, 2005; Huang & Krumhansl, 2011). Even among listeners with a musical background, differences exist. For example, musicians can detect subtle changes if they are listening to an instrument, they themselves play vs. other instruments (Wapnick et al., 2004; Broughton & Davidson, 2014). Other variables affecting the way music is perceived are familiarity with the music (Flôres & Ginsburgh, 1996), and enjoyment of the piece being heard (Thompson, 2006).

The emotional state of a performer influences their musical output (Dahl & Friberg, 2007; Vines et al., 2011; Van Zijl et al., 2014). Music performance anxiety (MPA) is a crucial emotion that may have considerable effects on musical outputs. In this thesis, the experience and perception of MPA is investigated. MPA has been defined by Kenny (2010) as

The experience of marked and persistent anxious apprehension related to musical performance that has arisen through specific anxiety-conditioning experiences. It is manifested through combinations of affective, cognitive, somatic and behavioural symptoms and may occur in a range of performance settings, but is usually more severe in settings involving high ego investment and evaluative threat. It may be focal (i.e. focused only on music performance), or occur comorbidly with other anxiety disorders, in particular social phobia. It affects musicians across the lifespan and is at least partially independent of years of training, practice, and level of musical accomplishment. It may or may not impair the quality of the musical performance. (p. 433).

Between 16.5 to 60% of musicians surveyed across multiple studies have reported experiencing MPA (Fernholz et al., 2019). To cope with the unpleasant and sometimes debilitating effects of MPA, musicians may incur in consuming drugs (Taylor & Wasley, 2004; West, 2004; Brugués, 2011 a,b; Hernández et al., 2018) or abandoning their profession altogether (Hernández et al., 2018; Fernholz et al., 2019). Experimental studies in MPA suggest that performing in front of audiences vs. no-audiences causes increases in performer's psychophysiological responses and self-reported anxiety (Brotons, 1994; Cox & Kenardy, 1993; Studer, et al., 2014; Yoshie et al., 2008; Yoshie et al., 2009). Some evidence suggests that performers with mid- and high-anxiety had their ratings of performance quality and expressiveness impaired (Kwan, 2016).

Music performance research has been mostly dichotomous (but see Schober & Spiro, 2014; 2016): on the one side, studying the variables influencing the listener's perception of musical outputs; on the other side, the effects of MPA on the performer's musical output. Yet, a music performance is a social interaction comprised of the musician-listener dyad, and so connecting both social actors seem like a logical next step. IA can provide a frame to investigate the interaction between musicians and listeners. In particular, how the listener's perception differs from a musician's experiences while performing. Exploring this question can shed lights on the contextual influences on complex forms of communication such as musical performances.

Defining the golden standards

In empathy in design, the dyad is composed by designers and users. In music performances, the dyad is composed by musicians and listeners. What criterium will allow evaluating the accuracy of inferences? The criterion which allows assessing the accuracy of inferences between a dyad is called golden standard (Hall et al., 2016). Golden standards are operationalization-dependent, that is, researchers will specify how the accuracy of inferences are measured in a particular study. For that reason, it is the norm that golden standards vary and are dependent of a study's objectives.

Two instantiations of IA studies, each with their own operationalization of golden standards, are empathic accuracy and emotional recognition accuracy (for a fuller account of IA paradigms, see Hall et al., 2016). Empathic accuracy is better suited for verbal interactions, as is often the case in design; whereas

emotional recognition accuracy is better suited for non-verbal interactions, as is often the case in music performance. Next, we will elaborate what these instantiations are in greater detail.

1.1 Empathic accuracy

Empathic accuracy (EA) is a set of paradigms to quantify the similarity between reported experiences of a target(s) and a perceiver(s) who infers those experiences (Ickes 2001; Levenson & Gottman, 1983; Marangoni, Garcia, Ickes, & Teng, 1995). EA has been applied in various contexts such as psychotherapy interactions (Marangoni et al., 1995), intercultural interactions (Soto & Levenson, 2009), intimate partner aggression (Cohen et al., 2015), and designer-user interactions (Study II). It differs from self-report empathy questionnaires in that it allows for in-context and almost real-time measures of shared understanding. There are three main types of EA paradigms: the unstructured dyadic interaction paradigm, the standard stimulus paradigm, and the shared physiology paradigm (Study II).

1.1.1. The unstructured dyadic interaction paradigm

In its original form (Ickes, Bissonette, Garcia, & Stinson, 1990), two participants (who may or may not be acquaintances) are brought to a waiting room by a researcher who then excuses themselves saying they must retrieve something missing for the experiment. The participants are recorded interacting spontaneously for some time. Once the researcher returns, they disclose to the participants that they were video-recorded and ask for their consent to use this material for the next stage of the study. Each participant adopts both roles as targets and perceivers.

In the thought/feeling recording phase (Ickes, 2016), participants re-watch separately their videoed interaction. They are instructed to stop the recording every time they remember having a specific thought or feeling. Making use of a standard response sheet, they write down the specific time on the recording where they remembered the thought or feeling, the content itself, and the emotional valence associated with it (positive, neutral, or negative).

In the thought/feeling inference phase (Ickes, 2016), participants once again re-watch the interaction. This time, the researcher pauses the recording at every time point specified by the other pair. Participants are instructed to infer what their pair was thinking or feeling at the specific times the researcher pauses the recording.

A group of independent judges rate the similarity of content between the remembered and inferred thoughts and feelings. Contents are rated using a scale ranging from 0-2, where 0 = “essentially different content”, 1 = “somehow similar, but not the same, content”, and 2 = “essentially the same content” (Ickes, 2001, p. 230). The obtained scores are then transformed into a percentage scale which allows an easier interpretation of each participant’s EA (0 = “none of the possible accuracy points”, 100 = “all of the possible accuracy points”, Ickes, 2016, p. 54).

1.1.2. The standard stimulus paradigm

In the standard stimulus paradigm (Marangoni et al., 1995), the recorded interaction between a dyad is presented to a set of participants. In the study by Marangoni et al. (1995), perceivers observed three recorded psychotherapy sessions featuring three different clients (targets) in session with the same therapist. Each target reported their thoughts and feelings, acting as the golden standards (Hall et al., 2016) to which the inferred thoughts and feelings of the perceivers were compared to. The standard stimulus paradigm has several advantages (Ickes, 2001). Its content is flexible: the dyadic interaction could be set in different contexts, depending on the study's field and purpose. It allows measuring and comparing the EA of numerous participants.

1.1.3. Shared physiology paradigm

This paradigm focuses on the continuous emotional inference as well as the associated physiological responses of the perceivers in relation to targets (Levenson & Gottman 1983; Levenson & Ruef, 1992). For Levenson and Ruef (1992) the key measurement to test the empathic proximity between two persons is their physiological linkage, that is, their shared emotional and physiological reactions: "empathy is associated with a state of shared physiology, such that when one person was most empathically (i.e., accurately) perceiving the feelings of another, the two would most likely be in a common physiological state" (p. 236). Levenson and Ruef (1992) found that accurate ratings of perceivers with targets, either for positive or negative emotions, were associated with differentiated physiological responses. When accurately rating positive emotions, low cardiovascular arousal was detected in the perceiver. The accurate rating of negative emotions was linked to a shared autonomic response between rater and target. The researchers concluded that physiological responses can be used as a measurement of empathic accuracy. Relatedly, more recent studies have incorporated brain function as an additional physiological indicator of accuracy (Zaki et al., 2009).

1.2 Emotion recognition accuracy

Emotion recognition accuracy (ERA) is the ability to correctly judge the non-verbal emotional expressions of a target (Bänziger, 2016). ERA is assumed as a necessary condition for satisfying social interactions. In fact, in certain psychopathologies characterized by the impairment in social interactions, deficits in processing negative valence emotions have been found (Griffiths & Ashwin, 2016). For example, people with schizophrenia have shown less accuracy when recognizing particularly negative valence emotions (Leppänen et al., 2008), such as fear, and disgust, but also neutral expressions (Kohler et al., 2003). The accuracy bias seems to be different in other psychopathologies. For instance, people with social anxiety disorder are more efficient than control or depressed groups in identifying facial expressions of anger (Joormann & Gotlib, 2006), while misattributing emotions of anger towards neutral expressions (Bell et al., 2011). Relatedly, people with depression are more accurate when identifying expressions of sadness (Gollan, McCloskey, Hoxha, & Coccato, 2010).

In a typical ERA study, a perceiver is presented with nonverbal expressions of a target in the form of photographs or video recordings of facial expressions, postures, gestures or audio recording for vocal expressions. Perceivers are asked to label those stimuli. Along with the stimuli, predetermined response categories to be chosen are often presented. The truth/accuracy criterion is related to the expressive intention, to the self-reported emotion of the target or to the theoretical expectations of the researchers. The most common assessment of ERA is discriminating basic emotions in prototypical facial expressions (Bänziger, 2016). Relatedly, a variety of standardized tests exist for measuring ERA and allowing comparisons between studies (for a review, see Bänziger, 2016). Our capacities to infer emotional states is not limited to other people's verbal or non-verbal expressions of emotions, though. Emotions can also be accurately detected through music. For instance, listeners have been able to accurately perceive intended emotions in music composed to communicate emotions such as joy, sorrow, dullness, anger, and peacefulness (Thompson & Robitaille, 1992).

2. Objectives

In this thesis we address the following questions:

1. How to implement an IA approach to quantify the understanding between dyads interacting in design and music?
2. How accurate are perceivers when inferring the mental experiences of targets?
3. What contextual factors affect IA?

The overall goal of this dissertation is to explore contextual factors which influence IA. These contextual factors include, for example, the content of interactions, multimodal perception, and the background of the perceivers. To investigate them, we design controlled conditions to approach naturalistic interactions in design and music performance. We chose these human activities because both have particularities which make them suited to explore the different contextual factors influencing IA.

A pre-condition to achieve our goals is to obtain conceptual clarity. As mentioned in the Empathy in Design section, designers use the word “empathy” to describe multiple phenomena. This wide scope of the word may difficult singling out the function of empathy in obtaining better design outcomes. Thus, Study I aimed at two things: positioning empathy in design as an umbrella construct by comparing the literature between design and psychology; and arguing for clearer definitions and operationalization of empathy in design. The emphasis on being careful on how we define and operationalize concepts in design is also applicable to music.

Having mapped the conceptual panorama in front of us, we addressed the first objective of how to create a controlled space where we can recreate a dyadic interaction and quantify it in a meaningful way that is as close as possible to real-life scenarios, but that allows us to extract relevant conclusions. This methodological challenge is transversal to Studies II-IV.

Study II implemented EA in a design case. The implementation began by identifying a real design challenge in a specific group of musicians. A designer interviewed relevant users with the aim of gaining further insight into their experiences with musical equipment. We adapted EA to suit this interaction and quantify the designers’ accuracy when inferring the thoughts and feelings of the musicians. Furthermore, we proxied a post-interview stage where designers synthesized the user’s needs, with the aim of bridging designers’ EA with the effectiveness of their design outcomes.

Study III further tested the flexibility of EA by applying it to three different design cases. Our experience with Study II taught us that the users' mental contents may not be all relevant for the purposes of design. EA originally categorizes mental contents as thoughts or feelings. In Study III we proposed a different categorization better tuned for the design context, by classifying mental contents as design-related mental contents from others. We further tested the accuracy of designers when inferring design-related contents vs. non-design-related ones.

Study IV recreated a music performance context where a pianist performed for a group of listeners. We made use of ERA instead of EA, as the former is more suitable for non-verbal interactions such as the ones occurring in music performance contexts. We measured the pianist's experienced anxiety while performing and compare it to the listener's perceived anxiety while listening to the same performances. This non-verbal musical interaction demanded considering other variables that may inform in relevant ways the performer and his experiences and the listeners and their perceptions. Thus, we considered the stimulus modality (audiovisual vs. audio-only), the performance context (rehearsal vs. recital), and the piece performed (slow vs. fast).

The epistemological framework of this thesis best fits in the world hypothesis of mechanism. According to Pepper (1942), the hypotheses we have about the world are "derived from certain masses of empirical evidence, originating in common sense, which become cognitively refined and may be codified into sets of categories that hang together" (p. 328). One such world view is mechanism. In mechanism, the root metaphor of knowledge (the common sense from which masses of empirical evidence are derived) is a machine. A mechanist would investigate the components of the machine which are supposedly related to others in a systematic way. Through the hypothetico-deductive methodology, the mechanist will test the adequacy of their model to reality (Hayes, Hayes, & Reese, 1988).

This thesis best fits within the mechanism world hypothesis because it focuses on two things. First, we present empathy as a complex conceptual "machine" and we make the claim that, to test its causal effect in design, we need to identify and delimit its components. Second, we explore the causal effects of some contextual variables on the individual's interpersonal accuracy.

3. Summary of studies

3.1 Study I – Under the umbrella: Components of empathy in psychology and design

Aims of the study

Empathy is a central concept in design. Designers emphasize its relevance in design outcomes and, so, much research efforts have been directed into understanding what it is and how it can be controlled. Yet, in this quest for understanding, the concept became too broad and unspecific. The lack of clarity threatens transforming the alleged causal properties of empathy untestable. The aim of this review is to gain perspective as to what does empathy in design mean, to single out more precise ways of investigating it. To do so, we review its definition in design and psychology, given the influence the latter has had on the former.

Methods

This review paper started in 2018 as a background literature search. Initial sources were reviews of empathy in design and psychology (e.g., Clarke, De-Nora, & Vuoskoski, 2015; Kouprie & Sleeswijk Visser, 2009; Mattelmäki, Vaajakallio, & Koskinen, 2014; Shamay-Tsoory, 2011; Zaki & Ochsner, 2012), as well as handbook chapters (e.g., Batson, 2009; Köppen & Meinel, 2015; Neumann & Westbury, 2011; Shamay-Tsoory, 2009). It has been expanded since then through a snowball procedure, by revising the sources cited by authors of the consulted texts. Additionally, the content was revised and corrected following the commentaries given by journal reviewers. Sources were found using search engines such as Google Scholar and looking directly into design and psychology journal websites. All sources included were in English language and restricted to the fields of psychology and design.

Results

In design, there is a shared understanding that empathy is key for successful businesses and projects. Accordingly, it has become central in design research and practice over the last two decades. Yet, despite being such an important concept, it is paradoxical that there is no consensual definition of what empathy means. We believe the context-dependent nature of design practice explains the multiple ways in which empathy is understood: different design challenges require different approaches to addressing them.

This has resulted in empathy being an umbrella construct or “broad concept or idea used loosely to encompass and account for a set of diverse phenomena” (Hirsch & Levin, 1999, p. 200). Umbrella constructs are useful to hold a research field together, especially during its initial phases of development.

