
PROFICIENCY-AWARE SYSTEMS

DESIGNING FOR USER SKILL AND EXPERTISE

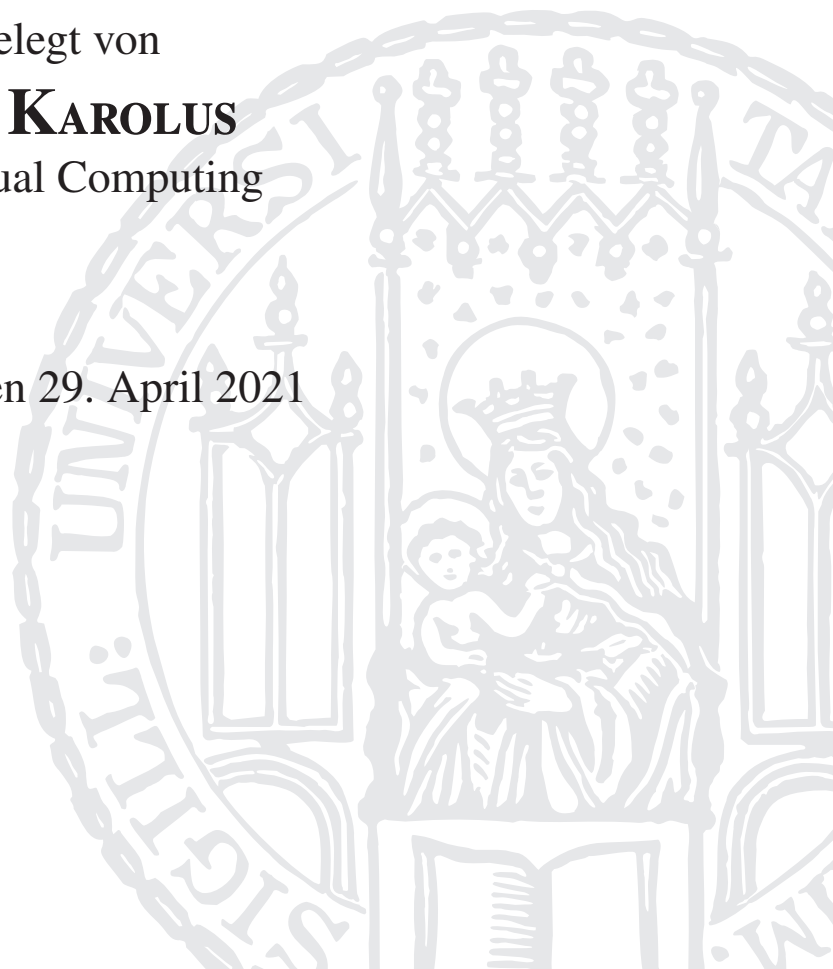
DISSERTATION

an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig-Maximilians-Universität München

vorgelegt von

JAKOB KAROLUS
M.Sc. Visual Computing

München, den 29. April 2021



Erstgutachter: Prof. Dr. Albrecht Schmidt
Zweitgutachter: Prof. Dr. Paweł W. Woźniak
Drittgutachter: Prof. Dr. Florian Michahelles

Tag der mündlichen Prüfung: 25. Juni 2021



ABSTRACT

In an increasingly digital world, technological developments such as data-driven algorithms and context-aware applications create opportunities for novel human-computer interaction (HCI). We argue that these systems have the latent potential to stimulate users and encourage personal growth. However, users increasingly rely on the intelligence of interactive systems. Thus, it remains a challenge to design for *proficiency awareness*, essentially demanding increased user attention whilst preserving user engagement. Designing and implementing systems that allow users to become aware of their own proficiency and encourage them to recognize learning benefits is the primary goal of this research.

In this thesis, we introduce the concept of *proficiency-aware systems* as one solution. In our definition, proficiency-aware systems use estimates of the user's proficiency to tailor the interaction in a domain and facilitate a reflective understanding for this proficiency. We envision that proficiency-aware systems leverage collected data for learning benefit. Here, we see self-reflection as a key for users to become aware of necessary efforts to advance their proficiency.

A key challenge for proficiency-aware systems is the fact that users often have a different self-perception of their proficiency. The benefits of personal growth and advancing one's repertoire might not necessarily be apparent to users, alienating them, and possibly leading to abandoning the system. To tackle this challenge, this work does not rely on learning strategies but rather focuses on the capabilities of interactive systems to provide users with the necessary means to reflect on their proficiency, such as showing calculated text difficulty to a newspaper editor or visualizing muscle activity to a passionate sportsperson.

We first elaborate on how proficiency can be detected and quantified in the context of interactive systems using physiological sensing technologies. Through developing interaction scenarios, we demonstrate the feasibility of gaze- and electromyography-based proficiency-aware systems by utilizing machine learning algorithms that can estimate users' proficiency levels for stationary vision-dominant tasks (reading, information intake) and dynamic manual tasks (playing instruments, fitness exercises). Secondly, we show how to facilitate proficiency awareness for users, including design challenges on when and how to communicate proficiency. We complement this second part by highlighting the necessity of toolkits for sensing modalities to enable the implementation of proficiency-aware systems for a wide audience.

In this thesis, we contribute a definition of proficiency-aware systems, which we illustrate by designing and implementing interactive systems. We derive technical requirements for real-time, objective proficiency assessment and identify design qualities of communicating proficiency through user reflection. We summarize our findings in a set of design and engineering guidelines for proficiency awareness in interactive systems, highlighting that proficiency feedback makes performance interpretable for the user.

ZUSAMMENFASSUNG

In einer zunehmend digitalen Welt schaffen technologische Entwicklungen - wie datengetriebene Algorithmen und kontextabhängige Anwendungen - neuartige Interaktionsmöglichkeiten mit digitalen Geräten. Jedoch verlassen sich Nutzer oftmals auf die Intelligenz dieser Systeme, ohne dabei selbst auf eine persönliche Weiterentwicklung hinzuwirken. Wird ein solches Vorgehen angestrebt, verlangt dies seitens der Anwender eine erhöhte Aufmerksamkeit. Es ist daher herausfordernd, ein entsprechendes Design für *Kompetenzbewusstsein* (*Proficiency Awareness*) zu etablieren. Das primäre Ziel dieser Arbeit ist es, eine Methodik für das Design und die Implementierung von interaktiven Systemen aufzustellen, die Nutzer dabei unterstützen über ihre eigene Kompetenz zu reflektieren, um dadurch Lerneffekte implizit wahrnehmen können.

Diese Arbeit stellt ein Konzept für *fähigkeitsbewusste Systeme* (*proficiency-aware systems*) vor, welche die Fähigkeiten von Nutzern abschätzen, die Interaktion entsprechend anpassen sowie das Bewusstsein der Nutzer über deren Fähigkeiten fördern. Hierzu sollten die Systeme gesammelte Daten von Nutzern einsetzen, um Lerneffekte sichtbar zu machen. Die Möglichkeit der Anwender zur Selbstreflexion ist hierbei als entscheidend anzusehen, um als Motivation zur Verbesserung der eigenen Fähigkeiten zu dienen.

Eine zentrale Herausforderung solcher Systeme ist die Tatsache, dass Nutzer - im Vergleich zur Abschätzung des Systems - oft eine divergierende Selbstwahrnehmung ihrer Kompetenz haben. Im ersten Moment sind daher die Vorteile einer persönlichen Weiterentwicklung nicht unbedingt ersichtlich. Daher baut diese Forschungsarbeit nicht darauf auf, Nutzer über vorgegebene Lernstrategien zu unterrichten, sondern sie bedient sich der Möglichkeiten interaktiver Systeme, die Anwendern die notwendigen Hilfsmittel zur Verfügung stellen, damit diese selbst über ihre Fähigkeiten reflektieren können. Einem Zeitungseditor könnte beispielsweise die aktuelle Textschwierigkeit angezeigt werden, während einem passionierten Sportler dessen Muskelaktivität veranschaulicht wird.

Zunächst wird herausgearbeitet, wie sich die Fähigkeiten der Nutzer mittels physiologischer Sensortechnologien erkennen und quantifizieren lassen. Die Evaluation von Interaktionsszenarien demonstriert die Umsetzbarkeit fähigkeitsbewusster Systeme, basierend auf der Analyse von Blickbewegungen und Muskelaktivität. Hierbei kommen Algorithmen des maschinellen Lernens zum Einsatz, die das Leistungsniveau der Anwender für verschiedene Tätigkeiten berechnen. Im Besonderen analysieren wir stationäre Aktivitäten, die hauptsächlich den Sehsinn ansprechen (Lesen, Aufnahme von Informationen), sowie dynamische Betätigungen, die die Motorik der Nutzer fordern (Spielen von Instrumenten, Fitnessübungen). Der zweite Teil zeigt auf, wie Systeme das Bewusstsein der Anwender für deren eigene Fähigkeiten fördern können, einschließlich der Designherausforderungen, wann und wie das System erkannte Fähigkeiten kommunizieren sollte. Abschließend wird die Notwendigkeit von Toolkits für Sensortechnologien hervorgehoben, um die Implementierung derartiger Systeme für ein breites Publikum zu ermöglichen.

Die Forschungsarbeit beinhaltet eine Definition für fähigkeitsbewusste Systeme und veranschaulicht dieses Konzept durch den Entwurf und die Implementierung interaktiver Systeme. Ferner werden technische Anforderungen objektiver Echtzeitabschätzung von Nutzerfähigkeiten erforscht und Designqualitäten für die Kommunikation dieser Abschätzungen mittels Selbstreflexion identifiziert. Zusammengefasst sind die Erkenntnisse in einer Reihe von Design- und Entwicklungsrichtlinien für derartige Systeme. Insbesondere die Kommunikation, der vom System erkannten Kompetenz, hilft Anwendern, die eigene Leistung zu interpretieren.

PREFACE

The work presented in this thesis is the result of my research that I conducted at the University of Stuttgart and the LMU Munich. Within the context of several research projects, colleagues and students from my home universities and external research institutes have shaped this thesis through collaborative publications. These publications are a core part of this thesis, and contributing authors are clearly stated at the beginning of each respective chapter. When applicable, a reference to the publication is provided. Further, a detailed list of my contribution to each of these publications is provided in the appendix (Contributing Publications). To emphasize the collaborative nature of this research, I use the scientific plural throughout this thesis.