The broaden scope of research on empathy in design has resulted in a series of implications that we categorized under four groups. The first one includes practical techniques meant to increase the designer’s understanding of the user’s circumstances and experiences. These tools can be distinguished by their level of “embodiment”. For instance, wearable simulations (e.g., sight-depriving glasses, movement-restriction suits, etc.) involve the designer’s physical sensations; whereas an empathy map requires an intellectual abstraction and classification of the verbalized or deduced user’s experiences into a visualization tool (see Table 1 for more examples).

Table 1. Examples of practical techniques for empathizing

Less embodied tools		More embodied tools	
Personas	“Abstractions of groups of real consumers who share common characteristics and needs” (Miaskiewicz & Kozar 2011, p. 418)	Wearable simulations	Equipment limiting body movements of the wearer to evoke the experience of completing some physical tasks under restricted conditions, thus inspiring designers to “engage more closely with user experiences” (Kullman 2016, p. 83).
Storytelling	“Mental story construction” (Madsen & Nielsen 2010, p. 59) used to interpret and communicate user data (Kankainen et al. 2012).	The Empathic Experience Design	A concept generation tool based on engaging designers in empathic experiences, which simulate disabilities or situational disabilities (Johnson et al. 2014; Raviselvam, Hölttä-Otto & Wood, 2016)
Empathy map	A visual tool containing four quadrants (say, do, think, and feel) utilised to summarise and synthesise key need-finding insights from users (Both & Baggeor n.d.; Study II).	Experience simulators	Facilitating the transfer of user insights empathically to design team members who were not in contact with users (Smeenk et al. 2017, 2019b).
		Role-playing	Often used to visualise a product from the user’s point of view (Medler & Magerko, 2010)
		Spatial augmented reality	Implemented as a way of understanding abilities, problems, insights, and strengths of, for instance, students with ASD, with the goal of providing practical tools for teaching and engaging ways for students to learn (Takahashi et al., 2018).

A second one is empathy as a process of interaction and reflection. This are step-by-step guidance on how to facilitate the understanding of a user's experiences and context. Two notable examples are frameworks by Kouprie & Sleeswijk Visser (2009) and Smeenk, Sturm & Eggen (2019a). The former is based on "the principle that a designer steps into the life of the user, wanders around for a while and then steps out of the life of the user with a deeper understanding of this user" (pp. 444–445). The latter "concerns the understanding of the formative process of becoming an empathic design professional who knows which attitude, skills and knowledge are applicable in an empathic design process" (p. 61) (Table 2).

Table 2. Examples of empathy-generating frameworks

Kouprie & Sleeswijk Visser (2009)	Smeenk, Sturm & Eggen (2019a)
1. Discovery	Is a design activity (a) more focused on self or other?
2. Immersion	Is a design activity (b) more affective or more cognitive?
3. Connection	Is a design activity (c) taken with a more participatory or expert mindset?
4. Detachment	Is a design activity (d) more design- or research-led?

A third one is limitations and user-designer empathy, acknowledging that there will always be limitations as to how well a designer can understand a user, regardless of the tools used to decrease the knowledge gaps. Lastly, cultivating ethical thinking through empathy or the ethical discussion around the pragmatic implications of empathy in design: is "empathy" the best term when considering utilitarian motivations of designers?

After mapping empathy in design, we rhetorically invite the reader to define it considering this conceptual discussion. We aim to highlight how complicated this is. We take this difficulty as a manifestation of empathy being an umbrella concept, a "broad concept or idea used loosely to encompass and account for a set of diverse phenomena" (Hirsch & Levin, 1999, p. 200). As the field develops and the possible meanings of empathy gain scope, causal claims like "empathy is key for successful business and projects" loose precision. The indetermined limits of an umbrella construct are opposed to the specificity needed for testing empirical statements.

To further gain perspective on the empathy maze, we bring in how empathy is understood in psychology. The reason being that design has borrowed much of its understanding of empathy from psychological literature but has missed precision in doing so. For instance, definitions of empathy in design tend to include the distinction between cognitive and affective empathy, but there are so many more mechanisms under the empathy umbrella in psychology which remain unexplored in design. Thus, we chose the definition of empathy by Cuff et al. (2016) for being the most recent and inclusive of previous definitions, and because it has proven a useful conceptual framework in other fields (Clark, Robertson & Young 2018). Thus, empathy is

An emotional response (affective), dependent upon the interaction between trait capacities and state influences. Empathic processes are automatically elicited but are also shaped by top-down control processes. The resulting emotion is similar to one's perception (directly experienced or imagined) and understanding (cognitive empathy) of the stimulus emotion, with recognition that the source of the emotion is not one's own. (Cuff et al., 2016, p. 150).

Lastly, we juxtaposed the conceptual landscapes of empathy in design and empathy in psychology into a single visual representation (Figure 1). The comparison reveals a complex and tangled scenario, where there is not a one-to-one correspondence between empathy in psychology and empathy in design. Every component of empathy in psychology is related to several understandings of empathy in design. Figure 1 illustrates why the causality properties of empathy cannot be tested the way it is currently understood in design. We hope future design researchers may be more aware of the empathy landscape and instead of stating they will investigate “empathy”, they carefully delimit *which* component of empathy are they interested investigating.

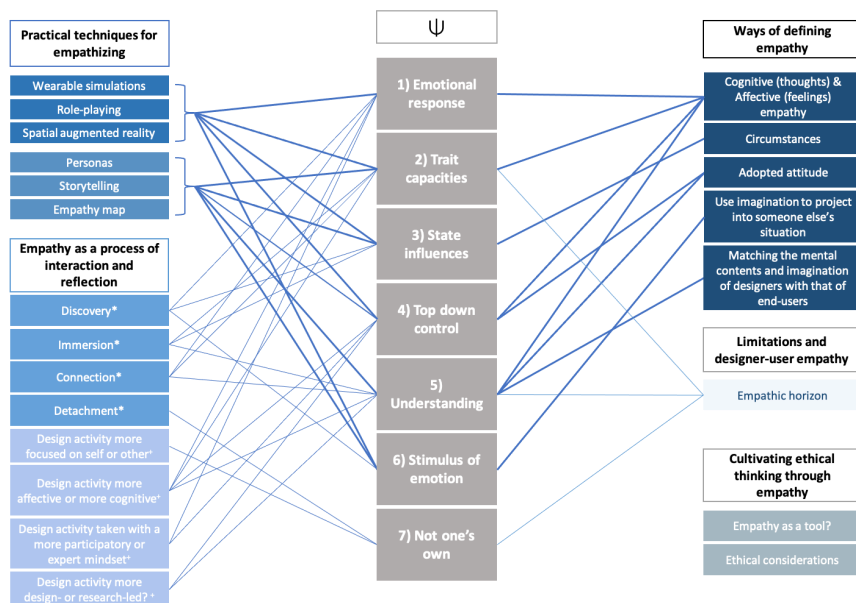


Figure 1. Overlaps of the components of the concept “empathy” in psychology (Ψ) and in design (on left and right sides of the figure). Under “practical techniques for empathizing” we grouped the six elements under two groups. The first one (darker blue) are techniques which have a higher embodied nature and are close to affective empathy. The second one (lighter blue) are less embodied techniques and closer related to top down control and understanding (cognitive empathy). Regarding “cultivating ethical thinking through empathy”, this implication is closely related to the behavioral outcomes of empathizing. Thus, we did not establish connections to Cuff et al. (2016) because they declined to incorporate behavioral outcomes into their definition: “Although it is to be noted that empathy may lead to behavioural outcomes, this definition of empathy purposefully avoids behavioural implications” (p. 150). Yet, we think that “cultivating ethical thinking through empathy” could be explored from other psychological perspectives on empathy (e.g., see Davis, 2006). *Kouprie and Sleeswijk Visser (2009). +Smeenk et al. (2019b) (from Chang-Arana, Á. M., Surma-aho, A., Hölttä-Otto, K., & Sams, M. (2022). Under the umbrella: Components of empathy in psychology and design. *Design Science*, 8, E20, p. 15. Copyright The Authors, 2022. Published by Cambridge University Press).

Conclusions

Despite empathy being a central concept in design, it is not clear what does it mean, let alone test the alleged causal properties of empathy in the success of design outcomes. We intended to uncover the multiple mechanisms within the umbrella construct, so designers interested in empathy can be aware of the complex landscape ahead. We expect that this review informs a more careful selection of the meaning designers will give to empathy in their work and research. Study I may inform future researchers of different approaches to study empathy. In our case, it allowed us to select the road that seemed most relevant to test its alleged casual properties: investigate the IA between designers and users. Study II and III are reports of our chosen path.

3.2 Study II – Empathic accuracy in design: Exploring design outcomes through empathic performance and physiology

Aims of the study

The main challenge of Study II was to find a method to quantify the IA between designers and users. Design literature on empathy is predominantly qualitative (except for studies such as Alzayed et al., 2021; Hess et al., 2017; Rasoal, Danielsson, & Jungert, 2012; and Surma-aho et al., 2018), since qualitative methods allow more flexibility to adapt to the natural conditions of a design context. Study II is a quantitative alternative to studying empathy while conserving some of the contextual nature of design. Furthermore, we explore whether our approach allows testing the role of IA on design outcomes. We addressed the following questions:

1. How to measure the interaction between a designer-user dyad exploring a design challenge?
2. How accurately can the designers understand the group of musicians?
3. Does the designers' accuracy in regard to the musicians' mental contents and emotions positively correlate with design outcomes?
4. Does the similarity of the facial emotional expressions of the designers and musicians correlate with the designer's empathic accuracy?

Methods

Participants

Five professional music students were recruited (mean age = 23.60, $SD = 1.52$, mean playing-time experience = 15 years, $SD = 1.41$). Their instrumental specialism was clarinet, saxophone, or oboe. We recruited one designer whose task was to interview all musicians (see Appendix A for interview model). Designer 1 had 13 years of experience in design which included study, industry, and research activities. A second designer with 5.5 years of experience in design was recruited to watch the interviews and perform the same EA tasks Designer 1 performed.

Design brief

Designer 1 conducted a user-centered interview with five woodwind musicians on their experiences with reeds. These wood or plastic artifacts allow the production of sound by vibrating due to air pressure. They can affect the tonal quality, expressive, and technical range of a musician (Thompson 1979; Ledet 1981; Almeida et al. 2013). Reeds are a recurrent time and monetary cost for most woodwind players, who need to replace them by buying or manufacturing them from scratch (Ledet 1981). Additionally, Designer 1 was allowed to explore other artifacts, such as instrumental transporting and cleaning equipment.

Tasks and procedures

Interviews lasted for approximately 30 minutes. During this time, we recorded electrocardiogram (ECG), facial electromyography (EMG) and galvanic skin responses (GSR) from both designers and the musicians. In this study we focused solely on facial EMG of the corrugator supercilii (eyebrow muscles) and zygomaticus major (cheek muscles). After each interview, musicians logged-in their mental contents recollected from the interview following Ickes' protocol (2001). Musicians paused the recorded interview every time they remembered having a thought or feeling. They were instructed to register their mental contents together with the exact timing when they occurred as well as whether it had a positive, neutral or negative valence, using Ickes' standard response sheet (see Table 3 for examples of remembered mental contents and Appendixes B and D in Study II for the instructions and response sheets used).

Table 3. Examples of remembered mental contents

Time	Thought or Feeling	+, 0, -
0.04	I was curious about what the interview was going to be about.	+
15.44	I was realizing I didn't demonstrate assembling at all.	0
29.09	I was feeling confident about my English.	+

Note: (Adapted from Chang-Arana, Á, M., Piispanen, M., Himberg, T., Surma-aho, A., Alho, J., Sams, M., & Hölttä-Otto, K. (2020). Empathic accuracy in design: Exploring design outcomes through empathic performance and physiology. *Design Science*, 6, E16, p. 29. Copyright The Authors, 2020. Published by Cambridge University Press).

The designers watched the video interviews musicians used to register their mental contents. The designers were instructed to infer as accurately as possible what each musician was thinking or feeling at every time pause using Ickes' standard inference sheet (see Table 4 for examples of inferences and Appendixes C and E in Study II for the instructions and response sheets used).

Table 4. Examples of inferred mental contents

Time	Thought or Feeling	+, 0, -
0.04	She was feeling slightly nervous about the interview.	0
15.44	She was feeling uncertain about what to show and explain.	0
29.09	She was feeling entertained knowing she has it easier with reeds than oboists.	+

Note: (Adapted from Chang-Arana, Á, M., Piispanen, M., Himberg, T., Surma-aho, A., Alho, J., Sams, M., & Hölttä-Otto, K. (2020). Empathic accuracy in design: Exploring design outcomes through empathic performance and physiology. *Design Science*, 6, E16, p. 29. Copyright The Authors, 2020. Published by Cambridge University Press).

The similarity of content between the remembered mental contents and the inferred mental contents were assessed by a group of 14 native speakers of English (eight for Designer 1 and six for Designer 2). Raters followed Ickes' protocol (2001), rating the similarity of each pair of entries using a three-point Likert

scale, where 0 = “essentially different content”, 1 = “somehow similar, but not the same content”, and 2 = “essentially the same content” (see Table 5 for an example of three entries and Appendix F in Study II for the instructions used).

Table 5. Example of three entries assessed by an external rater

Time	Actual thoughts or feelings	Inferred thoughts or feelings	How similar are they? (Max = 2, Min = 0)
0.04	I was curious about what the interview was going to be about.	She was feeling slightly nervous about the interview.	2
15.44	I was realizing I didn't demonstrate assembling at all.	She was feeling uncertain about what to show and explain.	1
29.09	I was feeling confident about my English.	She was feeling entertained knowing she has it easier with reeds than oboists.	0

Note: (Adapted from Chang-Arana, Á. M., Piispanen, M., Himberg, T., Surma-aho, A., Alho, J., Sams, M., & Hölttä-Otto, K. (2020). Empathic accuracy in design: Exploring design outcomes through empathic performance and physiology. *Design Science*, 6, E16, p. 29. Copyright The Authors, 2020. Published by Cambridge University Press).

To test for the reliability of the raters' scores, each rater was treated as a questionnaire item and every entry as a questionnaire response (Ickes, 2001). Cronbach's alpha was then calculated for each interview. Scores given by the raters were used to calculate an empathic accuracy score for the designers in the form of a percentage scale, following the procedure explained by Ickes (2001). The reliability values obtained were above Nunnally's (1967) .70 criterion (Table 6).