ACKNOWLEDGEMENTS

I am very grateful for being given the opportunity to conduct this thesis in such an inspiring environment. These past years are filled with great adventures, philosophical debates, and numerous collaborations with like-minded people that share the same enthusiasm about research as I do. Many of whom I became friends with over these past years.

—

First and foremost, I would like to thank my supervisor Albrecht Schmidt for giving me the possibility to freely pursue my research ideas, but never falling short of delivering provoking thoughts to contemplate on. I also would like to thank Paweł W. Woźniak for your contagious enthusiasm and for always providing me with solid council during this research adventure. Many thanks also go to Florian Michahelles for his time to serve as an external reviewer for this thesis. A special thank you goes to Tobias Grosse-Puppendahl for starting all of this.

I was fortunate to meet amazing colleagues that share the same vision and passion for research as I do. I am grateful for your friendship and many unforgettable moments, great journeys, and most importantly sharing all the ups and downs in life. Thank you, Florian, Sebastian, Thomas, Pascal, Paweł, Jasmin, Miriam, Matthias, Francisco, Dominik, Huy, Rufat. Further thanks go to all the people in Stuttgart and Munich who brightened up my days during these years, a list far too long to recite here. A particular thank you goes out to Anja, for not only always lending a helping hand but also always having an open ear.

This work would not have been possible without the excellent students that I had the pleasure of supervising. For me, the real joy in teaching is when I am able to allow others to follow their goals. Thank you for giving it your all.

Special thanks go to my friends from outside of research, keeping me level and connected to the real world. Knowing that you will always have my back gives me massive support. Thank you for supporting me through my life journey.

Last but certainly not least, I want to thank and honor my family for teaching me what is important in life. Thank you for your endless support, your love, and for always believing in me.

Danke.

TABLE OF CONTENTS

INTRODUCTION AND FOUNDATIONS	3
1 Introduction	3
1.1 Research Methodology	5
1.1.1 Research Questions	6
1.1.2 Research Context	7
1.2 Contributions to Human-Computer Interaction	8
1.3 Thesis Outline	9
1.3.1 Introduction and Foundations	10
1.3.2 Part I: Detecting and Quantifying Proficiency	10
1.3.3 Part II: Facilitating Proficiency Awareness	10
1.3.4 Conclusion and Future Work	11
2 Towards Proficiency-Aware Systems	13
2.1 A Definition of Proficiency-Aware Systems	14
2.1.1 Proficiency as an Aggregated Construct	15
2.2 The System Side	17
2.2.1 Adaptive Systems	17
2.2.2 From Calm to Engaging Ubiquitous Computing	18
2.2.3 Context-Aware Computing Applications	18
2.2.4 Physiological Sensing	20
2.3 The User Side	21
2.3.1 Reflection and Feedback	21
2.3.2 Learning and Skill Acquisition	21
2.3.3 Explainability and Interpretability	22
2.4 Challenges and Dimensions of Proficiency-Aware Systems	22
2.4.1 Interaction Duration and Frequency	23
2.4.2 Adaptation Control and Agency	23
2.4.3 Task Complexity And Domain	24
2.4.4 Feedback Types and Communicating Proficiency	24

I DETECTING AND QUANTIFYING PROFICIENCY **25**

3 Proficiency From Gaze **27**

3.1	Related Research	28
3.2	Use Case I: Robust Gaze Features for Enabling Language Proficiency Awareness	30
3.2.1	Methodology	31
3.2.2	Evaluation	33
3.2.3	Discussion	46
3.2.4	Conclusion	48
3.3	Use Case II: Assessment of Programming Proficiency using Gaze Properties	49
3.3.1	Methodology	49
3.3.2	Results	51
3.3.3	Conclusion	52
3.4	Use Case III: Gaze as a Predictor for the Proficiency of Piano Players	53
3.4.1	Methodology	53
3.4.2	Results and Conclusion	54
3.5	Chapter Summary	56

4 Proficiency From Electromyography **59**

4.1	Related Research and Background	61
4.2	Use Case I: EMGuitar - Recognizing Fine-Grained Motor Tasks	64
4.2.1	Methodology	65
4.2.2	Feasibility Study - Recognizing Guitar Chords via EMG	66
4.2.3	Main Study - Evaluating a Guitar Tutoring System	73
4.2.4	Requirements and Limitations	80
4.2.5	Conclusion	81
4.3	Use Case II: FitBack - Recognizing Exercise Form	82
4.3.1	Methodology	82
4.3.2	FitBack	83
4.3.3	Lab Study	85
4.3.4	Conclusion	89
4.4	Chapter Summary	89

II	FACILITATING PROFICIENCY AWARENESS	91
5	Designing Proficiency Awareness	93
5.1	Use Case I: Facilitating Bodily Insights Using Electromyography-Based Biofeedback during Physical Activity	95
5.1.1	Related Research	95
5.1.2	Methodology	97
5.1.3	Study I - Experimental Investigation of Biofeedback Modalities	99
5.1.4	Study II - Interviews With Coaches	106
5.1.5	Discussion	111
5.1.6	Conclusion	113
5.2	Use Case II: Leveraging Reflective Feedback in Text Production	113
5.2.1	Related Research	114
5.2.2	Methodology	116
5.2.3	Study I - Temporal Granularity of Proficiency Feedback	119
5.2.4	Study II - Designing Better Proficiency Feedback Through Gamification	125
5.2.5	Discussion	132
5.2.6	Conclusion	135
5.3	Use Case III: EMPiano - A Take on Artificial Proficiency	136
5.3.1	Methodology	136
5.3.2	EMPiano	137
5.3.3	Evaluation	138
5.3.4	Results	139
5.3.5	Conclusion	141
5.4	Chapter Summary	142
6	A Toolkit Approach to Implementing Proficiency-Aware Systems	145
6.1	Related Research	146
6.1.1	Toolkits	146
6.1.2	Pipeline Architectures	147
6.1.3	Physical Computing Toolkits	147
6.2	EMBody - A Data-Centric Toolkit for EMG-Based Interface Prototyping and Experimentation	148
6.2.1	Requirements	149
6.2.2	An Exploration-Centered Workflow	151
6.2.3	Hardware	154
6.2.4	Software	156

6.2.5	Technical Limitations	159
6.2.6	Target Audience	160
6.2.7	Formative Evaluation	163
6.2.8	Summative Evaluation	168
6.2.9	Discussion	171
6.2.10	Conclusion and Outlook	173
6.3	Chapter Summary	174
 CONCLUSION AND FUTURE WORK		179
7	Conclusion and Future Work	179
7.1	Designing Proficiency-Aware Systems	180
7.2	Engineering Proficiency-Aware Systems	182
7.3	Limitations	185
7.4	Future Work	186
7.5	Final Remarks	187
 LISTINGS AND BIBLIOGRAPHY		191
List of Acronyms		191
List of Figures		193
List of Tables		197
Bibliography		199
 APPENDIX		239
Contributing Publications		239

INTRODUCTION AND FOUNDATIONS

Chapter 1

Introduction

The ubiquity of interactive systems has substantially changed how we accomplish everyday tasks. They perpetuate our daily routine and support us in performing necessary tasks. Often, this change is displayed as an increase in comfort and efficiency. Machines take care of laborious tasks for us, and users increasingly depend on the *intelligence* [119] of interactive systems. There is no need for users to burden themselves with trifles, allowing us, in theory, to take a more holistic approach to task solving.

This technological progress opens up the opportunity for users to acquire novel skills that would otherwise be too cognitively demanding to develop. Here, interactive systems have the potential to enable users to advance their own proficiency by guaranteeing sufficient mastery of basic skills. Imagine an interactive system that supports a painter in perfecting their stroke work by assessing its fidelity. Such a system is unlikely to help them become a more creative painter but allows them to focus on the creative aspects rather than worry about their technique. Similarly, a translation aid should encourage the user to further their own language proficiency by calling their attention towards the smaller details of language that make up different cultures, thus leading to a better overall understanding of the foreign language.

However, this approach is challenging as users often have a different self-perception of their own proficiency, especially if proficiency is low [172]. This skewed reality can potentially be dangerous, e.g., "relying on Dr. Google" [2] and makes it all the more challenging for systems to encourage personal growth in users. While designers of interactive systems only have the best intentions in mind — providing pristine assistance to users — we argue that highly customized and *perfect* systems are seldom employed for the users' learning benefits. Further, the benefits of personal growth and advancing one's repertoire might not be apparent to users, especially if there are already systems that can take care of tedious tasks. This is unfortunate, as the collected user information could prove valuable in supporting the user to reflect on their own proficiency.

For example, smartphones give us access to an egocentric map that adapts to our current environment at any given time, dramatically boosting our ability to navigate and orient ourselves

in foreign areas. While this "resurrection" [155] of maps provides new opportunities for proficient users to develop their skills, less proficient users gladly accept the simplification provided by GPS-enabled smartphones. For this user group, there is just no need anymore to go through the painstaking process of acquiring this skill. In fact, the majority of young adults in the UK under the age of 38 would struggle to orient themselves without their mobile phone [129]. In response, there have been calls to teach basic navigation in schools [42], highlighting that less proficient users are especially prone to so-called *de-skilling*.

Consequently, we need to rethink how interactive systems can encourage users to advance their own proficiency. The challenge here lies in the fact that the system would need to convince users that increasing their proficiency will pay off in the long run; a fine line to walk between success and possible abandonment by the user. To address this challenge, this thesis proposes the concept of *proficiency-aware systems* as one solution. By incorporating proficiency awareness into their design, proficiency-aware systems are not only able to adapt interaction based on the user's proficiency but also allow users to reflect on their proficiency. We highlight the challenge of designing for proficiency by (1) investigating technical aspects to estimate user skill and expertise, and (2) providing guidelines on how to operationalize proficiency in interactive systems. Ultimately, the goal is to provide users with the necessary means to reflect on their own proficiency. In this aspect, proficiency-aware systems are unique compared to other context-aware systems in that they may systematically withhold better user support to encourage personal growth. It is thus essential that users become aware of their own proficiency to realize the benefits of the additional effort [278].