Table 6. The inter-rater reliability of the assessment of the similarity of content

	Designer 1		Designer 2	
	Cronbach's α	SEM	Cronbach's α	SEM
User 1	.90	0.09	.88	0.09
User 2	.90	0.14	.92	0.15
User 3	.88	0.15	.87	0.17
User 4	.86	0.14	.75	0.12
User 5	.90	0.12	.87	0.12

Note: User 1, entries = 45; User 2, entries = 18; User 3, entries = 15; User 4, entries = 17; User 5, entries = 22. Designer 1 was rated by eight external raters; Designer 2 by six external raters. SEM = Standard Error of Measurement. (From Chang-Arana, Á. M., Piispanen, M., Himberg, T., Surma-aho, A., Alho, J., Sams, M., & Hölttä-Otto, K. (2020). Empathic accuracy in design: Exploring design outcomes through empathic performance and physiology. *Design Science*, 6, E16, p. 12. Copyright The Authors, 2020. Published by Cambridge University Press).

To connect the designers' empathic accuracy with design outcomes, designers completed a modified empathy map (Both & Baggereor, n.d.) and generated ideas for new or improved accessories for musicians (see Appendix G for example of instructions). Insights from both tasks were sent back to the musicians for rating how close were the designers' insights to their own needs. This simulated the step of coming back to the user for obtaining feedback of ideas. See Figure 2 for a summary of the study's procedure.

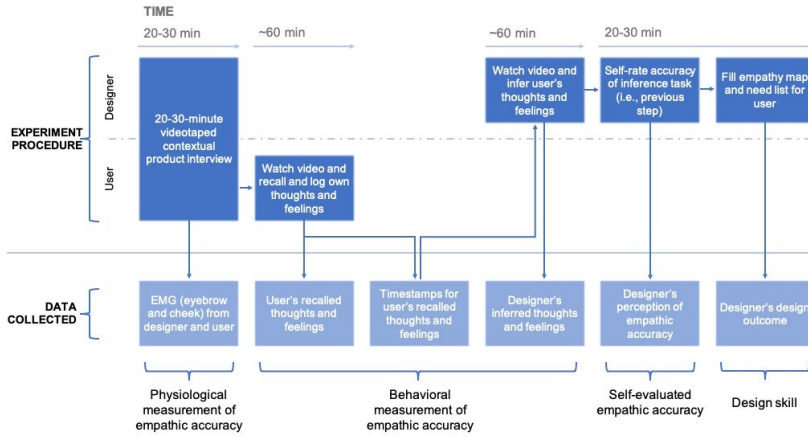


Figure 2. An overview of the study procedure (Adapted from Chang-Arana, Á, M., Piispanen, M., Himberg, T., Surma-aho, A., Alho, J., Sams, M., & Hölttä-Otto, K. (2020). Empathic accuracy in design: Exploring design outcomes through empathic performance and physiology. *Design Science*, 6, E16, p. 7. Copyright The Authors, 2020. Published by Cambridge University Press).

Results

Designers obtained approximately half of all the possible accuracy points (Table 7). Interestingly, Designer 1 reported a high self-rated accuracy on the task. It is important to highlight, though, that Designer 1 understood the self-rating question as how well he followed the instruction of inferring accurately what each musician was thinking or feeling at every time pause. Designer 2 had more accuracy variation on identifying the user's emotional valence than Designer 1. The inter-rater reliability of the external raters was above Nunnally's criterion of .70 (1967).

Table 7. The overall designers' empathic accuracy scores.

	Designer 1			Designer 2	
	Aggregated index of empathic accuracy (%)	Designer's reported self-rated accuracy (%)	Correct identification of user's emotional valence (%)	Aggregated index of empathic accuracy (%)	Correct identification of user's emotional valence (%)
User 1	45.42	90.00	42.22	42.22	40.00
User 2	50.35	80.00	55.56	55.09	50.00
User 3	48.75	60.00	40.00	49.44	20.00
User 4	44.49	80.00	41.18	55.88	35.29
User 5	45.17	80.00	50.00	53.41	40.91

Note: (From Chang-Arana, Á, M., Piispanen, M., Himberg, T., Surma-aho, A., Alho, J., Sams, M., & Hölttä-Otto, K. (2020). Empathic accuracy in design: Exploring design outcomes through empathic performance and physiology. *Design Science*, 6, E16, p. 13. Copyright The Authors, 2020. Published by Cambridge University Press).

In Table 8 we show examples of remembered and inferred mental contents from both designers. These correspond to different levels of empathic accuracy.

Table 8. Examples of high-, mid- and low-empathic accuracy.

	Musicians	Designers	Average EA
High-empathic accuracy	I was feeling unprofessional/ashamed because I do not use the 'proper' boxes to keep my reeds safe. (-)	D1: She was feeling ashamed about not taking as good care of the reeds as she knows she could. (-) D2: She was remembering someone struggle or complain about not being able to take their clarinet into pieces and having to carry around a big backpack for it. (0)	1.88
	I was feeling amused by the thought of carrying a giant case. (+)		1.83
Mid-empathic accuracy	I was thinking that I try to adapt to the reed, but I am not good enough yet. (-)	D1: She was feeling humble about not wanting to say she is good at adapting. (+)	1
	I was thinking that the neck strap prevents me to bow towards the audience in an elegant way. It could fall while bending forward. (0)	D2: She was recalling the relief of taking the shoulder strap off and being free of the sax's weight. (+)	1
Low-empathic accuracy	I was a bit frustrated as I started to struggle with English a bit in my head. (-)	D1: She was thinking back of how she got from hardly playing to playing more and more. (0) D2: She was thinking about the bad performance she had last week and how the reed felt then. (0)	0
	I was feeling happy, I like the word 'wizardry'. (+)		0

Note: D1 = Designer 1; D2 = Designer 2 (From Chang-Arana, Á, M., Piispanen, M., Himberg, T., Surma-aho, A., Alho, J., Sams, M., & Hölttä-Otto, K. (2020). Empathic accuracy in design: Exploring design outcomes through empathic performance and physiology. *Design Science*, 6, E16, p. 14. Copyright The Authors, 2020. Published by Cambridge University Press).

Musicians gave high scores to the designers on the empathy map and ideas for improvements (Table 9). The higher the score, the closer the insight was from the user's experience. Even though we correlated the aggregated index of empathic accuracy of each designer to their scores for the design tasks, the very low sample size makes these correlations meaningless to interpret, although indicative of what future studies could explore with a larger sample size.

Finally, the correlation between Designer 1 and the user's *zygomaticus major* muscle, and the EA scores reported across all interviews (117 events) was non-significant, $p = .51$, $r = -.06$. A similar non-significant result was obtained for Designer 2, $p = .76$, $r = -.03$. This means that the similarity in the emotional expression between the designers and musicians (smiling or the activity of the *zygomaticus major* muscle) had no relationship with the empathic accuracy of the designers on each inferred entry. Simply put, a similar activation of the designers and musicians' cheek muscles did not help designers infer correctly what the musicians were thinking at each time pause.

Table 9. The designers' performance in three design tasks.

	Designer 1			Designer 2		
	Empathy map		Ideas for improvements (%)	Empathy map		Ideas for improvements (%)
	Thoughts (%)	Feelings (%)		Thoughts (%)	Feelings (%)	
User 1	87.20	72.00	80.00	77.14	68.57	60.00
User 2	96.60	86.60	76.60	89.09	97.14	96.00
User 3	100.00	95.60	90.00	92.50	95.00	73.33
User 4	94.00	91.40	93.40	76.67	95.00	80.00
User 5	88.00	96.00	86.60	90.00	80.00	51.43

Note: (From Chang-Arana, Á. M., Piispanen, M., Himberg, T., Surma-aho, A., Alho, J., Sams, M., & Hölttä-Otto, K. (2020). Empathic accuracy in design: Exploring design outcomes through empathic performance and physiology. *Design Science*, 6, E16, p. 14. Copyright The Authors, 2020. Published by Cambridge University Press).

Conclusions

Study II is an adaptation of EA paradigms in design to measure the interaction between a designer-user dyad exploring a design challenge. In this way, design got added to a list of other areas of study where EA has proven useful. These include psychotherapy, relationship violence, brain research, spouse interaction, among others (Clements et al., 2007; Levenson & Gottman, 1983; Levenson & Ruef, 1992; Marangoni et al., 1995; Study III; Zaki et al., 2009).

This study was also an initial step towards testing whether empathy is necessary for successful design outcomes, by implementing EA paradigms in a design case. The two designers recruited for this study had a similar EA performance, independently of being directly or indirectly exposed to a group of users. Interestingly, their performance was considerably higher than previous EA studies that have measured the interpersonal accuracy of male strangers (EA = 24%) vs. male friends (EA = 36%) (Stinson & Ickes, 1992), and psychology students towards psychotherapy clients (23-24%) (Marangoni et al. 1995). Given our small sample size, it was not possible to draw meaningful conclusions about the influence of EA on design outcomes (i.e., empathy map and ideation in this study), yet provides an example of what future researchers could test at a larger scale. We failed at establishing a correlation between similarity of facial expressions, measured through EMG, of designers and musicians with the designers' EA on every inferred entry. This does not rule out the relevance of physiological data in EA research but suggests that it may not be relevant when combining it with the implementation of EA described in this study.

We faced challenges regarding the recording, preprocessing, and analysing our ECG, GSR, and EMG data. For instance, the software we used to collect GSR data was still under development, making it difficult to make sense of the data. In addition, the GSR metric devices presented some problems in data recording, missing bits of data. Since the analysis of physiological data in the context of design was new to us, we decided to focus solely on EMG, as it is often utilized

in emotional contagion research. Nonetheless, we encourage future researchers to consider these physiological metrics in future empathy studies should their research questions would demand so.

We hypothesized that the reason behind the high EA obtained by the designers, in comparison to previous studies, is the more structured context under which the dyad interacted. The designers and users were aware of the interview topic. It may have been a combination of mutual knowledge of the conversation topic, the demonstration of concrete objects, and the attentional resources of designers focused on design-relevant information that explains the high EA obtained by the designers. Thus, this combination of elements may have helped the designers correctly identify the mental contents of the musicians. Study III explores further this assumption by implementing EA in two additional design tasks and focusing on the designers' EA of design-related entries when compared to non-design-related entries.

3.3 Study III – Reading the user’s mind: Designers show high accuracy in inferring design-related thoughts and feelings

Aims of the study

Rarely EA studies report separately scores for thoughts and feelings (but see Hinnekens et al., 2018). Instead, they are reported as a combined score. We thought knowing the EA of designers for thoughts and feelings could be informative for different design decisions. For instance, negative feelings could inform about aspects to be improved in a design, while thoughts could inform of ways of improving it. Yet, evidence is limited on whether to expect differences in the EA for thoughts and feelings. Additionally, designers emphasize correctly identifying user’s needs in design education and practice (Hess et al., 2017; Kouprie & Sleeswijk Visser, 2009; Rasool et al., 2012; Surma-aho et al., 2018). To this end, we wondered if EA could allow quantifying design-related accuracy. This study implemented EA into three different design cases and raised two research questions and hypotheses:

1. How accurately can designers identify thoughts and feelings of users separately?
 - 1.1. There is no difference between how accurately designers infer thoughts and feelings of users.
2. How accurately can designers identify design related and non-design-related contents?
 - 2.1. Designers are more accurate in inferring design-related inferences than those that are not related to design.

Methods

Participants

Study III combines designers and users recruited from three sub-studies: musician sub-study, driver sub-study, and visual impairment sub-study. The musicians and designers of the musician sub-study were the same as the ones in Study II of this thesis, except for the exclusion of one musician due to technical problems with their recording. One driver and four designers were recruited for the driver sub-study. Lastly, one person with visual impairment and 18 design students were recruited for the visual impairment sub-study. Across sub-studies, there were 30 designer-level data points.

Design brief

Videoed semi-structured interviews were common to each sub-study. Interviews were meant to mimic the initial stages of a design process during which designers gather user information and identify potential future directions of a project. All interviewees could demonstrate their experiences with real objects during the interview (i.e., musical instruments and accessories, a car, braille reader and phone). For the musician sub-study, the design brief was to improve musicians’ experiences with reeds and other instrument accessories. For the driver sub-study, the design brief was to improve driving safety through car

accessory design. Finally, for the visual impairment sub-study, the design brief was to improve the experiences of people who have visual impairments.

After each interview, all six users logged in their mental contents recollected from the interview following Icke's protocol (2001), and as described in Study II. We coded all mental contents either as thoughts or feelings. Thoughts were defined as anything involving purposeful cognitive processing, and feelings as emotional reactions at any point in the interview. We estimated the inter-rater reliability using Cohen's kappa, obtaining a value of .91. Similarly, we coded all mental contents either as design-related or non-design-related. A design-related entry had to a) guide future research, b) inform the design of a solution, or c) help assess the truthfulness/relevance of gathered information. We estimated the inter-rater reliability using Cohen's kappa, obtaining a value of .81.

Inferred mental contents were obtained from the designers, following Ickes (2001) protocol, albeit with small variations within each sub-study. Given these differences in the inference stage as well as having different groups of users, we did not average designers' EA across sub-studies. The calculation of designers' EA followed Ickes (2001) protocol and as described in Study II. Following this procedure, we obtained five EA scores: (a) overall EA, and EA for (b) thoughts, (c) feelings, (d) design-related entries, and (e) non-design-related entries. The inter-rater reliability of the external raters was above Nunnally's criterion of .70 (1967), except for dyads Designer 1 – Musician 1 and Designer 2 – Musician 4.

Results

Designers' median EA for feelings ($Mdn = 46.67$) was significantly higher than for thoughts ($Mdn = 21.43$), $z = -3.94$, $p < .01$, $r = -.72$. We inspected the data grouping the designers according to the design case (Figure 3). The visual impairment sub-study showed the highest scores for feelings and had a higher sample than the other sub-studies, which could explain the higher median values of EA for feelings over thoughts. Thus, we take the significant difference between feelings and thoughts cautiously.

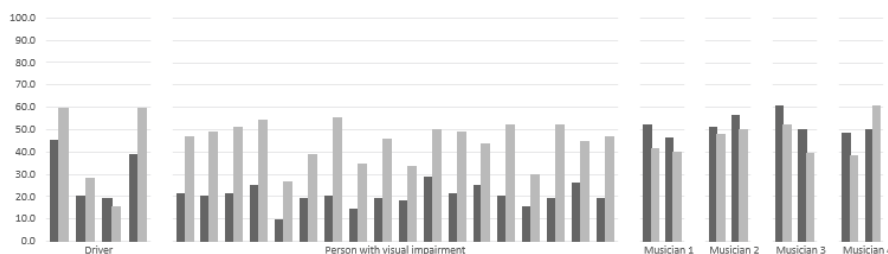


Figure 3. Designer-level accuracy for thoughts (dark grey) and feelings (light grey) (From Chang-Arana, Á. M., Surma-aho, A., Li, J., Yang, M. C., & Hölttä-Otto, K. (2020). Reading the users' mind: Designers show high accuracy in inferring design-related thoughts and feelings. *ASME 2020 IDETC/CIE Conference*, 17th-19th August. Paper No: DETC2020-22245, p. 13. Copyright, 2020. Published by the American Society of Mechanical Engineers).

Designers were more accurate when inferring design-related entries vs. non-design-related entries (Figure 4). Designers' median EA for design-related

entries ($Mdn = 56.11$) was higher than for non-design-related entries ($Mdn = 31.94$), $z = -4.60$, $p < .01$, $r = -.84$. When grouping the designers according to the interviews they watched, a clear trend is observed in favor of high design-related EA (Figure 5).

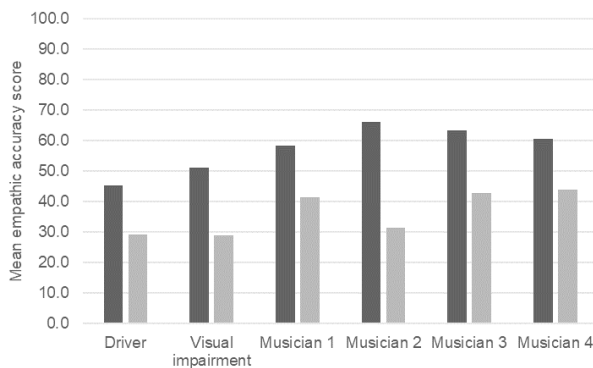


Figure 4. User-level empathic accuracy scores calculated from design-related (dark gray) and non-design-related (light gray) entries (From Chang-Arana, Á. M., Surma-aho, A., Li, J., Yang, M. C., & Hölttä-Otto, K. (2020). Reading the users' mind: Designers show high accuracy in inferring design-related thoughts and feelings. *ASME 2020 IDETC/CIE Conference*, 17th-19th August. Paper No: DETC2020-22245, p. 8. Copyright, 2020. Published by the American Society of Mechanical Engineers).

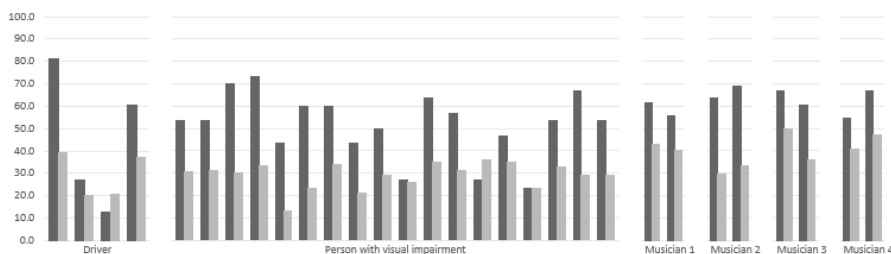


Figure 5. Designer-level accuracy for design-related (dark gray) and non-design-related (light gray) inferences (From Chang-Arana, Á. M., Surma-aho, A., Li, J., Yang, M. C., & Hölttä-Otto, K. (2020). Reading the users' mind: Designers show high accuracy in inferring design-related thoughts and feelings. *ASME 2020 IDETC/CIE Conference*, 17th-19th August. Paper No: DETC2020-22245, p. 13. Copyright, 2020. Published by the American Society of Mechanical Engineers).

Conclusions

Regarding our first hypothesis, there was no salient difference in designers' EA for thoughts and feelings. The higher median scores observed for feelings could have been carried by the sample size of the visual impairment sub-study. In that sub-study, an initial exploration of the user's experiences as a visually impaired person was the main focus, without a clear product or service in mind. Contrarily, the musicians' and driver's sub-studies were more focused on already existing products that could be referred to. Thus, it may be that designers' abilities to accurately detect thoughts or feelings change according to the design stage. Regarding our second hypothesis, designers had higher EA scores for design-related entries and for non-design-related entries across cases. We attributed

this to the design context. Designers were focused on identifying potential opportunities for future projects. Thus, their attentional resources are tuned to designer-related content the users may provide. Lastly, we tested the versatility of EA as a method for measuring empathic performance on three very different design cases. We also showed how EA can be adapted to fit demands from specific fields, by creating the categories of design-related and non-design-related mental contents.

3.4 Study IV – Exploring the interpersonal level of music performance anxiety: Online listener’s accuracy in detecting performer anxiety

Aims of the study

Like Studies II-III, the main challenge of Study IV was to find a method to quantify the IA between a dyad, in this case musician-listener. Contrarily to Studies II-III, though, these interactions were non-verbal. By removing verbal communication, other variables affecting the IA of a dyad should be carefully considered. Studying IA in non-verbal interactions also meant choosing ERA over EA as a better suited method for these situations. We explored the use of ERA in the context of MPA, due to the (understandably) highly intrapersonal focus of the field. A common symptom of MPA is fear of negative evaluation. Yet, it is often treated as a symptom to be treated rather than an empirical statement to be tested. Rarely experimental studies in MPA deepen into variables affecting the listener’s perception, and to the best of our knowledge, do not compare it to the musician’s actual MPA experience. The lack of feedback from the audience and its comparison to the musician’s own experiences, deprive the latter from information to evaluate more objectively the outcomes of their own performance. The first aim was to determine how listeners of different musical backgrounds perceive MPA, while considering different characteristics of the stimuli to be rated. The second was to evaluate the ERA of the listeners of different musical backgrounds and dispositional empathic concern, while considering different characteristics of the stimuli to be rated. Our research questions and hypotheses were as followed:

- 1) How is the listener’s perceived anxiety of the player influenced by the stimulus modality, performance condition, and piece performed, while considering the listeners’ musical background?

Hypothesis 1.1: There will be differences in perceived anxiety of the player depending on the stimulus modality, performance condition, and piece performed.

Hypothesis 1.2: The anxiety of the player will be perceived differently depending on the listener’s musical background.

- 2) Is the perceived vs. experienced anxiety influenced by the stimulus modality, performance condition, and piece performed, while considering the listener’s musical background and empathic concern?

Hypothesis 2.1: There will be differences in the perceived vs. experienced anxiety depending on the stimulus modality, performance condition, and piece performed.

Hypothesis 2.2: There will be differences in perceived vs. experienced anxiety depending on the listener’s musical background.

Hypothesis 2.3: The differences in the perceived vs. experienced anxiety will be related to the listener’s empathic concern.

Methods

The experimental design was based on Kwan (2016). The author of this thesis recorded the second (Allegretto) and third (Presto) movement of Beethoven’s piano sonata no. 6 in F major, op. 10, no. 2 from memory. These were performed in absence of an audience (rehearsal) and for an online audience (recital). Recordings were edited to obtain approximately 1-min clips in audiovisual and audio-only modalities. In total, the set of stimuli comprised of eight clips. Five professional pianists observing the pianist rated the Second movement as more ample in body movements than the Third movement. The pianist self-rated each of his performances in terms of how anxious he felt while playing (see Figure 6).

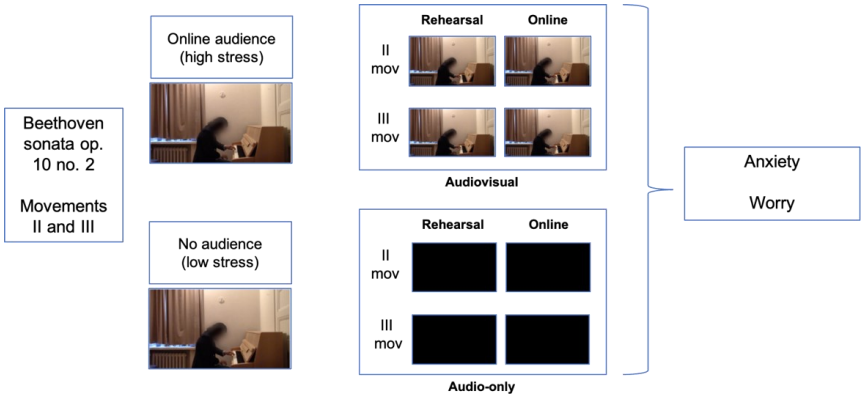


Figure 6. Stimuli creation.

Forty-eight participants were recruited online to rate the eight clips presented in randomized order. They were unaware that the pieces came from different performances. Participants rated each clip using the same scale the pianist used, thus allowing comparisons of both set of scores. After completing the perceptual study, participants filled the Interpersonal Reactivity Index (Davis, 1980) from which we obtained their empathic concern, the Goldsmith Music Sophistication Index (Müllensiefen et al., 2014) from which we divided the sample in musicians and non-musicians, and demographic questionnaires (see Figure 7).

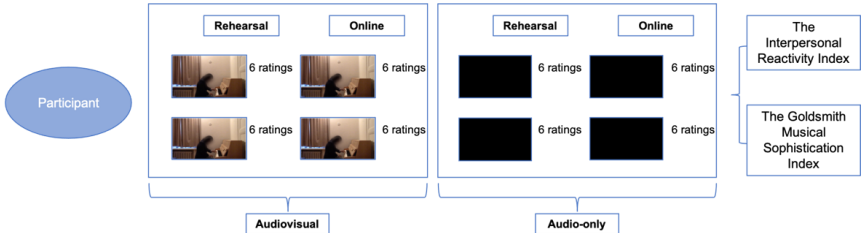


Figure 7. Structure of perceptual experiment.

Results

Regarding main effects, participants perceived higher anxiety in the Third movement ($M = 40.24$, $SE = 1.99$, 95% CI [36.23, 44.25]), when compared to the Second movement ($M = 35.50$, $SE = 1.76$, 95% CI [31.95, 39.04]), $F(1, 46) = 6.35$, $p = .015$, $r = .35$. In addition, audio-only stimuli received higher anxiety scores ($M = 39.62$, $SE = 1.89$, 95% CI [35.82, 43.42]), than audiovisual stimuli ($M = 36.12$, $SE = 1.72$, 95% CI [32.65, 39.59]), $F(1, 46) = 4.95$, $p = .031$, $r = .31$.

The listener's musical background became relevant when interacting with the piece performed (Figure 8). While non-musicians perceived approximately the same anxiety in the Second ($M = 37.21$, $SE = 2.64$, 95% CI [31.89, 42.52]) and Third movements ($M = 36.19$, $SE = 2.99$, 95% CI [30.18, 42.21]), musicians perceived higher anxiety in the Third movement ($M = 44.30$, $SE = 2.64$, 95% CI [38.99, 49.60]), when compared to the Second movement ($M = 33.79$, $SE = 2.33$, 95% CI [29.10, 38.48]), $F(1, 46) = 9.37$, $p = .004$, $r = .41$.

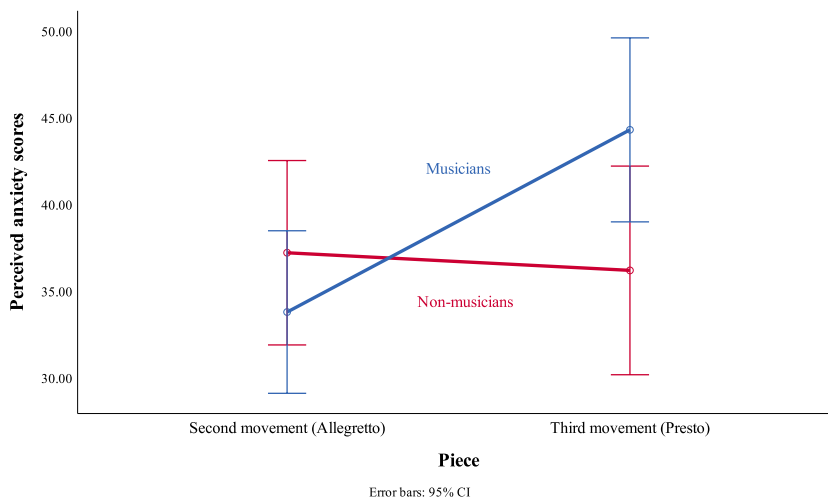


Figure 8. Interaction effect of piece performed and musical background (From Chang-Arana, Á. M., Mavrolampados, A., Pokki, N., & Thompson, M. R. (2022). Exploring the interpersonal level of music performance anxiety: Online listener's perception and accurate inference of anxiety. *Frontiers in Psychology* 13:838041, p. 7. Copyright The Authors, 2022. Published by Frontiers in Psychology).

An interaction was observed between the piece performed and the stimulus modality (Figure 9). Anxiety was perceived similarly when rating audio-only clips of the Second ($M = 39.19$, $SE = 2.17$, 95% CI [34.82, 43.55]) and Third ($M = 40.06$, $SE = 2.23$, 95% CI [35.56, 44.55]) movements. However, when rating audiovisual clips, more anxiety was perceived in the Third movement ($M = 40.43$, $SE = 2.12$, 95% CI [36.16, 44.70]) when compared to the Second movement ($M = 31.81$, $SE = 2.04$, 95% CI [27.70, 35.92]), $F(1, 46) = 8.66$, $p = .005$, $r = .40$.

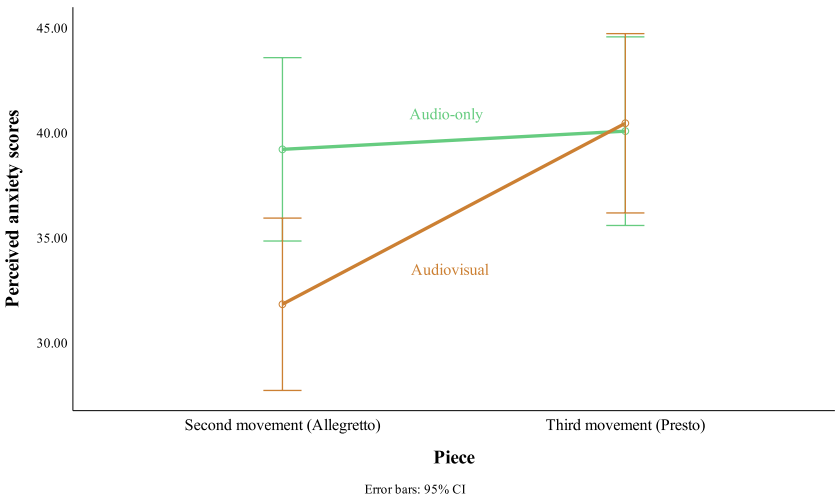


Figure 9. Interaction effect of piece performed and stimulus modality (From Chang-Arana, Á. M., Mavrolampados, A., Pokki, N., & Thompson, M. R. (2022). Exploring the interpersonal level of music performance anxiety: Online listener’s perception and accurate inference of anxiety. *Frontiers in Psychology* 13:838041, p. 7. Copyright The Authors, 2022. Published by Frontiers in Psychology).

The initial comparison between the listener’s perceived anxiety scores and the musician’s experienced anxiety scores revealed that the musician reported much higher anxiety than what listeners perceived (Table 10).

Table 10. Listener’s mean perceived anxiety scores and musician’s self-rated anxiety.