In this thesis, we define *proficiency awareness* as an interactive system's capability to identify and quantify a suitable proxy for the user's proficiency in a given task. Often, proficiency assessments involve lengthy tests that are not suitable for deployment in interactive systems. Consequently, this work makes use of ubiquitous sensing technologies to infer the user's proficiency. Classifying user states — and proficiency — leveraging physiological sensing is objective and allows real-time interaction with the system. Research has already shown that physiological markers are powerful indicators of user proficiency for different modalities [167, 338, 234].

Further, we identify design constraints of proficiency awareness. Users might easily get annoyed by a system that purposefully does not support them as well as it potentially could. Thus, it is vital that users understand this behavior and the system properly communicates its intent. Doing so allows users to become aware of their proficiency and ultimately see the benefits of investing in its advancement. As such, proficiency awareness is not just an additional dimension of input for interactive systems but is actively governed by the proficiency-aware system. A key goal of this thesis is to develop systems that allow for engaging user experiences [250], whilst users still have the option to identify necessary investments for their own personal growth. Here, we have identified the opportunity for reflection and properly curated feedback on estimated proficiency as essential. Allowing users to interpret their proficiency enables them to reflect on their skills and knowledge [278].

In the course of this research, we first identify suitable sensing modalities in detecting and quantifying proficiency. We focus on gaze- and electromyography-related metrics concerning their applicability for proficiency-aware systems. Gaze is a ubiquitous modality and offers rich insights into user perception. Electromyography is the measurement of muscle activity and allows distinctive insights into how users perform manual tasks. Here, we provide machine learning algorithms that can estimate users' proficiency levels in various domains, including stationary vision-dominant tasks (reading, information intake) and dynamic manual tasks (playing instruments, fitness exercises) involving body movements. Selected datasets and algorithms are publicly available. A set of engineering guidelines for proficiency-aware systems serves as an entry point for future research. Secondly, we highlight design constraints and opportunities of proficiency-aware systems through identifying how to design for proficiency awareness in interactive systems. We conclude with a set of design guidelines for future systems. This is complemented by our open-source toolkit EMBody, that we contribute to facilitate the implementation and design of proficiency-aware systems for a wider audience.

1.1 Research Methodology

This thesis presents the concept of proficiency-aware systems; interactive systems that adapt based on the user's estimated proficiency. We envision that proficiency-aware systems can provide engaging experiences for users, which has not yet been a prominent design goal for interactive systems that leverage context awareness. Consequently, we base this investigation on two main research pillars: **Detecting and Quantifying Proficiency** (Part I) and **Facilitating Proficiency Awareness** (Part II). This section provides a short introduction to the research methodology of both parts and how they are informed by existing research. A more detailed analysis is later provided in Chapter 2. Our research is based on user-centered design methods [177] and corresponding evaluations [80], offering empirical and artifact contributions through a number of use cases (Chapters 3 to 6).

Proficiency in Interactive Systems

We position this research within the area of adaptive systems, heavily drawing from established paradigms for user adaptation. This includes technical aspects as well as design considerations. Thereof results our first main research pillar for this thesis: **Detecting and Quantifying Proficiency** (Part I). During the course of this work, we identified new challenges and opportunities for interactive systems to detect and quantify proficiency.

For the purpose of this thesis, we divided typical user interaction into two classes: tasks that are predominantly governed by (1) the cognitive ability of users and (2) physical tasks, i.e., tasks that require users to perform body movements. For cognitive tasks, choices of sensing modalities include – among others – electroencephalography (EEG) [167], functional near-infrared spectroscopy (fNIRS) [338] and gaze [68]. We opted for gaze in our research

as setup costs are minimal, it is unobtrusive and provides sufficient insights into cognitive processes.

For physical tasks, we opted to employ electromyography (EMG), measuring the electric field during muscular activity. EMG provides superior distinctiveness and is privacy-sensitive in comparison to vision-based approaches [18] when detecting finer motor tasks. It further allowed us, compared to inertial measurement units (IMUs) [279], to provide actual insights of the muscle activity to users.

We acknowledge that this separation is theoretical, and most user interaction includes aspects from both cognitive and physical tasks. While it is impossible to apply a clear separation, this work also showcases that it is not necessary to do so. Proficiency-aware systems are effective even when focusing on one class. Often, it is even beneficial to focus on physical or cognitive aspects only, allowing precise feedback for the user. The choice of relevant proxies for proficiency estimation (cf. Chapter 2) is far more critical.

A User-Driven Approach to Data Interpretation

The second main pillar of the research in this thesis is: **Facilitating Proficiency Awareness** (Part II). Here, we investigate how proficiency-aware systems can facilitate the user’s understanding of their own proficiency. We envision that systems that foster personal growth provide engaging experiences for users [250]. In our work, we exploit user reflection and encourage awareness to enable user-driven interpretation of their own data. In other words, a proficiency-aware system communicates feedback on the collected data, but the users themselves need to make the final interpretation. In two use cases, we show that this strategy works for bodily signals (EMG) as well as for everyday computer work in a text production task.

1.1.1 Research Questions

To further structure our investigation, we introduce two research questions for each of the main pillars. **RQ1** and **RQ2** comprise the first part on **Detecting and Quantifying Proficiency**, while **RQ3** and **RQ4** provide guidance for **Facilitating Proficiency Awareness**. Each research question is addressed in a chapter in this thesis as outlined in Table 1.1.

In the first pillar, **RQ1** guides our investigation of gaze as a modality to estimate a user’s proficiency, including different use case scenarios and interaction forms. These include ad hoc systems, such as language proficiency detection, and proficiency-aware systems with prolonged user interaction, such as detection of programming proficiency of students and the proficiency of piano players. Analogously, **RQ2** addresses technical challenges and opportunities for EMG-based proficiency estimation of motor tasks. Here, we focus on the technology’s capabilities in detecting fine motor tasks in a guitar tutoring and fitness exercise scenario.

RQ	Research Questions	Chapter
Detecting and Quantifying Proficiency		
RQ1	What are suitable metrics inferred from a user's gaze for proficiency estimation?	3
RQ2	How accurately can we estimate proficiency from a user's muscle activity?	4
Facilitating Proficiency Awareness		
RQ3	What are design criteria for proficiency awareness in interactive systems?	5
RQ4	How can we support a wide audience in realizing proficiency-aware systems?	6

Table 1.1: Overview of the research questions that guided the work in this thesis.

In the second pillar, **RQ3** guides establishing guidelines for designing proficiency awareness. Here, we study an EMG-based feedback system and a system that provides proficiency feedback for a text production task. As an outlook, we probe the use of artificial proficiency for music improvisation in a study with piano players, deriving findings on how to seamlessly integrate proficiency enhancements. Complementing this part is **RQ4**, where we inquire how to support a wider audience in implementing proficiency-aware systems.

1.1.2 Research Context

The work in this thesis was conducted over the course of about five years in the Human-Computer Interaction and Cognitive Systems group at the University of Stuttgart and the Human-Centered Ubiquitous Media group at the Ludwig Maximilian University of Munich. Both groups are supervised by Prof. Albrecht Schmidt. During this time, several collaborations with project partners and researchers have shaped this thesis. A detailed list of contributing publications, involved co-authors, and my contribution to each of these publications is provided in the appendix (Contributing Publications).

SFB TRR-161 - Quantitative Methods for Visual Computing

The first parts of this thesis were conducted within the Collaborative Research Center SFB-TRR 161: Quantitative Methods for Visual Computing¹, connecting 17 project teams of the University of Stuttgart, University of Constance, and the Max-Planck Institute for Biological Cybernetics. Within the project C02: Physiologically Based Sensing and Adaptive Visualization, we researched new methods and techniques for cognition-aware visualizations. The collaboration within the research center resulted in joint publications [145, 146, 183, 97, 147].

¹ <http://www.sfbtrr161.de/>

ERC Amplify - Digital Technologies for Amplifying Human Perception

The major part of the research in this thesis was conducted within the European Union's Horizon 2020 Programme under ERCEA grant no. 683008 AMPLIFY² at the Ludwig Maximilian University of Munich. The overall objective of this project is to systematically research, explore, and model new means for increasing the human intake of information in order to lay the foundation for new and improved human senses enabled through digital technologies and to enable artificial reflexes. The work resulted in several publications [143, 148, 168, 123, 224, 62, 149, 331, 159, 151, 144, 150].

1.2 Contributions to Human-Computer Interaction

Proficiency-aware systems support the user in recognizing their own proficiency or lack thereof. While existing interactive systems readily adapt to a user's context, possibly including their proficiency, designing for the user's awareness of their proficiency is usually not the primary goal. This thesis investigates feasible proxies for estimating a user's proficiency and how these can be exploited to facilitate proficiency awareness in these systems. In the context of this research, we explore related aspects of the grand challenges for human-computer interaction (HCI) [292].

Calm Computing Contradicts Engaging Experiences

Ubiquitous technologies allow us to capture more and more user data. Context-aware applications leverage this information and provide tailored adaptation for the user [61]. Here, interactive systems have the potential to be perfect assistants, seamlessly integrating themselves into everyday tasks [324]. However, is "calm technology" [325] something users want and expect of future computing systems?

While perfect systems offer seamless interaction and maximize performance, we argue that optimizing for performance is not always beneficial for the user. Involved stakeholders might prefer the fastest, most efficient system operation, yet this goal is often dictated by short-term success and profit. Long-term benefits for users are ignored, although they could easily be realized with the collected data at hand. We believe that engaging experiences [250], where users are challenged and ultimately advance their own skills and knowledge, are much more rewarding. However, it is a fine line between motivating users through challenges while simultaneously not demanding too much of them, as this could potentially lead to abandonment.

Physiological Data Is Powerful but Complex

Leveraging physiological data from users in context-aware systems has the potential to innovate human-computer interaction. Through extending the available communication band-

² <https://amp.ubicomp.net/>

width between the user and the system, physiological computing allows for more natural and implicit interaction with machines [75]. Physiological data enables us to determine user states objectively and in real-time; no explicit response from the user is required [35]. Rather than being directed by users, systems can learn from users shifting to a more symbiotic relationship where systems leverage user awareness in real-time [161].

However, there are still fundamental challenges that need to be fully addressed before physiological computing can achieve its full potential, as identified by Fairclough [75]. Among these are the complexity of the psychophysiological inference and the appropriate user state representation. Especially in HCI research, the necessary expertise to develop such systems is sparse, inhibiting the widespread use of physiological computing in HCI applications. Yet, the potential of physiological data is most promising in the context of estimating users' proficiency.

Thesis Contributions

To advance research in tackling these challenges, this thesis introduces the concept of proficiency-aware systems. More specifically, we contribute:

1. A definition and presentation of proficiency-aware systems through a series of use cases.
2. An analysis of technical requirements for real-time, objective proficiency assessment for gaze- and EMG-based interactive systems.
3. Design guidelines for proficiency awareness in interactive systems, detailing *when* and *how* to communicate proficiency.
4. A toolkit approach to implementing EMG-based interactive systems and corresponding engineering guidelines for proficiency-aware systems.

1.3 Thesis Outline

The content of this thesis is distributed over seven chapters. Chapters 1 and 2 introduce the topic of this work and the vision of proficiency-aware systems. The two main parts of this thesis — Parts I and II — cover the two main research pillars in the context of proficiency-aware systems: detecting and quantifying proficiency as well as facilitating proficiency awareness. Here, Part I focuses on technical aspects, highlighting algorithmic work on gaze- and electromyography-based interactive systems. Part II complements this investigation by illustrating how proficiency awareness can be facilitated through design recommendations and toolkit approaches. Within these two main parts, chapters introduce related research and contain a number of use cases based on conducted research works. A summary highlights implications for each chapter. The thesis concludes with Chapter 7, which

summarizes design and engineering guidelines for proficiency-aware systems and presents opportunities for future work.

1.3.1 Introduction and Foundations

Chapter 1 – Introduction

The first chapter introduces the context, motivation and topic of this thesis. It further introduces the vision of proficiency awareness in interactive systems and summarizes challenges and resulting contributions of this research.

Chapter 2 – Towards Proficiency-Aware Systems

This chapter provides a detailed definition of proficiency-aware systems and how it is positioned within the research corpus of interactive systems. On the one hand, the chapter details how the conducted research draws from knowledge of context-aware applications, ubiquitous and physiological computing to address technical challenges of the presented vision. On the other hand, connections to research about user reflection and feedback are outlined to highlight their importance in creating engaging user experiences with proficiency-aware systems.

1.3.2 Part I: Detecting and Quantifying Proficiency

Chapter 3 – Proficiency From Gaze

Gaze is presented and evaluated as a sensing modality for proficiency-aware systems. In this chapter, we focus on interaction scenarios that predominantly include cognitive tasks. To do so, we leverage gaze-based estimation of user proficiency and present technical requirements and opportunities.

Chapter 4 – Proficiency From Electromyography

Contrary to Chapter 3, this chapter focuses on physical tasks involving motor skills and muscle movements. In this area, we identified electromyography as a suitable sensing modality and investigated technical limitations and challenges of proficiency estimation for fine motor tasks.

1.3.3 Part II: Facilitating Proficiency Awareness

Chapter 5 – Designing Proficiency Awareness

Drawing from the findings in Part I, knowing that proficiency estimation is technically feasible, this chapter illustrates how to subsequently design for proficiency awareness. Here, we

present three cases, concluding with design guidelines for EMG-based systems, everyday computer work and artificial proficiency.

Chapter 6 – A Toolkit Approach to Implementing Proficiency-Aware Systems

Complementing the previous chapter, Chapter 6 introduces a toolkit approach to implementing proficiency awareness. Having established that complexity of sensing modalities inhibits their applicability in proficiency-aware systems, we present EMBody, a data-centric toolkit for EMG-based interface prototyping and experimentation.

1.3.4 Conclusion and Future Work

Chapter 7 – Conclusion and Future Work

This chapter concludes the research conducted within this thesis. We summarize the findings through a set of design and engineering guidelines for proficiency-aware systems, addressing challenges that we identified at the outset of this work. Additionally, we discuss opportunities for future work and the further development of proficiency-aware systems.

Chapter 2

Towards Proficiency-Aware Systems

In everyday tasks, we are supported by interactive systems that adapt to our actions and environment. While these systems are intended to make our lives easier, designers of such systems tend to overlook how the observed context could be used to advance a user's skill and expertise. Allegedly, there are valid design reasons for doing so, such as (1) the inability to provide meaningful information for the user, (2) the increased cognitive effort required of the user, or (3) the absence of a suitable learning domain and established procedures.

In this thesis, we showcase that communicating proficiency to users is feasible and beneficial, regardless of task complexity or domain. We also show the inherent potential of interactive systems to support users in advancing their proficiency. The following definition of proficiency-aware systems provides the framework for our investigation. It is informed by research in the area of adaptive systems and context-aware applications, as well as research on user reflection and feedback within the domain of human-computer interaction. The former provides us with valuable insights into how to measure and use proficiency as input for proficiency awareness. The latter informs our research on how to communicate and leverage proficiency for the user's benefits. The duality of this approach is detailed in this chapter.

This chapter is based on the following publications.

Jakob Karolus, and Albrecht Schmidt. 2018. Proficiency-Aware Systems: Adapting to the User's Skills and Expertise. In *Proceedings of the 7th ACM International Symposium on Pervasive Displays (PerDis '18)*. ACM, New York, NY, USA.

<https://doi.org/10.1145/3205873.3210708>

Jakob Karolus, and Paweł W. Woźniak. 2021. Proficiency-Aware Systems: Designing for User Reflection in Context-Aware Systems. *it - Information Technology*.

<https://doi.org/10.1515/ITIT-2020-0039>

2.1 A Definition of Proficiency-Aware Systems

The term proficiency is a multifaceted construct incorporating, among others, the user's inherent abilities, experience, knowledge, and acquired skill. The presented definition is concerned with an operationalization of this construct and its role within proficiency-aware systems. We first provide a clear definition of the term *proficiency*, which will form the basis for the presented definition of proficiency-aware systems.

The Oxford English Dictionary [225] defines proficiency as:

- a. The quality or fact of being proficient; the degree of competence attained; adeptness, skill *in* a particular field.
- b. A skill, a talent; (now frequently) a certain standard of skill acquired after a period of education or training.

Note that proficiency is bound to a particular field of application. We represent this fact in our definition by allowing proficiency to be an aggregation of skills in associated task domains. For proficiency-aware systems, we define proficiency as:

Definition - Proficiency

Proficiency is the aggregated construct of any skills, knowledge, competence, or experience of a person relevant to the interaction between the person and a system (the task domain).

If we define proficiency as an aggregated construct, it is easier to operationalize and quantify proficiency. Specific tasks can be broken down and the associated skill levels can be obtained with ease. For example, it might be challenging to create a metric to ascertain a user's proficiency in "write an essay about American history in plain English". However, it is feasible to combine their skill in "write in plain English" and their competence in American history.

Accordingly, we define proficiency awareness and proficiency-aware systems as:

Definition - Proficiency Awareness

Proficiency Awareness of (computing) systems is a functionality of estimating a user's proficiency in the associated task domain and facilitating its understanding for the user.

Definition - Proficiency-Aware Systems

Proficiency-Aware Systems are systems that use estimates of a user's proficiency to tailor the interaction in their associated task domain(s). A **proficiency-aware system** is able to adapt content, presentation, and interaction to

1. support the user in accomplishing their task and
2. facilitate the user's understanding of their own (lack of) proficiency.

Hence, the term proficiency-aware system refers to a dynamic interconnection between system and user that continuously assesses the user's ability to perform tasks using the system and adapts accordingly. Such adaptations will enable a user experience tailored to the specific user's proficiency. The proficiency hereby refers to the knowledge and skills relevant for using the system, such as reading skills, background knowledge, motoric skills, and cognitive abilities.

Notably, this definition allows us to draw parallels to context and context-aware systems (cf. Section 2.2.3). At its core, a proficiency-aware system is a context-aware system that focuses on the user's proficiency. Challenging the user to improve their proficiency is a novel aspect of proficiency-aware systems and contrasts to the striving for conformity of other assistive systems. This potentially lets users reflect on their work — a vital design goal in HCI [16] — facilitating an understanding of their own proficiency (cf. Section 2.3.1).

In our vision, a proficiency-aware system provides the means for users to become better in the task domain. We envision that this creates engaging experiences [250], where users can fully reflect on their proficiency and its progress. These proficiency metrics do not necessarily have to be performance-oriented, as long as the user knows how to connect them to the task. Instead of developing highly customized systems with delicate algorithms, proficiency-aware systems rely on a user-driven interpretation of context, where the user is included in the process and provides the final interpretation of available data.

An envisioned model of a proficiency-aware system is depicted in Figure 2.1. The system is governed by a continuous cycle involving proficiency estimation on the one side and appropriate adaption on the other side. Both sides mutually influence each other until an equilibrium is reached. Relevant skills are inferred and supplied to the system. Here, the system itself does not necessarily have to be responsible for estimating individual factors contributing to a user's proficiency but aggregates factors to a suitable proxy for adaptation. Additionally, a proficiency-aware system facilitates a user's understanding of their proficiency through communicating the estimation back to the user.