Movement	Performance condition and stimulus modality	Listeners' mean perceived anxiety scores	Pianist's self-rated anxiety	Difference	SD	95 % CI	
						Lower	Upper
Second	Practice Audio	37.95	50.00	-12.05	18.82	-17.52	-6.59
	Practice Audiovisual	32.50	57.00	-24.50	18.25	-29.80	-19.20
	Recital Audio	40.03	44.50	-4.47	15.51	-8.97	0.04
	Recital Audiovisual	30.66	60.50	-29.84	16.98	-34.77	-24.91
Third	Practice Audio	40.67	44.00	-3.33	15.16	-7.73	1.07
	Practice Audiovisual	40.94	44.50	-3.56	17.50	-8.64	1.52
	Recital Audio	40.53	67.00	-26.47	22.49	-33.00	-19.94
	Recital Audiovisual	40.86	71.50	-30.64	17.70	-35.78	-25.49

Note: (From Chang-Arana, Á. M., Mavrolampados, A., Pokki, N., & Thompson, M. R. (2022). Exploring the interpersonal level of music performance anxiety: Online listener’s perception and accurate inference of anxiety. *Frontiers in Psychology* 13:838041, p. 8. Copyright The Authors, 2022. Published by Frontiers in Psychology).

The ERA was the difference between the listener’s perceived anxiety scores and the musician’s experienced anxiety scores. This value was squared and squared rooted to obtain positive values, where values closer to 0 indicated higher accuracy and values larger than 0, lower accuracy.

Similarly to the listener’s perception analyses, the listener’s musical background became relevant when interacting with the piece performed (Figure 10). After adjusting for the empathic concern of the listeners, non-musicians ($M = 22.11, SE = 1.99, 95\% \text{ CI } [18.11, 26.11]$) and musicians ($M = 21.79, SE = 1.75, 95\%$

CI [18.27, 25.32]) had a similar ERA in the Second movement. However, in the Third movement, the ERA of musicians increased ($M = 18.51$, $SE = 1.80$, 95% CI [14.90, 22.13]), while the ERA of non-musicians decreased ($M = 26.17$, $SE = 2.04$, 95% CI [22.07, 30.27]), $F(1, 45) = 7.60$, $p = .008$, $r = .38$. The listener's empathic concern did not have an influence on their ERA, $F(1, 45) = 4.75$, $p = .035$, $\eta p^2 = .10$.

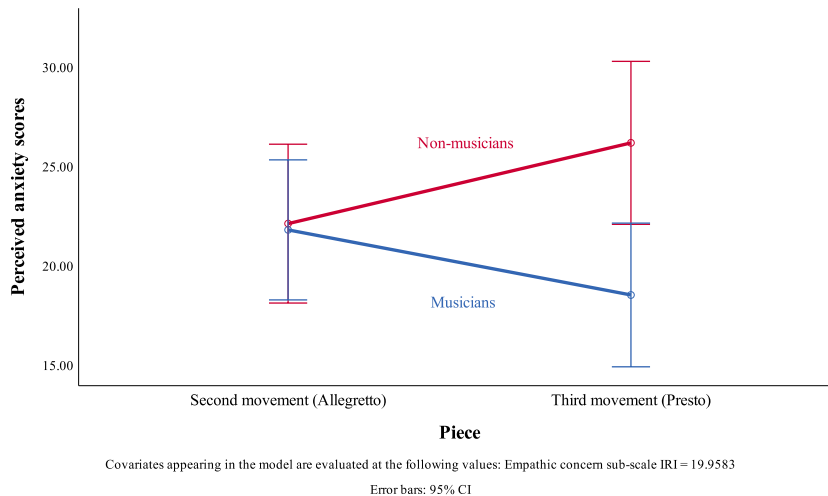


Figure 10. Interaction effect of piece performed and musical background (From Chang-Arana, Á. M., Mavrolampados, A., Pokki, N., & Thompson, M. R. (2022). Exploring the interpersonal level of music performance anxiety: Online listener's perception and accurate inference of anxiety. *Frontiers in Psychology* 13:838041, p. 9. Copyright The Authors, 2022. Published by Frontiers in Psychology).

Conclusions

This study focused on the fear of negative evaluation, a cognitive symptom of MPA. The listener's perception of anxiety is complex and is affected by the piece rated, the stimulus modality and the interaction with the listener's musical background. Listeners with no musical background perceived similar anxiety when rating both pieces. However, listeners with musical background perceived higher anxiety in the Third movement vs. the Second movement. This judgment coincided with the pianist's own anxiety experience. This suggests that listeners with a musical background may have had the skills to recognize the technical challenges of a piece. The perception of anxiety increased when the listener could not see the pianist. This is indicative of the "protective effect" of ancillary gestures. The presence of these secondary body movements which are not primarily responsible to produce sound are associated with expressiveness, phrasing, positive musical experiences, etc. The reason why the Third movement received higher anxiety scores than the Second movement could be explained by the amplitude of the pianist's ancillary gestures. The pianist's movements during the Third movement were more constraint, thus limiting the body expressivity. It was also interesting that the listeners did not perceive anxiety differently according to the performance condition, even though the pianist was more anxious in the recital condition vs. rehearsal condition. Music information

retrieval analysis also revealed that during the recital conditions tempo was slightly increased in comparison to the rehearsal condition. Yet, this slight increase in tempo may not have been enough for listeners to perceive a noticeable difference.

The listener's ERA revealed two facts. First, the musician experienced much higher anxiety across conditions than what the listeners perceived. That is, the musician's experienced anxiety was underestimated by the listeners. The listener's musical background became relevant for a better ERA when it came to the Third movement. The anxiety perceived by listeners with a musical background was closest to the experienced anxiety of the pianist in the Third movement when compared to the Second movement. Once again, perhaps the musical background of listeners allowed them to identify the technical challenges and the struggle the pianist had with the fast movement, while this remained unnoticeable for listeners without a musical background.

Finally, experimental MPA studies include listeners to create a performance context and thus induce anxiety in the musicians, or, in addition, to obtain judgments of performance quality. This study expanded the role of listeners in experimental MPA studies by studying more in depth their perception.

4. Discussion

The overall goal of this dissertation was to explore the contextual factors which influence IA. To do so, we address the following questions:

1. How to implement an IA approach to quantify the understanding between dyads interacting in design and music?
2. How accurate are perceivers when inferring the mental experiences of targets?
3. What contextual factors affect IA?

4.1 How to implement an IA approach to quantify the understanding between dyads interacting in design and music?

A precondition to implement IA into design and music was to recognize IA as an adequate approach. The conception of this thesis began with exploring what do designers meant by empathy. Designers emphasize the importance of understanding users correctly to obtain desirable outcomes. That is, empathy is deemed key to successful design (Brown 2009; Kramer et al., 2016; Leonard & Rayport, 1997). While designers are aware of the importance of empathy for their practice, one struggles to find a consensual and clear definition of what do they mean by it. This situation has resulted in a wide scope of definitions and working tools for fostering empathy towards a user yet lacking specificity to actually answer the question (Study I). IA, and EA in particular, provided an interesting approach to empathy in design from yet an unexplored angle. Later, IA triggered our curiosity to study the interactions between listeners and musicians from a different perspective. Most literature on music performance have investigated listeners and performers separately, but IA allowed connecting both social agents to understand further what happens in the complex context of music performance.

Study II was the first time we adapted EA to design. For this purpose, several steps were taken. The implementation began by identifying a real design challenge in a specific group of musicians. We focused the case on the problems that reeds and other instrumental equipment posit for woodwind musicians (Thompson, 1979; Ledet, 1981; Almeida et al., 2013). A designer interviewed relevant users with the aim of gaining further insight into their experiences with these musical equipments. We adapted EA to suit this interaction and quantify the designers' accuracy when inferring the thoughts and feelings of the musicians. Furthermore, we proxied a post-interview stage where designers synthesized the user's needs. To test the alleged causal properties of empathy in design outcomes, we adapted the "empathy map" (Both & Baggereor, n.d.) from its four

quadrants (i.e., “say”, “do”, “think” and “feel”) to two (i.e., “think” and “feel”). Simultaneously, a detailed understanding of the different EA paradigms was needed to adapt them to the design context. Researchers interested in the specific tools/instructions used in this study are encouraged to revise the Appendix section of Study II.

Study III was an opportunity to implement EA into two additional design cases, further showing the flexibility of EA for the contextual nature of design. The “thoughts and feelings” categories were rephrased to “design-related and non-design-related” categories. This modification allowed us to further test the EA of designers towards mental contents related to design vs. those which were not. Thus, Studies II and III have included new elements into the EA paradigms, such as supporting the conversation by displaying objects, conducting a semi-structured interview on a very specific topic, and thinking of mental contents in other terms than thoughts and feelings.

For Study IV, we had to consider a different approach than EA to studying IA, given that the communication between listeners and musicians is predominantly non-verbal. Thus, a suitable method for such interaction was ERA. The author of this thesis performed and recorded a piano recital in rehearsal and recital conditions. The recordings were edited into audiovisual and audio-only conditions, given the influence stimuli modality has on perception (Davidson, 1993; Huang & Krumhansl, 2011; Juchniewicz, 2008; Kwan, 2016; Thompson & Luck, 2012; Tsay, 2013, 2014; Vines et al., 2006; Vuoskoski et al., 2016). We operationalized ERA as the difference between the musician’s experienced anxiety while performing to the listener’s perceived anxiety while listening. In the literature of MPA, no previous attempts to do such comparison have been found. Study IV expands the methods available to study ERA beyond the use of static facial expressions to naturalistic music performances.

4.2 How accurate are perceivers when inferring the mental experiences of targets?

Perceivers were rather inaccurate when inferring the mental experiences of targets. Implementing EA in a design context allowed us to observe that designers understood with about 50 % EA the mental contents of the musicians. This percentage is surprisingly high when compared to previous EA studies done in different contexts and dyads, where accuracy ranges between 20 % – 30 % (Clements, Holtzworth-Munroe, Schweinle, & Ickes, 2007; Ickes & Hodges, 2013; Marangoni et al., 1995). We believe this relatively high percentage may be due to the highly specific context of the conversations. The interviews focused on the musicians’ experiences with reeds and instrumental accessories, rather than open-ended conversations. Such narrowed focus of conversation where objects could be manipulated and shown may have assisted designers in inferring the mental contents of the users. Yet, the average EA observed in design suggests that IA in design interactions could be further improved.

We failed at establishing a link between EA and design outcomes. Our small sample (two designers and five users) is a serious limitation. Yet, we hope they are indicative for future researchers of what studies with larger samples could

test. Relatedly, to extract meaningful conclusions of the relationship between EA and design outcomes, future researchers should improve how design outcomes are defined and measured. Design techniques have been investigated as potential tools to foster “empathy” towards users (as we reviewed in Study I). However, their predictive properties as empathy-generating tools need to be further explored. One approach towards this exploration would be applying contemporary psychometric research into the analysis of empathy-generating tools, particularly by analyzing validity evidence based on content, based on the relationship with other variables and based on the consequences of using these techniques (American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 2014). The creation of solid design-outcomes measuring tools are of great importance for future explorations of the link between empathy in design and design outcomes.

The relation of facial expressions and physiological signals with design outcomes should be retested with an adequate paradigm. In Study II we made a crossover between two forms of EA paradigms. This combination did not show meaningful results. Thus, the study of designer’s EA using physiological and continuous EA paradigms remains unexplored. Additional physiological indices, like skin conductance, could be relevant to include in such paradigms. Facial reactions, alone, are not sufficient indices of participant’s automatic emotions and, therefore, other physiological indices should be considered.

Study III allowed us to dwell deeper into one conclusion of Study II: users’ mental contents may not be all relevant for the purposes of design. EA originally categorizes mental contents as thoughts or feelings. In Study III we proposed a different categorization better tuned for the design context, by classifying mental contents as design-related mental contents from others. Design-related entries were around a median of 56.11 vs. 31.94 of all other entries. The high accuracy of designers may be due to the highly specific context of the design task. These numbers must be interpreted cautiously, though, as the design cases were different between each other, and participants were not exposed to all three design cases. Yet, even observing each design case individually, we found support to our claim on the influence of the design context.

Study III contributes in two other ways. First, adapting Icke’s thought-feeling categories to design-related and non-design-related categories. This simple change, although unprecedented, allowed uncovering the interesting results previously pointed at. Second, reporting EA for thoughts and feelings separately. This seldomly happens in the literature of EA. The higher accuracy for feelings in the visual impairment sub-study may reflect how designers inference skills change depending on the design stage. In this sub-study, the focus was mostly on the general experience of a person with visual impairment. On the other two sub-studies, the focus was narrowed to specific products and less to general experiences.

Results of Study IV showed that in the context of music performance, perceivers were also inaccurate when inferring a target’s anxiety experience. Table 10 reveals that listeners consistently underestimated the actual anxiety experience reported by the musician. Yet, listeners were right in perceiving higher

anxiety in the Third movement when compared to the Second movement, as the pianist reported more concerns performing the Third movement in contrast to the Second movement. Differences in ERA were also identified depending on the stimulus modality and the listener's musical background. But these latter two will be further discussed in sub-sections 4.3 and 4.4.

4.3 What contextual factors affect IA?

Multimodal perception did not affect the perceiver's ERA. However, we think the effect of multimodal perception cannot be discarded yet. It is worth pointing at some methodological choices in Study IV that could account for the null results. In Study IV we adjusted our p value to .013. The reason being that we controlled for family-wise error by dividing .05 times the number of hypotheses (i.e., four) (Rubin, 2017). In addition, including empathic concern from the Interpersonal Reactivity Index (Davis, 1980) as a covariate into the model testing the effect of multimodal perception on ERA could have withdrawn power to detect an effect. Given the exploratory nature of Study IV, in our supplementary results we excluded empathic concern as a covariate. The reason being that, at least in the EA literature, perceiver's self-reported empathic dispositions do not correlate with their empathic performance skills (Ickes, 2001; Stueber, 2019). As a result of removing empathic concern from our model, our p value increased to .05 in the supplementary text.

Supplementary results suggest that listeners obtained higher ERA in the audio-only conditions when compared to the audiovisual conditions. That is, in the absence of visual stimulation, listeners were more accurate when inferring the anxiety experienced by the musician. Why did their ERA decreased when they could see the musician in addition to hearing him? Would it not be expected that the more sensorial information a perceiver has at their disposal, the higher IA they will obtain? This difference could be explained by the "protective effects" of the ancillary gestures that perceivers could observe in the audiovisual stimuli. An increase in ancillary gestures of a performer have been associated with higher perception of expressiveness, interest, dynamics, phrasing, and overall musical performance (Juchniewicz, 2008; Broughton & Stevens, 2009). Thus, the ancillary gestures displayed by the pianist could have persuaded the listeners that he was more in control and absorbed in the performance than what he was, hiding the negative experience of anxiety. In fact, in the absence of the "distraction" of ancillary gestures, participants gave more accurate ratings of the pianist's anxiety experience.