2.1.1 Proficiency as an Aggregated Construct

In our conceptualization for proficiency-aware systems, proficiency is an aggregated construct of relevant user skills, knowledge, experience, and competence that may be inferred

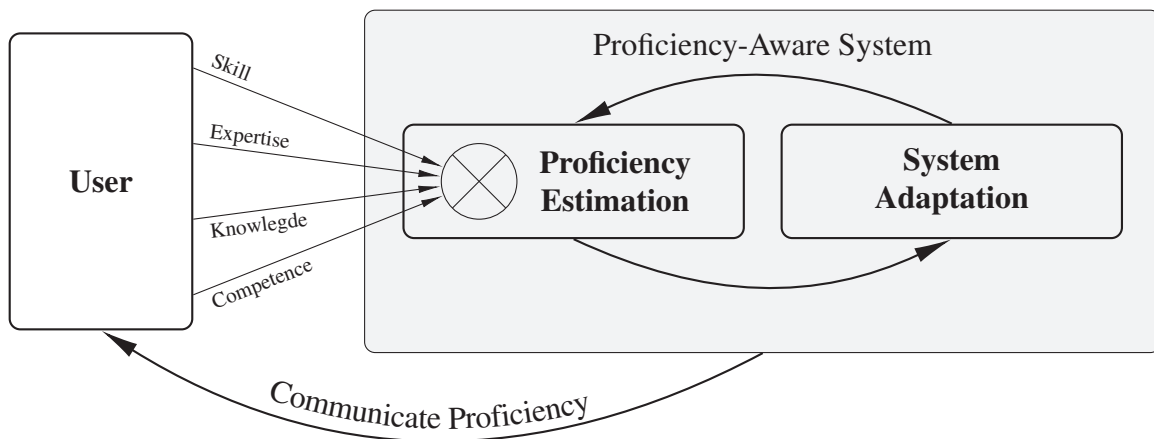


Figure 2.1: Depiction of a proficiency-aware system. Relevant skills are inferred and supplied to the system. Adaptation and estimation mutually influence each other until an equilibrium is reached. Additionally, a proficiency-aware system communicates proficiency back to the user.

from measured metrics. The aggregation of relevant user skills contributes to an understanding of the user’s proficiency in the actual task at hand. While overall proficiency is usually difficult to measure accurately, individual contributing skill levels may be easier to estimate. In practice, it is often sufficient to consider only relevant skills that might influence a person’s proficiency for the given task. Identifying these proxies is a key design task when building a proficiency-aware system.

We illustrate this with an example: a professional editor for a newspaper is given the task to proofread an article about a breakthrough discovery in computer science. In this case, we can assume a high skill level in writing, a skill inherent to this particular user. Though, unless the editor has a computer science background, their competence about the topic is limited. This means that the editor is already at a disadvantage, as their limited knowledge about the topic impacts upon the proofreading task. Here, a proficiency-aware system might offer explanations for unfamiliar terms or suggest articles to study to prepare for the task.

Further, let us consider two different contexts: a typical workday around 11 am and a stressful day just before the end of work. The editor’s proofreading skill is unchanged, yet, the editor will have a harder time with the task within the second context. This is a typical example of a context-aware system and situational impairments [327]. Notice that we explicitly did not include short-term impairment factors in our definition of proficiency. Such factors should be modeled through context. Hence, a proficiency-aware system can make use of the context by deciding to support the editor for the second context, focusing on task completion as the editor is probably not keen on sitting through an extended learning session. In contrast, if the context allows, the proficiency-aware system points out ways to increase the editor’s knowledge about computer science.

For some systems, this set of skills is readily available and straightforward to measure. However, it is often impossible to quantify the skill level to a degree necessary to provide effective

adaptation. Consider this next example: you are tasked with writing a proposal for a broad audience on how to make your city greener. This task requires skills in several areas, such as writing in plain and understandable English, structuring one's ideas concisely, and providing convincing arguments. The latter two skills are inherently difficult to estimate. Yet, it may not be necessary to model the complete proficiency to support the user. Here, measuring the skill of being able to write in understandable English could already be sufficient. This puts the writer in a better position to write the required proposal, not only for this specific topic but also for other topics.

Part I of this thesis will focus on how we can detect and quantify users' proficiency, addressing the technical challenges of proficiency-aware systems. To complement this research, Part II will focus on how we can provide adequate proficiency support for the user and facilitate an understanding of proficiency through feedback and reflection. The following sections will depict relevant foundations in each research area.

2.2 The System Side

Proficiency-aware systems are inherently adaptive systems that react to user input during runtime. More specifically, proficiency-aware systems are closely related to context-aware applications and may base proficiency estimation on ubiquitous sensing modalities. This section highlights necessary foundations in these areas and how they inform the design of proficiency-aware systems.

2.2.1 Adaptive Systems

Addressing user variance through adaptive user interfaces that allow users to customize and personalize the system to their needs has been in the focus of human-computer interaction for decades [96]. Adapting to and predicting [116] user needs dates back to early user interfaces, such as the command-line [174]. Common approaches focus on adding intelligence [22] and user modeling [84].

Proficiency is closely related to a person's inherent cognitive and physical abilities. While there exists work that investigates "ability-based optimization" [264, 265], long-term or permanent inhibitors of proficiency should be considered separately and modeled appropriately. The main strengths of proficiency-aware systems lie in the adaptation during runtime. Consequently, proficiency-aware systems are designated as adaptive systems, not adaptable systems.

Interestingly, adapting user interfaces may not always be beneficial for the user. Predictability and accuracy of adaptation play a vital role in user acceptance [88]. Users tend to be reluctant to undergo customization, as it requires time and cognitive effort while immediate benefits are not visible [187, 176]: the "paradox of the active user" [36]. As outlined,

a proficiency-aware system should confer the benefits of undergoing a change of task approach and modality. The system needs to convince the user that increasing their proficiency will pay off in the long run, a key design goal for proficiency-aware systems. Sometimes this may even require withholding possible adaptations, contrarily to classical adaptive systems.

2.2.2 From Calm to Engaging Ubiquitous Computing

The omnipresence of interconnected systems, from computers to smartwatches, intelligent wearables, and smallest computing units, marks the three parts of technology required for ubiquitous computing as envisioned by Mark Weiser in 1991: accessible computing units (with displays), appropriate software applications, and a network tying them together [324].

"The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it."

Weiser [324]

Weiser believed that pushing computers into the background would allow us to become more aware of the people on the other end, especially for computer-based communication; an early take on the unhealthy impact of computerized workplaces. While contradictory at first, his concept of "calm technology" [325] achieves its merits by providing us with more information. The idea is to provide this additional information in a way that seamlessly integrates into the workplace and everyday life: technology that is "tuned into what is happening around us" [325].

While HCI research has made extensive progress towards adaptive mechanisms and systems, it can be argued that calm technology might not necessarily be most beneficial for users of adaptive systems, challenging Weiser's original vision. Instead, supporting systems should be visible and provide meaningful engagement [250]. This is a key goal for proficiency-aware systems, so that users can fully reflect on their proficiency and use of the system.

2.2.3 Context-Aware Computing Applications

With a multitude of devices comes the significant challenge of interconnection and cooperating on the user's behalf, especially in constantly changing environments. In a seminal work from 1994, Schilit et al. [272] already formulated this challenge, but also saw an opportunity in it:

"One challenge of mobile distributed computing is to exploit the changing environment with a new class of applications that are aware of the context in which they are run."

Schilit et al. [272]

Technological advances have developed the original definition of context-aware applications and the four categories as proposed by Schilit. While the user's location: "where you are, who you are with, and what resources are nearby" [272] played a vital role in early systems and Schilit's definition, other sources of information have become equally important.

A more recent definition by Anind K. Dey provides a slight generalization to the definition of context:

"Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves."

Dey [60]

In this thesis, we draw from this definition for our own definition of proficiency-aware systems (cf. Section 2.1). In essence, proficiency can be modeled as context as it characterizes the situation of the user. Consequently, proficiency awareness is based on context awareness:

"A system is context-aware if it uses context to provide relevant information and/or services to the user, where relevancy depends on the user's task."

Dey [60]

This aspect is reflected in the first part of our definition of proficiency-aware systems: "support the user in accomplishing their task". However, while Dey's definition above already considers user skill as relevant information, we argue that this is rarely employed for the user's benefit besides supporting task completion. Given that proficiency assessment is technically feasible, it remains a challenge to investigate how being aware of one's proficiency affects everyday tasks and how systems can effectively communicate proficiency metrics to assure an optimal work experience.

While system adaptation and change of context go hand in hand, there is more to this reciprocal relationship:

"A change in the application will inevitably lead to a change in the context (...)"

Schmidt [273]

We believe that explicitly communicating proficiency in addition to implicit system adaptation is worthwhile to investigate. In essence, this thesis focuses not only on how the system's adaptation influences the current context but also how communicating this proficiency estimation can benefit the user apart from the current task at hand.

Consequently, there are unexplored opportunities in understanding the user's skill level and using that information to improve the user experience of everyday computer work. In this thesis, we highlight proficiency as an underexplored aspect of the context in context-aware systems. Understanding how we can leverage proficiency to assist users in their primary tasks and support users in understanding their own proficiency is a crucial aspect of proficiency-aware systems.

Yet, this form of implicit interaction is inherently flawed if users question and disagree with the system's decisions. This issue is especially prominent for creative tasks, art, or ludic interaction. By simply providing means for proficiency advancement, e.g. a stroke analysis for a painter, we envision that proficiency-aware systems are able to support creative tasks. Here, users have full control in deciding whether those means are important for them.

2.2.4 Physiological Sensing

Being able to assess user states objectively alleviates the need for explicit customization and prevents subjective user bias. A specific method within this paradigm is physiological computing [75], which aims for real-time adaptation through measuring psychophysiological user parameters.

"Physiological computing has the potential to render human-machine dialogue as dynamic, collaborative, spontaneous and effortless."

Fairclough [75]

Physiological sensing offers opportunities for interactive systems, yet there are fundamental challenges that inhibit this potential [75]. Among others, the complexity of the psychophysiological inference and the appropriate user state representation often require extensive expertise, posing a high entry barrier for use in HCI applications. However, with the proliferation of ubiquitous sensing technologies, the setup of sensors and associated algorithms are becoming more accessible [30, 156].

In this thesis, we look at two specific modalities: gaze and electromyography in the context of physiological sensing and their applicability for proficiency-aware systems. Gaze is an unobtrusive sensing modality that provides insights into cognitive processes of users, while electromyography exhibits high distinctiveness for recognizing motor task (cf. Section 1.1). Besides looking at their capabilities in distinguishing proficiency levels, we investigate how these technologies can be made accessible to less technically adept users by pursuing a toolkit approach.

2.3 The User Side

The other key design goal of proficiency-aware systems is to facilitate a user's understanding of their own proficiency. Rather than employing a user modeling approach, proficiency-aware systems offer means for reflection and learning through context awareness. Thus, we need to investigate *where*, *when*, and *how* to facilitate this understanding.

2.3.1 Reflection and Feedback

Reflection has been a recurring theme in HCI. Schön's [278] work on the nature of reflection has been particularly influential [16]. He distinguishes between reflection-in-action and reflection-on-action. Reflection-in-action happens when performing a task and noticing unexpected outcomes. Reflection-on-action is retrospective. While most systems focus on reflection in a holistic interpretation, focused on daily life patterns [253], wellbeing [11] or crowdsourced tasks [66], this thesis uses a more atomistic approach to reflection. The concept of proficiency-aware systems supports reflection-in-action through making the consequences of users' actions easily visible, exploring ways to foster reflection through interface elements.