Based on the results observed in the design and music studies, the perceiver's background influences their IA. As mentioned in Section 4.2, designers were more accurate when inferring design-related entries vs. non-design-related entries. This may suggest that the designers' background allowed them to identify design-relevant cues. Future studies could further explore whether perceivers without a design background do not show differences in EA between design-related and non-design-related entries.

Study IV suggests that musical background influenced ERA when interacting with the piece performed. The ERA obtained by musicians and non-musicians

in the Second movement was roughly the same. However, the ERA of perceivers with musical background increased in the Third movement, while the ERA of perceivers without a musical background decreased for the same movement. Perceivers with a musical background could have been more aware of the technical difficulties posed by the Third movement and deem these as more challenging than the ones of the Second movement. This interpretation is aligned with the pianist's own appreciation of the higher difficulties posed by the Third movement vs. the Second movement. It is unclear why the ERA of listeners without a musical background decreased in the Third movement. Taken together, Studies II-IV suggest that the perceiver's background and previous experiences are relevant to consider for future IA studies.

4.4 Future directions and limitations

In Study I we compiled relevant literature of empathy from psychology and design. However, other relevant literature could have been omitted. Relatedly, the selection of the sources did not follow a random sampling or meta-analytic approaches. Instead, we built our reference list starting from review articles and book chapters in psychology and design, following a snowball procedure, modifying the text following peer reviewing process, and incorporating new sources as we came across them in a three-years interval. Nonetheless, we believe readers will appreciate the compiling effort, the mapping of empathy as an umbrella construct, and the advice to approach this concept carefully in future research.

In Study II the important limitation to mention is the small sample size. Our findings are exploratory at this stage and should be interpreted cautiously. However, future researchers interested in EA should take care of the duration of the tasks since each user required approximately four hours to complete all the tasks described in Study II. We also deviated slightly from Ickes' protocol (2001), particularly on the role of the designer after each interview. A strict adherence of Ickes' instructions would have required the designer to infer the mental contents of the user right after they reported their remembered thoughts and feelings. Yet, since Designer 1 had five musicians to interview, the inference step was postponed until completion of interviews to avoid a practice effect. It is also important to underscore that user-understanding is a complex process where interviews are one of the methods employed (Sanders & Stappers 2014; Oygür 2018). Study II had a simplified version of need-finding and so it is not very representative of how designers operate in real life. Another important limitation of Study II was the chosen design task. We asked designers to complete an empathy map, again as a simplified recreation of tasks designers may carry after interacting with a user. Although we collected physiological data using EMG, GSR, and ECG, we reported only EMG data due to two reasons: First, the GSR sensors used did not allow us to obtain adequate data for pre-processing the signal. Second, given time constraints and already long extension of the study, we decided to focus only on EMG, given that previous literature have used facial muscle movements as an indicator of emotional contagion. Future researchers interested in expanding research of EA in design may revise the appendices of Study II. We included detailed documentation that allows replications of our findings or applying our methodology to future studies.

The main limitation of Study III is that portraying a unified picture is not possible, as there were three very different design cases with different users and participants. Yet, this limitation also exhibits the flexibility of EA to adapt to different design cases. In Study III we raised two questions which remain open for future researchers. One is what constitutes a design-related content. We proposed our own operationalization of what constitutes such content. Independently of whether readers agree or not with it, we encourage future designers to take care of clearly defining their variables of study. The second question is what constitutes a designer. This methodological question is particularly important for future studies about the effects of design education, comparing designers vs. non-designers, transference of design skills, etc. The realization of comparing designers vs. non-designers occur after completing Study IV, where we did compare perceivers with musical vs. non-musical background.

Studies II and III were our first steps into incorporating a quantitative paradigm to measuring mutual understanding of designers and users in a user-centered interview. As such, we encountered a series of challenges and limitations reported in the articles.

Lastly, in Study IV we reported limitations on data collection, stimuli's design, and generalizability of results. This study, conceived during the COVID-19 pandemic, focused on how online listeners perceive anxiety. Thus, a careful control of the listening conditions was not feasible, even though participants were instructed to use of headphones, laptop/computers, and avoid distractions. Regarding the stimuli, a few observations must be raised. Contrarily to most multimodal music perception studies, we decided not to include video-only conditions; a decision reported previously in the literature (Broughton & Stevens, 2009; Huang & Krumhansl, 2011). One reason was due to ecological validity. It is uncommon to observe performance without any music. The other reason was due to time constraints. Just by including audio-only and audiovisual conditions, the study took one hour. Given that it was fully online, we had a concern about participant attrition, and so we decided to keep the essential stimuli. Relatedly, the audiovisual stimuli did not show the face of the pianist. This deprived participants from an important source of expressive information. Yet, we decided to blur the face of the pianist to diminish the considerations of data protection and because a similar ongoing study will recruit more pianists whose faces will be blurred, allowing comparisons with Study IV. Additionally, rating 1-min musical clips hinders a full appreciation of the piece, yet a compromise had to be done to allow feasibility of the study. Finally, researchers interested in using our methodology for replication or other future studies may find the stimuli used, as well as the MATLAB code used for analyzing their sound properties.

4.5 Recommendations

We conclude Section 4 with a few general recommendations based on the results of Studies I-IV and previous literature on IA.

4.5.1 Ask. Do not assume.

Asking about someone else's perspective is an effective way of increasing IA. Yet, perspective getting (Damen et al., 2021; Damen, van Amelsvoort et al., 2021) is rarely adopted by perceivers. Perceivers confident they will accurately infer a target's mental contents are more likely to be less accurate because they will rely on their own knowledge and stereotypes about the target, rather than asking directly from them (Damen et al., 2021; Eyal, Steffel, & Epley, 2018). A perceiver should be particularly careful when they share similar experiences with the target because, rather than asking, the perceiver is likely to use their own experiences as a frame of reference when processing the target's memories, likely biasing this perception, and interfering with processing the new information from the target (Damen et al., 2021; Damen, van Amelsvoort et al., 2021; Kang et al., 2022). The high EA displayed by designers in comparison to previous EA studies may be explained by the fact that the interaction was not a casual conversation, but rather an interview where designers were interested in getting the perspective of the users. Designers could benefit from approaching a user with curiosity and openness, especially if they are approaching a context, they have had previous experience with. Assuming they will know better than a user will lead to inaccuracies.

4.5.2 Exemplify and structure.

Relatedly to Sub-section 4.6.1, we recommend supporting a conversation with different resources to exemplify. The display of accessories and instruments may have supported the designers' inferences. Yet, ways of improving their IA should be further investigated as the average 50 % EA allows room for improvement. In addition to supporting the conversation with objects, structuring an interview may facilitate paying attention towards relevant information.

4.5.3 Context matters.

To conduct IA research in design and music, it was necessary to understand how interactions occur in these two contexts. For design, that meant understanding what designers meant by empathizing with a user, how designers approach them, and what are relevant questions and cues for designers. For music performance, it meant considering the effect multimodal perception and musical background have on the listener's perception and choosing an appropriate IA measuring approach for the predominantly non-verbal nature of music performance. We encourage future IA researchers to be aware of the particularities affecting their context of interest.

5. General conclusions

Human cooperation has been key in our evolution. It would not have been successful unless we had certain degree of accuracy in understanding each other's mental states. This thesis focused on how to quantify the IA of people interacting in two contexts of interaction: design and music.

We exemplified how to implement different paradigms of IA into contexts as different as design and music performance. We described a series of steps necessary to recreate design and music performance interactions in controlled settings, which allowed quantifying the IA of different dyads. These methods may be useful for future researchers interested in investigating IA in other contexts.

A common finding across contexts is that we are rather inaccurate at inferring the mental states of others. Our findings suggest that IA is influenced by contextual variables such as the sensorial channel through which dyads interact, as well as the background of the perceivers.

6. References

- Almeida, A., George, D., Smith, J., & Wolfe, J. (2013). The clarinet: How blowing pressure, lip force, lip position and reed “hardness” affect pitch, sound level, and spectrum. *J. Acoust. Soc. Am.*, 134(3). <https://doi.org/10.1121/1.4816538>
- Alzayed, M. A., McComb, C., Menold, J., Huff, J., & Miller, S. R. (2021). Are you feeling me? An exploration of empathy development in engineering design education. *ASME. Journal of Mechanical Design*, 143(11), 112301. <https://doi.org/10.1115/1.4048624>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington DC: American Educational Research Association.
- Ball, L. J. & Christensen, B. T. (2019). Advancing an understanding of design cognition and design metacognition: Progress and prospects. *Design Studies*, 65, 35–59. <https://doi.org/10.1016/j.destud.2019.10.003>
- Bänziger, T. (2016). Accuracy of judging emotions. In J. A. Hall, M. S. Mast, & T. V. West (Eds.), *The social psychology of perceiving others accurately* (pp. 23–51). Cambridge University Press. <https://psycnet.apa.org/doi/10.1017/CBO9781316181959.002>
- Bänziger, T., Scherer, K. R., Hall, J. A., & Rosenthal, R. (2011). Introducing the MiniPONS: A short multichannel version of the Profile of Nonverbal Sensitivity (PONS). *Journal of Nonverbal Behavior*, 35, 189–204. <https://doi.org/10.1007/s10919-011-0108-3>
- Batson, C. D. (2009). These things called empathy: Eight related but distinct phenomena. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 3–15). MIT Press. <https://doi.org/10.7551/mitpress/9780262012973.003.0002>
- Battarbee, K. (2004). Co-experience: Understanding user experiences in interaction (doctoral dissertation). Aalto University. Retrieved from <https://aaltodoc.aalto.fi/handle/123456789/11462>

- Battarbee, K., Baerten, N., Hinfelaar, M., Irvine, P., Loeber, S., Munro, A., & Pederson, T. (2002). Pools and satellites: Intimacy in the city. *DIS '02: Proceedings of the 4th conference on designing interactive systems: Processes, practices, methods, and techniques, USA*, 237-245. ACM Press, New York. <http://dx.doi.org/10.1145/778712.778746>
- Bell, C., Bourke, C., Colhoun, H., Carter, F., Frampton, C., & Porter, R. (2011). The misclassification of facial expressions in generalised social phobia. *Journal of Anxiety Disorders*, 25, 278–83. <https://doi.org/10.1016/j.janxdis.2010.10.001>
- Bernstein, M. J., & Claypool, H. M. (2012). Social exclusion and pain sensitivity: Why exclusion sometimes hurts and sometimes numbs. *Personality and Social Psychology Bulletin*, 38(2), 185–196. <https://doi.org/10.1177/0146167211422449>
- Bjornsdottir, R. T., Alaei, R., & Rule, N. O. (2017). The perceptive proletarian: Subjective social class predicts interpersonal accuracy. *J Nonverbal Behav*, 41, 185–201. <https://doi.org/10.1007/s10919-016-0248-6>
- Both, T. & Baggereor, D. (n.d.) Bootcamp Bootleg [Online], online document (downloadable on July 15th 2019) Retrieved from <https://dschool.stanford.edu/resources/the-bootcamp- bootleg>
- Brotons, M. (1994). Effects of performing conditions on music performance anxiety and performance quality. *J. Music. Ther.* 31, 63–81. <https://psycnet.apa.org/doi/10.1093/jmt/31.1.63>
- Broughton, M. C. & Davidson, J. W. (2014). Action and familiarity effects on self and other expert musician's Laban effort-shape analyses of expressive bodily behaviors instrumental music performance: a case study. *Front. Psychol.* 5,1201. <https://doi.org/10.3389/fpsyg.2014.01201>
- Broughton, M. C. & Stevens, C. (2009). Music, movement and marimba: an investigation of the role of movement and gesture in communicating musical expression to an audience. *Psychol. Music* 37, 137–153. <https://doi.org/10.1177/0305735608094511>
- Brown, T. (2009). *Change by design: How Design Thinking transforms organizations and inspires innovation*. Harper Business.
- Brugués, A. O. (2011a). Music performance anxiety—part 1: a review of its epidemiology. *Med. Probl. Perform. Art.* 26, 102–105. <https://www.proquest.com/scholarly-journals/music-performance-anxiety-part-1-review/docview/873493908/se-2?accountid=27468>
- Brugués, A. O. (2011b). Music performance anxiety—part 2: a review of treatment options. *Med. Probl. Perform. Art.* 26, 164–171.

<https://www.proquest.com/openview/43d0b27ac19aacf8754602c6fc61fee2/1?pq-origsite=gscholar&cbl=30487>

- Cacioppo, J. T. & Patrick, W. (2008). *Loneliness: Human nature and the need for social connection*. U.S.A.: W. W. Norton & Company.
- Cash, P. J. (2018). Developing theory-driven design research. *Design Studies*, 56, 84–119. <https://doi.org/10.1016/j.destud.2018.03.002>
- Chang-Arana, Á. M., Mavrolampados, A., Thompson, M. R., Pokki, N., & Sams, M. (2022). Exploring the interpersonal level of music performance anxiety: Online listener's perception and accurate inference of anxiety. *Frontiers in Psychology*, 13:838041. <https://doi.org/10.3389/fpsyg.2022.838041>
- Chang-Arana, Á. M., Piispanen, M., Himberg, T., Surma-aho, A., Alho, J., Sams, M., & Hölttä-Otto, K. (2020). Empathic accuracy in design: Exploring design outcomes through empathic performance and physiology, *Design Science*, 6. <https://doi.org/10.1017/dsj.2020.14>
- Chang-Arana, Á. M., Surma-aho, A., Hölttä-Otto, K., & Sams, M. (2022). Under the umbrella: Components of empathy in psychology and design. *Design Science*, 8. <https://doi.org/10.1017/dsj.2022.13>
- Chang, A., Livingstone, S. R., Bosnyak, D. J., & Trainor, L. J. (2017). Body sway reflects leadership in joint music performance. *Proc. Natl. Acad. Sci. U. S. A.* 114, E4134–E4141. <https://doi.org/10.1073/pnas.1617657114>
- Clark, M. A., Robertson, M. M., & Young, S. (2018) “I feel your pain”: A critical review of organizational research on empathy. *Journal of Organizational Behavior*, 40, 166–192. <https://doi.org/10.1002/job.2348>
- Clarke, E., DeNora, T., & Vuoskoski, J. (2015). Music, empathy and cultural understanding. *Physics of Life Reviews* 15, 61–88. <https://doi.org/10.1016/j.plrev.2015.09.001>
- Clements, K., Holtzworth-Munroe, A., Schweinle, W., & Ickes, W. (2007). Empathic accuracy of intimate partners in violent versus nonviolent relationships. *Pers. Relatsh.*, 14(3), pp. 369–388. <https://doi.org/10.1111/j.1475-6811.2007.00161.x>
- Cohen, S., Schulz, M. S., Liu, S. R., Halassa, M., & Waldinger, R. J. (2015), Empathic accuracy and aggression in couples: Individual and dyadic links. *Fam Relat*, 77, 697–711. <https://doi.org/10.1111/jomf.12184>
- Costanza-Chock, S. (2020). *Design Justice*. The MIT Press.