While many explorations have been conducted, it remains a challenge to understand the design qualities that may foster personal reflection in interactive systems [15]. Pointers from related research, such as design dimension of personal visualizations [126] and self-reflection with ubicomp technologies [182], guide the research in this thesis (cf. Chapter 5). Notably, this thesis draws from related work to guide the design of *task-oriented reflection*, i.e., systems enable users to reflect on how they accomplish a task, rather than systems encouraging personal self-reflection.

2.3.2 Learning and Skill Acquisition

Learning is the process of acquiring new knowledge and skills, either through experience or by being taught. There are many theories about how we acquire and consolidate new knowledge, such as Sweller's cognitive workload theory [301]. This theory explains which

cognitive factors can inhibit or foster learning. Out of the three components (intrinsic, extraneous, and germane), germane workload is associated with the user's effort to recognize patterns and structure within the task, which might foster learning [302]. However, this separation of workload components is highly disputed [277] and challenging to operationalize for interactive systems.

Analogously to the acquisition of cognitive skills, the exact mechanisms of consolidation for physical skills — so-called motor memory — are not entirely understood either [171]. To learn new motor tasks, we rely on mimicking movements to get a sensation of how the movements feel, relying on proprioception to orient ourselves in physical spaces [238].

Consequently, the work in this thesis focuses on how we can facilitate insights for users into their proficiency, in contrast to providing an optimal learning strategy for the given tasks. As learning strategies are highly person-dependent, a more abstract approach is favored in this investigation by providing users with the means to engage in reflection but not with a dictated method on how to do so.

2.3.3 Explainability and Interpretability

This aforementioned approach is also favorable in light of recent advances in machine learning. Powerful deep learning applications significantly advanced this research area. Yet, the complexity of these algorithms comes at a price. The so-called black-box algorithms often fail to explain their decision-making sufficiently. This is not only a problem for the developer (cf. algorithmic bias), but also for the user, especially if decisions are unpredictable and questionable [37, 92]. The consequence is a loss of user trust.

As mentioned earlier, this work takes a different approach by (1) leveraging established machine learning applications that allow for sufficient data transparency (Part I) and (2) by investigating an approach for user-driven data interpretation (Part II), allowing the user to make sense of the collected data.

2.4 Challenges and Dimensions of Proficiency-Aware Systems

Proficiency-aware systems can be found everywhere and rely on a variety of sensors. They can range from very simple to highly complex systems. Yet, developers rarely explicitly consider designing for proficiency and the associated user benefits.

We identified a number of related challenges for the design and implementation of proficiency-aware systems, which inspire the research parts of this thesis. The following section highlights the challenges and how we addressed those in the presented research. We note that this set is not exhaustive and only highlights particular dimensions identified by the work in this thesis.

2.4.1 Interaction Duration and Frequency

A proficiency-aware system might be used once for a short duration or be employed for a longer duration, such as a student learning system [63]. Consequently, the interaction duration of a user with a proficiency-aware system is often predetermined by its usage scenario and cannot be altered. A short duration must rely on local adaptation based on ad hoc context, while prolonged interaction allows for a much more customized user experience, as systems potentially have access to an extensive history of data.

In this thesis, we investigated systems that employ short-term and ad hoc user interaction (Section 3.2). We identified that especially gaze as a modality is suited for this form of interaction. Other gaze-related use cases (Sections 3.3 and 3.4) investigated prolonged interaction.

For electromyography (Chapter 4), we identified prolonged and multiple use of a proficiency-aware system as more beneficial, as the data collected through this modality is highly person-dependent. By allowing for sufficient observation time or even continuous observation, precise proficiency assessments are possible.

Ultimately, this thesis reports on the feasibility of modality choices with regard to proficiency-aware systems and highlights favorable interaction forms.

2.4.2 Adaptation Control and Agency

When interacting with a system, users perceive a sense of control when they feel that actions of the system are evoked by them [209]. Similarly, a proficiency-aware systems can present its users with a choice of adaptations by providing a recommendation, giving full control to the user. However, this method requires an explicit action by the user. In contrast, implicit adaptation allows for a seamless user experience while potentially leaving the user with a feeling of a loss of control.

If the system has full control over the adaptation, users may perceive a loss of agency. Here, adaptation frequency and noticeability of changes are crucial [38]. Our research is guided by related works in this area, noticeably in the area of physiological computing. BACH [338] is a system that automatically adjusts the playing difficulty of piano pieces based on the user's cognitive workload. Here, players reported more efficient learning with BACH, although they had to relinquish control over how they approach learning the piece. Research by Ewing et al. [74] takes this to the extreme by implementing a biocybernetic loop that feeds EEG measures directly to the system allowing continuous adaption of the difficulty in a Tetris game.

Consequently, it can be beneficial to restrict user control if the task can be divided into ordered levels of proficiency. We include aspects of this approach throughout our use cases, where the user relinquishes control to the system for some parts of the task (Sections 3.4, 4.2 and 4.3).

In contrast, full user control requires explicit intervention. In Section 4.2, we evaluated both approaches' pros and cons by comparing automatic to manual adaptation in a guitar tutoring system. The system only suggested a new difficulty level in the form of tempo adjustments. The user was free to adhere to this recommendation or to choose their own tempo. Players reported that the suggested tempi provided a good challenge and trusted the system's assessment. This showcases that especially beginner and amateur players appreciate a helping hand in curating their learning process.

2.4.3 Task Complexity And Domain

To deliver effective feedback, it is often essential to incorporate task domain knowledge. Informed feedback tailored to the task domain can provide a more efficient learning experience and show users how to improve. Yet, this knowledge might not be available.

In this thesis, we demonstrate that the concept of proficiency awareness is not limited to complex systems but can be applied for everyday tasks (cf. writing task in Section 5.2). Even if the task domain is complex (learning an instrument), there exist feedback choices that are beneficial for the user (cf. Sections 3.4 and 4.2). One of the main goals of this work is to depict that even simple metrics can often stimulate reflection in users. This strategy potentially allows proficiency-aware systems to scale better than highly customized adaptive systems.

2.4.4 Feedback Types and Communicating Proficiency

It remains a challenge how to properly convey motivation and avoid judgment. For users, it might not be evident that the provided feedback is task-oriented, reducing its effectiveness. Similar to most interactive systems, proficiency-aware systems suffer from a lack of transparency during the decision-making process [37]. Automated analysis of complex signals is often challenging to understand for laymen. Hence, users view those decisions with skepticism. Developers and engineers find sensing modalities too complex [75]. It should be evident to the user what and why data is collected and, most importantly, for what it is being used. Consequently, transparency of communicated proficiency should be carefully considered to gain the user's trust in the proficiency estimation.

This thesis addresses this challenge in two ways. First, we investigate different dimensions of feedback that communicate proficiency back to the user for a variety of use cases (cf. FitBack use case in Sections 4.3 and 5.1 as an example). Secondly, we believe that familiarity with sensing modalities is essential for their acceptance. Chapter 6 directly addresses this issue by proposing a toolkit for EMG-based interactive systems.

I

DETECTING AND QUANTIFYING PROFICIENCY

Chapter 3

Proficiency From Gaze

Seminal works by Buswell [33] and Yarbus [334] have first investigated a relationship between high-level cognitive processes and gaze behavior, indicating a top-down approach of viewers in gaze-related tasks. More importantly, Yarbus [334] showcased that even when confronted with the same stimulus, the eye movements of viewers were task-specific, i.e., they exhibited distinctive gaze behavior during free viewing vs. when given tasks related to the stimulus, such as searching for an object.

Consequently, our eye movements and gaze behavior convey a lot of information about our mental processes, including perception and cognition. As outlined in Section 1.1, this thesis leverages gaze as a ubiquitous sensing modality for user proficiency in tasks that are predominantly of cognitive nature. This chapter explores how user interfaces can use information, such as its users' gaze, to recognize the users' proficiency. We contribute to the understanding of human gaze properties to build interfaces in proficiency-aware systems.

The investigation in this chapter is strongly informed by existing research in connecting eye movements and gaze patterns to user activities, skills, and knowledge. Seminal works include Rayner's [245] work on eye movements and reading, as well as Jacob and Karn's [134] review on eye-tracking metrics in HCI. We research the applicability of established gaze metrics for proficiency-aware systems with a focus on specific challenges as outlined in Section 2.4 in three use cases, guided by **RQ1**:

RQ1

What are suitable metrics inferred from a user's gaze for proficiency estimation?

Language Proficiency

In this use case, we focus on interaction duration and frequency. Since detecting language proficiency is a solved problem if enough gaze data is available (cf. Section 3.1), we look at specific task requirements for ad hoc systems that only have limited access to a user's gaze data. This covers our dimension of ad hoc proficiency-aware system (cf. Section 2.4). In

particular, the experimental study focuses on a task with low complexity (the visual stimulus is simple) and already established metrics.

Programming Proficiency

Contrary to the first use case, estimating users' programming proficiency is complex as the teaching background greatly influences the user's task-solving strategy. In this experimental study, we opted for a longer interaction duration with the system. Consequently, we analyzed top-down task-solving strategies of users and evaluated whether proficiency influences these strategies.

Proficiency of Piano Play

In this last use case, we again opted for a complex task scenario (cf. Section 2.4). Like the previous two use cases, we looked at established gaze metrics and extracted patterns prominent for piano play. This use case additionally focuses on the user's gaze metrics in a continuous tutoring scenario. The system has detailed access to the user's history of gaze data, complementing our dimension of interaction duration (cf. Section 2.4).

This chapter is based on the following publications.

Jakob Karolus, Paweł W. Woźniak, Lewis L. Chuang, and Albrecht Schmidt. 2017. Robust Gaze Features for Enabling Language Proficiency Awareness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA.

<https://doi.org/10.1145/3025453.3025601>

Calvin Liang, Jakob Karolus, Thomas Kosch, and Albrecht Schmidt. 2018. On the Suitability of Real-Time Assessment of Programming Proficiency using Gaze Properties. In *Proceedings of the 7th ACM International Symposium on Pervasive Displays (PerDis '18)*. ACM, New York, NY, USA.

<https://doi.org/10.1145/3205873.3210702>

Submitted for review.

Jakob Karolus, Johannes Sylupp, Albrecht Schmidt, and Paweł W. Woźniak. EyePiano: Leveraging Gaze for Data-Driven Piano Tutoring.

3.1 Related Research

Gaze has played a major role, both as input modality and evaluation tool in HCI. The following section summarizes relevant research in this area. For an extensive overview of gaze metrics in HCI and gaze-based interaction, we refer the reader to works by Jacob and Karn [134] and Duchowski [69].

Explicit Gaze Input in HCI

A user's gaze can be used as an explicit input to computing interfaces, albeit to varying levels of effectiveness. Most prominently, gaze has been used as a substitute or supplement for manual cursor control [288, 255], e.g. in MAGIC-Pointing [339] and typing [189]. Interfaces have also been developed that respond to distinctive and contrived gaze gestures [67] as a form of explicit interaction.

Context-Awareness Through Gaze Data

Besides this, gaze can also serve as an implicit input for context-aware applications, such as activity recognition [31] while wearing custom-made electrooculography (EOG)-glasses [30]. iTourist [241] successfully plans city trips based on recorded gaze patterns. Gaze contingent displays [70] use the gaze point of the user to define a region of interest. The information is used to highlight specific elements in the user's view [6] or to selectively render foveated regions at high resolution in order to save computing costs [100, 201].

Gaze Cues in Conversations

Our daily communications afford us interpersonal cues that allow us to moderate our language and responses to our conversation partners. Telltale facial expressions and body language can indicate the interest and engagement levels of our conversation partners. Gaze can readily indicate whether they are focused on the conversation or distracted by something else, such as looking at their phone or recognizing another person in the background [86, 118]. As humans, we are readily able to interpret these implicit signals and adapt our conversation strategy by changing the topic or requesting attention.

Other work in the domain of implicit interaction focuses on human-like interaction in virtual environments, e.g. for conversational agents [293, 318]. It has been shown that gaze directional cues can serve as a predictor for conversational attention. Moreover, virtual agents that respond to users' gaze have been shown to increase their users' emotional response and allocated attention [197]. This suggests that gaze-responsive systems are perceived as being more human-like in their interactions and, hence, elicit more user-attention.

Gaze and Reading

The connection between eye movements and reading has already been researched thoroughly [245, 247]. There exist psychologically plausible models that describe many phenomena in reading [141, 248]. Hence, computer systems can also rely on gaze inputs to assist users in accomplishing their task. In tutoring systems, the user's gaze information can be relied on by the system to determine when it is necessary to provide feedback and guidance to the user [63]. For example, patterns in gaze behavior could reveal that the user is confused by a given topic [3], which prompts the system to provide further guidance [221]. Learning new topics often involves reading provided material, such as a book, a document or even a simple time-table of a bus station.

Gaze as an Indicator of Personal Interest

Yet, representing information in textual form is inherently limited in terms of data dimensionality and scalability. Walber et al. [321] have identified factors in gaze data that are connected to our personal interests and used it to create a personalized photo collection without the need for manual interaction. Furthermore, work from Santella et al. [260] showcased that this idea can even be applied to single images by identifying interesting image parts and thus realizing an automatic photo cropping algorithm.

Gaze and Language

In recent years, researchers built upon these findings to examine language proficiency [199, 335], as well as develop systems that have educational purposes. These include e.g. real-time annotations [41, 23, 157] or translations during reading [310]. In this scenario the users' eye movements provide contextual information about what they are reading, how fast [173] or even how much they understand [131, 257] and provide assistance accordingly [287, 132]. These systems showcase the capabilities of gaze-assisted language detection and/or translation. A vital part of real-time assistive systems is the exact point in time when to assist.

3.2 Use Case I: Robust Gaze Features for Enabling Language Proficiency Awareness

In an increasingly globalized world, users are constantly exposed to unfamiliar cultures and languages. Concurrently, our lives are increasing reliant on digital technology as our environments - in public, at work or at home - are pervaded by digital artifacts such as public displays and shared mobile devices. When visiting foreign countries, especially those that are multilingual, users are often faced with the challenge of navigating an interface in an unfamiliar language. While alternative language versions are often available, accessing them (usually through a poorly visible button or a submenu) tend to be cumbersome. The problem is compounded by screen space limitations.

Consequently, we are often confronted with information interfaces designed in an unfamiliar language, where the language barrier inhibits interaction with the system. In this use case, we look at how a user's gaze properties can be used to detect whether the interface is presented in a language they understand, in other words, whether they are proficient in the given language. We examined users' gaze properties (gathered through an embedded eye tracker) when reading short sentences in multiple languages. Afterwards, we evaluated how gaze properties can be interpreted to detect if the interface is presented in a language that the user understands.

We evaluate the feasibility of language awareness in proficiency-aware systems leveraging short time gaze data. First, we identified that users exhibited shorter average fixation duration

as well as longer average and summed blink duration, when presented with languages that they were proficient in. Based on these results, this use case outlines a classification strategy and provides some technical guidelines to facilitate real-time language proficiency detection.

3.2.1 Methodology

Eye-tracking is widely used in the domain of reading analysis [245]. However, existing approaches have usually focused on one specific language and are used for posthoc analysis [199, 335]. We propose an approach that relies on characteristic gaze properties to infer a user's requirement for real-time assistance. The goal is to recognize a user's likelihood of comprehending displayed text, given limited gaze data based on reading a short sentence in a given language. We anticipate that users are unlikely to persist in using a system beyond a single sentence in a language that they lack proficiency in. In this regard, the current work contrasts with eye movement research based on reading entire documents. When working with eye gaze interaction, prior calibration often is a hindrance for readily available interaction. We will discuss this limitation and analyze its impact on user experience given our approach.

For our study design, we considered previous research findings on eye movements and reading to identify gaze characteristics that could be viable candidates for inferring language proficiency in our application context. Research reports an average duration for fixations of about $200ms$ to $250ms$ and an average saccade length of 7 to 9 letters, but fixation duration can range from $50ms$ to $500ms$ depending on the context of the task and user state. Pertinently, the fixation duration increase with conceptually more difficult text, leading us to hypothesize that fixation duration should vary with a user's language proficiency. In addition, text difficulty is reported to correlate with saccade length as well as the frequency of saccade regressions¹ and refixations. Hence, difficult languages should exhibit lower saccade length and a higher frequency² of regressions and refixations. [245, 94]

Moreover, blink rate and pupil diameters have been reported to be associated with cognitive load and human information processing. For example, Siegle et al. [289] demonstrated a phasic increase of blinks prior and pursuant to the anticipation of an increase in the load of information processing, which was manipulated with basic psychological tests (i.e., Stroop task, digit-sorting task). Pupil dilation, on the other hand, reflected sustained information processing over longer periods, making this metrics less suitable for our inquiry [289].

On the basis of **RQ1** as reiterated below, we formulate two related research questions tailored towards this use case:

¹ Saccades opposite to the reading direction

² Standard frequency is about 10 to 15%.

RQ1

What are suitable metrics inferred from a user's gaze for proficiency estimation?

RQ1a: Can we determine whether a user is able to use the interface in a given language from gaze properties?

To realize near real-time proficiency estimation, it is vital that the system is able to tell a user's proficiency in the display language in a mere seconds. This implies that the users should only be presented with at most a few sentences of an unknown language before the system decides on the respective proficiency. This is the basis for the second research question:

RQ1b: What are technical requirements when determining whether a user understands the current interface language?

Based on these questions, we investigate the following hypotheses in our study:

H1: Increased language proficiency level will result in a lower average gaze fixation duration.

H2: Increased language proficiency level will decrease the number of refixations in a given time period.

Lower average gaze fixation duration as well as lower refixation ratio is connected to text difficulty when reading [245]. In our study we aim to vary text difficulty by changing the display language, hence creating several different text difficulty levels. Participants should therefore find it easier to answer questions in their proficient languages.

H3: Increased language proficiency level will increase the average blink duration.

H4: Increased language proficiency level will increase the total blink ratio.

Proficient users of a language should experience lower cognitive load during reading than non-proficient users. We assume that this increases blink duration and total blink ratio when presented with a language in which the user is proficient. During unknown languages cognitive load increases as the user is trying to figure out the question, hence blinks are less frequent.

If we reflect on one primary scenario – public displays in commuting areas – the user usually looks for predefined information (e.g. bus departure, flight details). The method of presenting this information is often ordered and predictable, such as a time-table, and does

Language	Example questions	
English	How many days are within a week?	What is the first letter of your first name?
French	Combien de jours y a-t-il dans une semaine?	Quelle est la première lettre de votre prénom?
Danish	Hvor mange dage er der i en uge?	Hvad er det første bogstav i dit fornavn?
Finnish	Kuinka monta päivää on viikossa?	Mikä on etunimesi ensimmäinen kirjain?

Table 3.1: Example questions used in our study in four different languages.

not change across languages. Yet this is of no help for the user if he cannot locate the name of his destination or instructions for the payment process.

Information like this is expressed in simple sentences as well as simple language, to ensure readability. Hence, we chose our questions to be of “Basic User” ([51], p. 23) level with regard to the *Common European Framework of Reference for Languages* (CEFR) [51]. This framework defines six levels of foreign language proficiency: basic (A1, A2), independent (B1, B2) and proficient user (C1, C2) [51]. For our study we added an extra non-proficient level (X).

We collected questions in 13 different languages (15 questions each). By choosing to display simple questions to the participants, we provide an engaging task [120] as reading the text is required to answer correctly. The respective translations were provided by either native or highly proficient users of that language. Table 3.1 shows a few example questions. Most are part of the Indo-European language family [59], yet we included some outliers such as Finnish and Hungarian as well as languages not using the Latin alphabet (Greek, Arabic). See Table 3.2 for a complete overview on the used languages and the respective proficiency levels exhibited by our participants.

3.2.2 Evaluation

We conducted an experimental study following a within-subject design for the factor of language proficiency. This section first describes the study procedure and measures as well as later highlights results including classification performance for language proficiency.