- Cox, W. J. & Kenardy, J. (1993). Performance anxiety, social phobia, and setting effects in instrumental music students. *Journal of Anxiety Disorders*, 7, 49-60. [https://doi.org/10.1016/0887-6185\(93\)90020-L](https://doi.org/10.1016/0887-6185(93)90020-L)
- Crits-Christoph, P., Gibbons, M. B. C., Temes, C. M., Elkin, I., & Gallop, R. (2010). Interpersonal accuracy of interventions and the outcome of cognitive and interpersonal therapies for depression. *Journal of Consulting and Clinical Psychology*, 78(3), 420-428. <https://doi.org/10.1037/a0019549>
- Cuff, B. M. P., Brown, S. J., Taylor, L., & Gowat, D. J. (2016). Empathy: A review of the concept, *Emotion Review*, 8(2), 144-153. <https://doi.org/10.1177/1754073914558466>
- Dahl, S., & Friberg, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music Percept.* 24, 433-454. <https://doi.org/10.1525/mp.2007.24.5.433>
- Damen, D., Pollmann, M. M. H., & Grassow, T-L. (2021). The benefits and obstacles to perspective getting. *Front. Commun.* 6:611187. <https://doi.org/10.3389/fcomm.2021.611187>
- Damen, D., van Amelsvoort, M., van der Wijst, P., Pollmann, M., & Krahmer, E. (2021). Lifting the curse of knowing: How feedback improves perspective-taking. *Quarterly Journal of Experimental Psychology*, 74(6), 1054-1069. <https://doi.org/10.1177/1747021820987080>
- Darwin, C. (1871/1981). *The descent of man, and selection in relation to sex*. U.S.A.: Princeton University Press.
- Davidson, J. W. (1993). Visual perception of performance manner in the movements of solo musicians. *Psychol. Music*, 21, 103-113. <https://doi.org/10.1177/030573569302100201>
- Davidson, J. W. (2005). "Bodily communication in musical performance," in D. Miell, R. MacDonald & D. J. Hargreaves (eds.), *Musical Communication* (pp. 215-237). Oxford: Oxford University Press.
- Davidson, J. W. (2007). Qualitative insights into the use of expressive body movement in solo piano performance: A case study approach. *Psychology of Music*, 35(3), 381-401. <https://doi.org/10.1177/0305735607072652>
- Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*. 10(85). Retrieved from http://www.uv.es/~friasnav/Davis_1980.pdf (Accessed April 26, 2022).
- Davis, M. H. (2006). Empathy. In J. E. Sets, and J. H. Turner (Eds.). *Handbook of the sociology of emotions* (pp. 443-466). U.S.A.: Springer.

- Devecchi, A. & Guerrini, L. (2017). Empathy and design. A new perspective. *The Design Journal*, 20(1), S4357–S4364. <https://doi.org/10.1080/14606925.2017.1352932>
- Eyal, T., Steffel, M., & Epley, N. (2018). Perspective mistaking: Accurately understanding the mind of another requires getting perspective, not taking perspective. *Journal of Personality and Social Psychology*, 114(4), 547–571. <https://doi.org/10.1037/pspa0000115>
- Fernholz, I., Mumm, J. L. M., Plag, J., Noeres, K., Rotter, G., Willich, S. N., et al. (2019). Performance anxiety in professional musicians: A systematic review on prevalence, risk factors and clinical treatment effects. *Psychol. Med.*, 49, 2287–2306. <https://doi.org/10.1017/S0033291719001910>
- Fila, N. D., & Hess, J. L. (2014). Exploring the role of empathy in a service-learning design project, *Design Thinking Research Symposium*, 10. Retrieved from <https://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1037&context=dtrs>
- Flôres, R. G. & Ginsburgh, V. A. (1996). The Queen Elisabeth musical competition: How fair is the final ranking? *Journal of the Royal Statistical Society. Series D (The Statistician)*, 45, 97–104. <https://doi.org/10.2307/2348415>
- Fujiwara, K. & Daibo, I. (2022). Empathic accuracy and interpersonal coordination: Behavior matching can enhance accuracy but interactional synchrony may not. *The Journal of Social Psychology*, 162(1), 71–88. <https://doi.org/10.1080/00224545.2021.1983509>
- Gerber, J., & Wheeler, L. (2009). On Being Rejected: A Meta-Analysis of Experimental Research on Rejection. *Perspectives on Psychological Science*, 4(5), 468–488. <https://doi.org/10.1111/j.1745-6924.2009.01158.x>
- Gollan, J. K., McCloskey, M., Hoxha, D., & Coccaro, E. F. (2010). How do depressed and healthy adults interpret nuanced facial expressions? *Journal of Abnormal Psychology*, 119, 804–810. <https://psycnet.apa.org/doi/10.1037/a0020234>
- Griffiths, P. & Ashwin, C. (2016). Accuracy in perceiving facial expressions of emotions in psychopathology. In J. A. Hall, M. S. Mast, & T. V. West (Eds.), *The social psychology of perceiving others accurately* (pp. 185–205). Cambridge University Press. <https://doi.org/10.1017/CBO9781316181959.009>
- Hall, J. A., Mast, M. S., & West, T. V. (2016). Accurate interpersonal perception: Many traditions, one topic. In J. A. Hall, M. S. Mast, & T. V. West (Eds.), *The social psychology of perceiving others accurately* (pp. 3–22). Cambridge University Press. <https://doi.org/10.1017/CBO9781316181959.001>

- Hayes, S. C., Hayes, L. J., & Reese, H. W. (1988). Finding the philosophical core: A review of Stephen C. Pepper's World Hypotheses. *Journal of Experimental Analysis of Behavior*, 50, 97-111. <https://doi.org/10.1901/jeab.1988.50-97>
- Hernández, S. O., Zarza-Alzugaray, F. J., & Casanova, O. (2018). Music performance anxiety. Substance use and career abandonment in Spanish music students. *Int. J. Music. Educ.* 36, 460–472. <https://doi.org/10.1177/0255761418763903>
- Hess, J. L., Miller, S., Higbee, S., Fore, G. A., & Wallace, J. (2021). Empathy and ethical becoming in biomedical engineering education: A mixed methods study of an animal tissue harvesting laboratory. *Australasian Journal of Engineering Education*, 26(1), 127–137. <https://doi.org/10.1080/22054952.2020.1796045>
- Hess, J. L., Strobel, J., Pan, R. C., & Wachter Morris, C. A. (2017). Insights from industry: A quantitative analysis of engineers' perceptions of empathy and care within their practice, *Eur. J. Eng. Educ.*, 42(6), pp. 1128–1153. <https://doi.org/10.1080/03043797.2016.1267717>
- Heylighen, A. & Dong, A. (2019). To empathise or not to empathise? Empathy and its limits in design, *Design Studies*, 65, 107-124. <https://doi.org/10.1016/j.destud.2019.10.007>
- Hinneken, C., Loeys, T., De Schryver, M., & Verhofstadt, L. L. (2018). The manageability of empathic (in)accuracy during couples' conflict: Relationship-protection or self-protection? *Motiv. Emot.*, 42(3), 403–418. <https://doi.org/10.1007/s11031-018-9689-z>
- Hirsch, P. M. & Levin, D. Z. (1999). Umbrella advocates versus validity police: A life-cycle model, *Organization Science*, 10(2), 199-212. <https://doi.org/10.1287/orsc.10.2.199>
- Huang, J., & Krumhansl, C. L. (2011). What does seeing the performer add? It depends on musical style, amount of stage behavior, and audience expertise. *Music. Sci.* 15, 343–364. <https://doi.org/10.1177/1029864911414172>
- Ickes, W. (2001). Measuring empathic accuracy. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal Sensitivity: Theory and Measurement* (pp. 219–241). U.S.A.: Lawrence Erlbaum.
- Ickes, W. (2016). Empathic accuracy: Judging thoughts and feelings. In J. A. Hall, M. S. Mast, & T. V. West (Eds.), *The social psychology of perceiving others accurately* (pp. 52–70). Cambridge University Press. <https://doi.org/10.1017/CBO9781316181959.003>
- Ickes, W., & Hodges, S. D. (2013). Empathic accuracy in close relationships. In J. A. Simpson & L. Campbell (Eds.), *Oxford library of psychology: The*

Oxford handbook of close relationships (pp. 348–373). U.S.A.: Oxford University Press.

- Ickes, W., Bissonnette, V., Garcia, S., & Stinson, L. L. (1990). Implementing and using the dyadic interaction paradigm. In C. Hendrick & M. S. Clark (Eds.), *Research methods in personality and social psychology* (pp. 16–44). Sage Publications, Inc.
- Johnson, D. G., Genco, N., Saunders, M. N., Williams, P., Seepersad, C. C., & Hölttä-Otto, K. (2014). An experimental investigation of the effectiveness of empathic experience design for innovative concept generation. *Journal of Mechanical Design*, 136(5), 051009. <https://doi.org/10.1115/1.4026951>
- Joormann, J., & Gotlib, I. H. (2006). Is this happiness I see? Biases in the identification of emotional facial expressions in depression and social phobia. *Journal of Abnormal Psychology*, 115, 705–714. <https://psycnet.apa.org/doi/10.1037/0021-843X.115.4.705>
- Juchniewicz, J. (2008). The influence of physical movement on the perception of musical performance. *Psychol. Music*, 36, 417–427. <https://doi.org/10.1177/0305735607086046>
- Kang, Y., Cakar, M.E., Shumaker, K., Brook O'Donnell, M., & Falk, E. B. (2022). Experience similarity, mindful awareness, and accurate interpersonal understanding. *Mindfulness*. <https://doi.org/10.1007/s12671-022-01859-x>
- Kankainen, A., Vaajakallio, K., Vesa Kantola, V., & Mattelmäki, T. (2012). Storytelling group – a co-design method for service design. *Behaviour and Information Technology*, 31(3), 221–230. <https://doi.org/10.1080/0144929X.2011.563794>
- Kenny, D. T. (2010). The role of negative emotions in performance anxiety. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of Music and Emotion: Theory, Research, and Applications* (pp. 425–452). U.K.: Oxford University Press.
- Kohler, C. G., Turner, T. H., Bilker, W. B., Brensinger, C. M., Siegel, S. J., Kanes, S. J., Gur, R. E., & Gur, R. C. (2003). Facial emotion recognition in schizophrenia: intensity effects and error pattern. *American Journal of Psychiatry*, 160, 1768–1774. <https://psycnet.apa.org/doi/10.1176/appi.ajp.160.10.1768>
- Köppen, E. & Meinel, C. (2015). Empathy via design thinking: Creation of sense and knowledge. In H. Plattner, C. Meinel, & L. Leifer (Eds.), *Design thinking research* (pp. 15–28). https://doi.org/10.1007/978-3-319-06823-7_2

- Kouprie, M. & Sleeswijk Visser, F. (2009). A framework for empathy in design: Stepping into and out of the user's life, *Journal of Engineering Design*, 20(5), 437-448. <https://doi.org/10.1080/09544820902875033>
- Kramer, J., Agogino, A. M. & Roschuni, C. (2016). Characterizing competencies for human-centered design. *Proceedings of the ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Volume 7: 28th International Conference on Design Theory and Methodology*. Charlotte, North Carolina, USA. August 21–24. V007T06A026. ASME. <https://doi.org/10.1115/DETC2016-60085>
- Kullman, K. (2016). Prototyping bodies: A post-phenomenology of wearable simulations. *Design Studies*, 47, 73–90. <https://doi.org/10.1016/j.des-tud.2016.08.004>
- Kwan, P. Y. (2016). *The effect of music performance anxiety, context, modality and observers' music expertise on judgment of musical performances* (Unpublished master thesis). University of Jyväskylä. Retrieved from: <https://jyx.jyu.fi/handle/123456789/50620>
- Ledet, D. A. (1981). *Oboe reed styles, theory and practice*. Indiana University Press.
- Leonard, D. & Rayport, J. F. (1997). Spark innovation through empathic design. *Harvard Business Review*, 75(6), 102–113.
- Leppänen, J. M., Niehaus, D. J., Koen, L., Du Toit, E., Schoeman, R., & Emsley, R. (2008). Deficits in facial affect recognition in unaffected siblings of schizophrenia patients: Evidence for a neurocognitive endophenotype. *Schizophrenia Research*, 99, 270–273. <https://doi.org/10.1016/j.schres.2007.11.003>
- Levenson, R. W. & Gottman, J. M. (1983). Marital interaction: Physiological linkage and affective exchange. *Journal of Personality and Social Psychology*, 45(3), 587–597. <https://psycnet.apa.org/doi/10.1037/0022-3514.45.3.587>
- Levenson, R. W. & Ruef, A. M. (1992). Empathy: A physiological substrate. *Journal of Personality and Social Psychology*, 63(2), 234–246. <https://psycnet.apa.org/doi/10.1037/0022-3514.63.2.234>
- MacDonald, G., & Leary, M. R. (2005). Why does social exclusion hurt? The relationship between social and physical pain. *Psychological Bulletin*, 131(2), 202–223. <https://doi.org/10.1037/0033-2909.131.2.202>
- Madsen, S. & Nielsen, L. (2010). Exploring persona-scenarios – using storytelling to create design ideas. In *Human Work Interaction Design: Usability in Social, Cultural and Organizational Contexts - Second IFIP WG 13.6 Conference, HWID 2009 Pune, India, October 7–8, 2009*.

- (ed. D. Kater, R. Orngreen, P. Yammiyavar & T. Clemmensen), pp. 57–66; Retrieved from https://link.springer.com/content/pdf/10.1007/978-3-642-11762-6_5.pdf
- Marangoni, C., Garcia, S., Ickes, W., & Teng, G. (1995). Empathic accuracy in a clinically relevant setting. *Journal of Personality and Social Psychology*, 68(5), 854–869. <https://doi.org/10.1037/0022-3514.68.5.854>
- Mattelmäki, T., Vaajakallio, K., & Koskinen, I. (2014). What happened to empathic design? *Design Issues*, 30(1), 67–77. Retrieved from <http://ilpo-koskinen.com/media/Empathicdesign.PDF>
- McDonagh, D., Woodcock, A., & Iqbal, S. (2018). *Experience + experience + experience = Empathy design for health*. Sheffield Hallam University.
- Medler, B. & Magerko, B. (2010). The implications of improvisational acting and role-playing on design methodologies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. Association for Computing Machinery, pp. 483–492. ACM; <https://doi.org/10.1145/1753326.1753398>
- Miaskiewicz, T. & Kozar, K. A. (2011). Personas and user-centered design: How can personas benefit product design processes? *Design Studies*, 32, 417–430. <https://doi.org/10.1016/j.destud.2011.03.003>
- Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: an index for assessing musical sophistication in the general population. *PLoS One* 9:e89642. <https://doi.org/10.1371/journal.pone.0089642>
- Neumann, D. L. & Westbury, H. R. (2011). The psychophysiological measurement of empathy. In D. J. Scapaletti (ed.), *Psychology of Empathy* (pp. 119–142). Retrieved from https://research-repository.griffith.edu.au/bitstream/handle/10072/44076/77247_1.pdf%3Bsequence=1
- Nunnally, J. C. (1967). *Psychometric Theory*. McGraw Hill.
- Nussek, M., & Wanderley, M. M. (2009). Music and motion—how music-related ancillary body movements contribute to the experience of music. *Music. Percept.*, 26, 335–353. <https://doi.org/10.1525/mp.2009.26.4.335>
- Oygür, I. (2018). The machineries of user knowledge production. *Design Studies*, 54, 23–49. <https://doi.org/10.1016/j.destud.2017.10.002>
- Pepper, S. C. (1942). *World hypotheses: A study in evidence*. U.S.A.: University of California Press.

- Pigliucci, M. (2017). *How to be a stoic: Using ancient philosophy to live a modern life*. U.S.A.: Basic Books.
- Rasool, C., Danielsson, H., & Jungert, T. (2012). Empathy among students in engineering programmes, *Eur. J. Eng. Educ.*, 37(5), 427–435. <https://doi.org/10.1080/03043797.2012.708720>
- Raviselvam, S., Hölttä-Otto, K., & Wood, K. L. (2016). User extreme conditions to enhance designer empathy and creativity: Applications using visual impairment. In *International Design Engineering Technical Conferences, August 21–24. American Society of Mechanical Engineers*. <https://doi.org/10.1080/21650349.2021.2024093>
- Ruben, M. A. (2016). Interpersonal accuracy in the clinical setting. In J. A. Hall, M. S. Mast, & T. V. West (Eds.), *The social psychology of perceiving others accurately* (pp. 287–308). Cambridge University Press. <https://psycnet.apa.org/doi/10.1017/CBO9781316181959.014>
- Rubin, M. (2017). Do *p* values lose their meaning in exploratory analyses? It depends how you define the familywise error rate. *Rev. Gen. Psychol.*, 21, 269–275. <https://doi.org/10.1037/gpr0000123>
- Ryan, R. M. & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. U.S.A.: The Guilford Press.
- Sanders, E. B. N. & Stappers, P. J. (2014). Probes, toolkits and prototypes: Three approaches to making in codesigning. *CoDesign*, 10(1), 5–14. <http://dx.doi.org/10.1080/15710882.2014.888183>
- Schaller, M., Park, J. H., & Kenrick, D. T. (2007). Human evolution and social cognition. In R. I. M. Dunbar & L. Barrett (Eds.), *Oxford handbook of evolutionary psychology* (pp. 491–504). U.S.A.: Oxford University Press.
- Schmid Mast, M., & Hall, J. A. (2018). The impact of interpersonal accuracy on behavioral outcomes. *Current Directions in Psychological Science*, 27(5), 309–314. <https://doi.org/10.1177/0963721418758437>
- Schmid Mast, M., & Latu, I. (2016). Interpersonal accuracy in relation to the workplace, leadership, and hierarchy. In J. A. Hall, M. S. Mast, & T. V. West (Eds.), *The social psychology of perceiving others accurately* (pp. 270–286). Cambridge University Press. <https://psycnet.apa.org/doi/10.1017/CBO9781316181959.013>
- Schmid, P. C. (2016). Situational influences on interpersonal accuracy. In J. A. Hall, M. S. Mast, & T. V. West (Eds.), *The social psychology of perceiving others accurately* (pp. 230–252). Cambridge University Press. <https://doi.org/10.1017/CBO9781316181959.011>

- Schober, M. F., & Spiro, N. (2014). Jazz improvisers' shared understanding: A case study. *Frontiers in Psychology*, 5(AUG), 1–21. <https://doi.org/10.3389/fpsyg.2014.00808>
- Schober, M. F., & Spiro, N. (2016). Listeners' and performers' shared understanding of jazz improvisations. *Frontiers in Psychology*, 7(NOV), 1–20. <https://doi.org/10.3389/fpsyg.2016.01629>
- Scriven, M. (1969). Psychology without a paradigm. In L. Breger (Ed.), *Clinical-cognitive psychology: Models and integrations* (pp. 9–24). New Jersey: Prentice-Hall.
- Sened, H., Lavidor, M., Lazarus, G., Bar-Kalifa, E., Rafaeli, E., & Ickes, W. (2017). Empathic accuracy and relationship satisfaction: A meta-analytic review. *Journal of Family Psychology*, 31(6), 742–752. <https://doi.org/10.1037/fam0000320>
- Shamay-Tsoory, S. G. (2009). Empathic processing: Its cognitive and affective dimensions and neuroanatomical basis. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 215–232). MIT Press. <https://doi.org/10.7551/mitpress/9780262012973.003.0017>
- Shamay-Tsoory, S. G. (2011). The neural bases for empathy. *The Neuroscientist* 17(1), 18–24. <https://doi.org/10.1177/1073858410379268>
- Smeenk, W., Sturm, J., & Eggen, B. (2017). Empathic handover: How would you feel? Handing over dementia experiences and feelings in empathic co-design. *CoDesign*, 14, 259–274. <https://doi.org/10.1080/15710882.2017.1301960>
- Smeenk, W., Sturm, J., & Eggen, B. (2019a). A comparison of existing frameworks leading to an empathic formation compass for co-design. *International Journal of Design*, 13(3), 53–68. <http://www.ijdesign.org/index.php/IJDesign/article/view/3406>.
- Smeenk, W., Sturm, J., & Eggen, B. (2019b). A comparison of existing frameworks leading to an empathic formation compass for co-design, *International Journal of Design*, 13(3), 53–68. <http://www.ijdesign.org/index.php/IJDesign/article/view/3406>
- Soto, J. A. & Levenson, R. W. (2009). Emotion recognition across cultures: The influence of ethnicity on empathic accuracy and physiological linkage, *Emotion*, 9(6), 874–884. <https://psycnet.apa.org/doi/10.1037/a0017399>
- Stanley, M., Brooker, R., & Gilbert, R. (2002). Examiner perceptions of using criteria in music performance assessment. *Res. Stud. Music Educ.*, 18, 46–56. <https://doi.org/10.1177/1321103X020180010601>

- Stinson, L. & Ickes, W. (1992). Empathic accuracy in the interactions of male friends versus male strangers. *Journal of Personality and Social Psychology*, 62(5), 787–797. <https://doi.org/10.1037/0022-3514.62.5.787>
- Studer, R.K., Danuser, B., Wild, P., Hildebrandt, H., & Gomez, P. (2014). Psychophysiological activation during preparation, performance, and recovery in high- and low-anxious music students. *Appl Psychophysiol Biofeedback*, 39, 45–57. <https://doi.org/10.1007/s10484-014-9240-2>
- Stueber, K. (2019) Empathy. In *The Stanford Encyclopedia of Philosophy* (ed. E. N. Zalta) [Online], online document, Accessed July 8, 2022. Retrieved from <https://plato.stanford.edu/archives/fall2019/entries/empathy/>
- Surma-aho, A. & Hölttä-Otto, K. (2022). Conceptualization and operationalization of empathy in design research. *Design Studies*, 78, 101075. <https://doi.org/10.1016/j.destud.2021.101075>
- Surma-aho, A., Björklund, T., & Hölttä-Otto, K. (2018). Assessing the development of empathy and innovation attitudes in a project-based Design Thinking course. *2018 ASEE annual conference & exposition proceedings. ASEE annual conference & exposition proceedings, American Society for Engineering Education, ASEE Annual Conference, Salt Lake City, Utah, United States, June 23*. <https://peer.asee.org/29826>
- Takahashi, I., Oki, M., Bourreau, B., Kitahara, I., & Suzuki, K. (2018). An empathic design approach to an augmented gymnasium in a special needs school setting. *International Journal of Design*, 12(3), 111–125. Retrieved from <http://ijdesign.org/index.php/IJDesign/article/view/2949>
- Taylor, A., & Wasley, D. (2004). Physical fitness. In A. Williamon (Ed.), *Musical Excellence: Strategies and Techniques to Enhance Performance* (pp. 163-178). New York: Oxford University Press.
- Thompson, M. R., & Luck, G. (2012). Exploring relationships between pianists' body movements, their expressive intentions, and structural elements of the music. *Musicae Scientiae*, 16(1), 19–40. <https://doi.org/10.1177/1029864911423457>
- Thompson, S. (2006). Audience responses to a live orchestral concert. *Music Sci.*, 10, 215–244. <https://doi.org/10.1177/102986490601000204>
- Thompson, S. C. (1979). The effect of the reed resonance on woodwind tone production. *The Journal of the Acoustical Society of America*, 66, 1299–1307. <https://doi.org/10.1121/1.383448>
- Thompson, W. F. & Robitaille, B. (1992). Can composers express emotions through music? *Empirical Studies of the Arts*, 10(1), 79-89. <https://doi.org/10.2190/NBNY-AKDK-GW58-MTEL>

- Tsay, C. J. (2013). Sight over sound in the judgment of music performance. *Proc. Natl. Acad. Sci. U. S. A.*, *110*, 14580–14585. <https://doi.org/10.1073/pnas.1221454110>
- Tsay, C. J. (2014). The vision heuristic: judging music ensembles by sight alone. *Organ. Behav. Hum. Decis. Process.*, *124*, 24–33. <https://doi.org/10.1016/j.obhdp.2013.10.003>
- Van Zijl, A. G., Toiviainen, P., Lartillot, O., & Luck, G. (2014). The sound of emotion. *Music Perception: An Interdisciplinary Journal*, *32*(1), 33–50. <https://doi.org/10.1525/mp.2014.32.1.33>
- Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition*, *101*, 80–113. <https://doi.org/10.1016/j.cognition.2005.09.003>
- Vines, B. W., Krumhansl, C. L., Wanderley, M. M., Dalca, I. M., & Levitin, D. J. (2011). Music to my eyes: cross-modal interactions in the perception of emotions in musical performance. *Cognition*, *118*, 157–170. <https://doi.org/10.1016/j.cognition.2010.11.010>
- Vuoskoski, J. K., Gatti, E., Spence, C., & Clarke, E. F. (2016). Do visual cues intensify the emotional responses evoked by musical performance? A psychophysiological investigation. *Psychomusicol. Music Mind Brain*, *26*, 179–188. <https://psycnet.apa.org/doi/10.1037/pmu0000142>
- Vuoskoski, J. K., Thompson, M. R., Clarke, E. F., & Spence, C. (2014). Cross-modal interactions in the perception of expressivity in musical performance. *Atten. Percept. Psychophysiol.*, *76*, 591–604. Retrieved from <https://link.springer.com/content/pdf/10.3758/s13414-013-0582-2.pdf>
- Waddell, G., & Williamon, A. (2017). Eye of the beholder: stage entrance behavior and facial expression affect continuous quality ratings in music performance. *Front. Psychol.*, *8*:513. <https://doi.org/10.3389/fpsyg.2017.00513>
- Walker, A. (1994). *Franz Liszt: The Weimar years 1848-1861* (Vol. II). U.S.A.: Faber and Faber.
- Walther, J., Miller, S. E., & Sochacka, N. W. (2017). A model of empathy in engineering as a core skill, practice orientation, and professional way of being. *Journal of Engineering Education*, *106*(1), 123–148. <https://doi.org/10.1002/jee.20159>
- Wanderley, M. M., Vines, B. W., Middleton, N., McKay, C., & Hatch, W. (2005). The musical significance of clarinetists' ancillary gestures: An exploration of the field. *Journal of New Music Research*, *34*(1), 97–113. <https://doi.org/10.1080/09298210500124208>

- Wapnick, J., Ryan, C., Lacaille, N., & Darrow, A. A. (2004). Effects of selected variables on musicians' ratings of high-level piano performances. *Int. J. Music. Educ.*, 22, 7–20. <https://doi.org/10.1177/0255761404042371>
- West, R. (2004). Drugs and musical performance. In A. Williamon (Ed.), *Musical excellence: Strategies and techniques to enhance performance* (pp. 271-290). New York: Oxford University Press.
- Yoshie, M., Kudo, K., & Ohtsuki, T. (2008). Effects of psychological stress on state anxiety, electromyographic activity, and arpeggio performance in pianists. *Med. Probl. Perform. Art.*, 23, 120–132. <https://doi.org/10.21091/mppa.2008.3024>
- Yoshie, M., Kudo, K., Murakoshi, T., & Ohtsuki, T. (2009). Music performance anxiety in skilled pianists: effects of social-evaluative performance situation on subjective, autonomic, and electromyographic reactions. *Exp. Brain Res.*, 199, 117–126. Retrieved from <https://link.springer.com/content/pdf/10.1007/s00221-009-1979-y.pdf>
- Zaki, J. & Ochsner, K. (2012). The neuroscience of empathy: Progress, pitfalls and promise. *Nature Neuroscience* 15, 675–680. <https://doi.org/10.1038/nn.3085>
- Zaki, J., Weber, J., Bolger, N. & Ochsner, K. N. (2009). The neural basis of empathic accuracy. *Proceedings of the National Academy of Sciences of the United States of America*, 106(27), 11382–11387. <https://doi.org/10.1073/pnas.0902666106>

Human cooperation has been key in our evolutionary history. This thesis explores a key social skill: interpersonal accuracy (IA) or the correct understanding of other's mental states such as thoughts and feelings. We investigate how IA is affected by a person's background and perception. To do so, we apply methods from social psychology adapted to the contexts of design and music performance. Our results indicate that we are rather inaccurate when inferring other's mental states. Readers interested in human cooperation may benefit from our methodological adaptations, as well as some suggestions that could improve IA between individuals.



ISBN 978-952-64-1156-9 (printed)

ISBN 978-952-64-1157-6 (pdf)

ISSN 1799-4934 (printed)

ISSN 1799-4942 (pdf)

Aalto University

School of Science

Department of Neuroscience and Biomedical Engineering

www.aalto.fi

**BUSINESS +
ECONOMY**

**ART +
DESIGN +
ARCHITECTURE**

**SCIENCE +
TECHNOLOGY**

CROSSOVER

**DOCTORAL
THESES**