Apparatus

Our setup consisted of a 22 inch LCD display (resolution: 1680x1050) and a remote eye tracker (SMI RED 250; 250 Hz sampling frequency) that was positioned below the display. Our participants were seated at a distance that felt comfortable for them and well within the reliable tracking range of the system (0.5 m – 0.7 m) in an enclosed cubicle. Figure 3.1 shows a picture of the apparatus.

Language	Proficiencies of participants
English	B2, C1, C2
German	X, B2, C1, C2
Danish	X
Dutch	X, A1
Finnish	X
French	X, A1, A2, B1, B2
Greek	X, A1, C1
Romanian	X
Spanish	X, A1, A2, B1, C2
Turkish	X
Slovenian	X
Arabic	X, A1
Hungarian	X

Table 3.2: Languages used in this study and respective proficiencies present in our participant base.



Figure 3.1: Apparatus showing LCD monitor with attached eye-tracking device (left) and participant during the study (right).

Participants

We recruited twenty-nine participants from the University of Stuttgart and the Stuttgart Media University via mailing lists. The data of 21 participants (12 female, 9 male; Age: $\bar{x} = 24.9y$, $s = 3.9y$) were used for further analysis. The eye-tracking data of three participants were removed because glasses and make-up interfered with the reliability of eye-movement recording. We excluded the data of five more participants, as the recording was not stable enough during the whole study³. Two participants had prior experience with eye-tracking studies. Out of the 21 participants, 17 were native German readers, two were native

³ Participants started to move more towards the end of the study.

English readers as well as one native Spanish and Indonesian participant. All participants had normal or corrected-to-normal vision. Each participant was paid an allowance of 10 Euros.

Procedure

After introducing the prospective participants to our study, we handed them a detailed study description. It stated to follow on-screen instructions worded as simple questions, that each require a single key-press to answer. The study description sheet additionally contained an example question. Prospective participants were made aware to expect these questions in different languages and to answer to the best of their knowledge, yet favoring correctness over speed. The respective answer could be given in any preferred language⁴. If a question was impossible to answer, prospective participants were made aware not to press anything and wait for the next question. Timings between each question (10 s) and the overall study time (two sessions of 12.5 min each) were provided on the information sheet. The available answer time of ten seconds per question was conservatively chosen, preventing unnecessary pressure due to lack of time. After providing informed consent, the participants were asked to complete a demographic questionnaire, polling sex, age, work field, highest educational qualification, their native language and their eyesight. They also rated their reading level for specific languages based on the provided CEFR's self-assessment test⁵ ([51], pp. 26-27). Additionally, the participants were asked to provide their proficiency level for languages that were not listed. The experimenter was present to answer any questions that might arise.

Before starting the experiment, the eye-tracking device was calibrated using a five-point calibration. We only accepted calibration accuracy below one degree of visual angle. During the experiment, we sequentially displayed 150 questions to the participants, in random order. Each question was visible for exactly ten seconds and could be answered by a single key press, such as one letter or one number (see Table 3.1). Participant keystrokes were collected during this period. The experiment was conducted in two sessions of 12.5 minutes with an intervening rest period. It was possible to ask questions during the experiment, which occurred two times. The respective experiment question was hence marked for deletion. Before resuming with session two, another calibration analog to the start of the experiment was performed. Ethical approval for this study was obtained from the Ethics Committee at the University of Constance.

Post-Processing

We applied the following post processing steps to the obtained eye-tracking data including event detection using a velocity-based fixation algorithm [256] with a velocity threshold of 35 degrees of visual angle per second and blink detection based on pupil diameter change. Eye movement events were allocated to the respective question and any data after the point

⁴ E.g. answering in one's native language or in English.

⁵ If needed. Most participants were familiar with the framework and provided information based on language tests.

in time when the participants provided an answer for the respective question, as indicated by the recorded keystrokes, was discarded. Note that we did not distinguish between wrong and correct answers, since language proficiency as stated by the participants was used as ground truth. We observed a mean answer time for all participants of $\bar{x} = 4.9 s$ with a standard deviation of $s = 0.2 s$. Thus, we conservatively limited the overall observation time for each question to a maximum of $4.5 s$. Additionally, we discarded fixations with a duration of less than $50 ms$ and more than $600 ms$, that were outside the range of reported values in reading research [245, 247, 283].

Measures

All measures were collected on a per question basis. We derived the following metrics from the eye-tracking data. For fixations, we examined the average fixation duration. We also recorded refixations that occurred when the user's gaze revisited the location of a previous fixation. This area was bounded by a 30 pixel radius⁶ around the previous fixation's location. The chosen radius is a conservative interpretation of the distance of consecutive fixations during reading [245]. The amount of refixations was normalized by the total fixation count for the respective question, yielding the refixation ratio. For blinks, we calculated average duration as well. Total blink ratio relates the total blink duration to the respective answer time of each question. All the measures were grouped by the reported language proficiency based on the CEFR [51]. Hence, we end up with a total of seven groups: non-proficient (X), basic user (A1, A2), independent user (B1, B2) and proficient user (C1, C2).

Statistical Results

As mentioned before, we looked at data available after having interacted with the screen for a maximum of $4.5 s$ or the user's respective answer time, whichever was shorter. Sample size did vary for different metrics, e.g. if the user did not blink during one particular question.

Average Fixation Duration (AFD)

The grand mean of average fixation duration (AFD) was $\bar{x} = 239.86 ms$ ($s = 53.22 ms$). Participants with C2 proficiency in the tested language exhibited the shortest average fixation duration ($\bar{x} = 194.02 ms$, $s = 37.24 ms$) while no knowledge of a given language produced the longest fixations ($\bar{x} = 249.19 ms$, $s = 53.04 ms$). As all metrics in our experiment are unevenly distributed in terms of proficiencies (participants have varying levels of proficiency in different languages), we decided to compute the slope coefficient for each of the participants using ordinal logistic regression. We then conducted a t-test comparing the obtained slope coefficients with a constant function. The test shows that language proficiency has a significant effect on AFD, $t(20) = -3.78$, $p < 0.01$. Figure 3.2 shows the distribution for average fixation duration.

⁶ Covers roughly 1-2 letters.

Refixation Ratio (RFR)

The grand mean of refixation ratio RFR was $\bar{x} = 0.24$ ($s = 0.15$). Participants with C2 proficiency in the tested language exhibited the lowest refixation ratio ($\bar{x} = 0.15$, $s = 0.13$) while participants who did not know the given language had the highest ratio ($\bar{x} = 0.26$, $s = 0.15$). Analogously to AFD, slope coefficients were computed and compared with $y = 0$ using a t-test. The test showed that language proficiency had a significant effect on RFR, $t(20) = -2.48$, $p < 0.05$. Figure 3.3 shows the distribution for refixation ratio.

Average Blink Duration (ABD)

The grand mean of average blink duration (ABD) was $\bar{x} = 413.47\text{ms}$ ($s = 395.52\text{ms}$). Participants with C2 proficiency in the tested language exhibited the longest average blink time ($\bar{x} = 752.05\text{ms}$, $s = 391.00\text{ms}$) while participants who did not know the given language blinked the shortest ($\bar{x} = 306.41\text{ms}$, $s = 332.29\text{ms}$). Again, we calculated slope coefficients for each participant and used a t-test to compare with no slope yielding a result of $t(20) = 2.71$, $p < 0.05$. Figure 3.4 shows the distribution for average blink duration.

Total Blink Ratio (TBR)

The grand mean of total blink ratio (TBR) was $\bar{x} = 0.15$ ($s = 0.14$). Participants with C2 proficiency in the tested language exhibited the highest total blink ratio ($\bar{x} = 0.30$, $s = 0.14$) while participants who did not know the given language had the lowest ratio ($\bar{x} = 0.11$, $s = 0.11$). Using a method analogous to the previous metrics, we obtained $t(20) = 2.75$, $p < 0.05$. Figure 3.5 shows the distribution for total blink ratio.

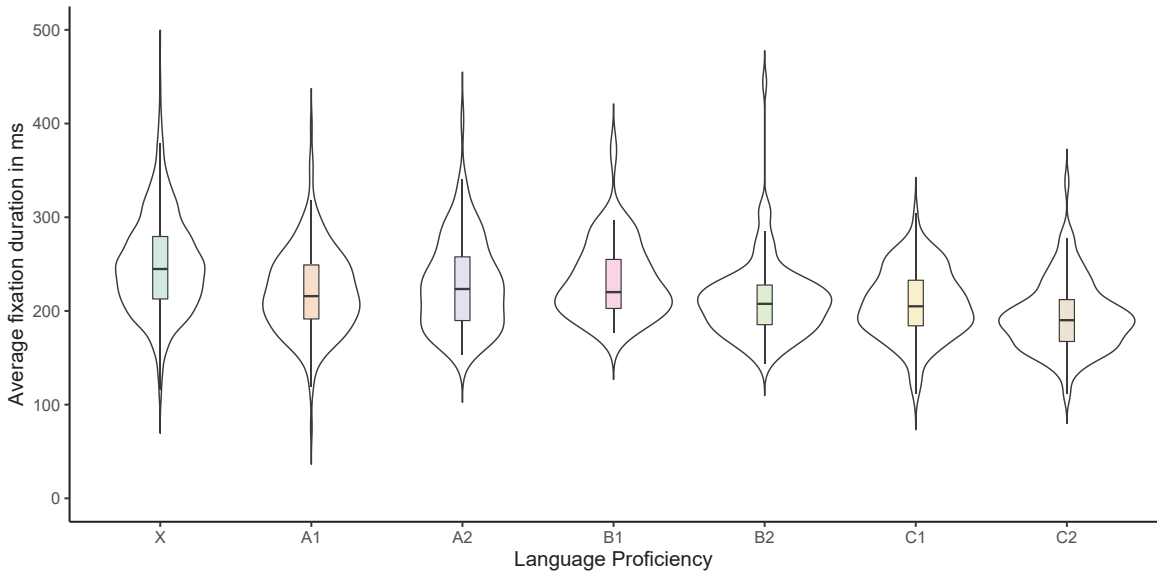


Figure 3.2: Violin plots (width normalized per group) showing the distribution of the average fixation duration (AFD) grouped by proficiency according to CEFR [51]. Density plots depict the sample distribution over the proficiency groups.

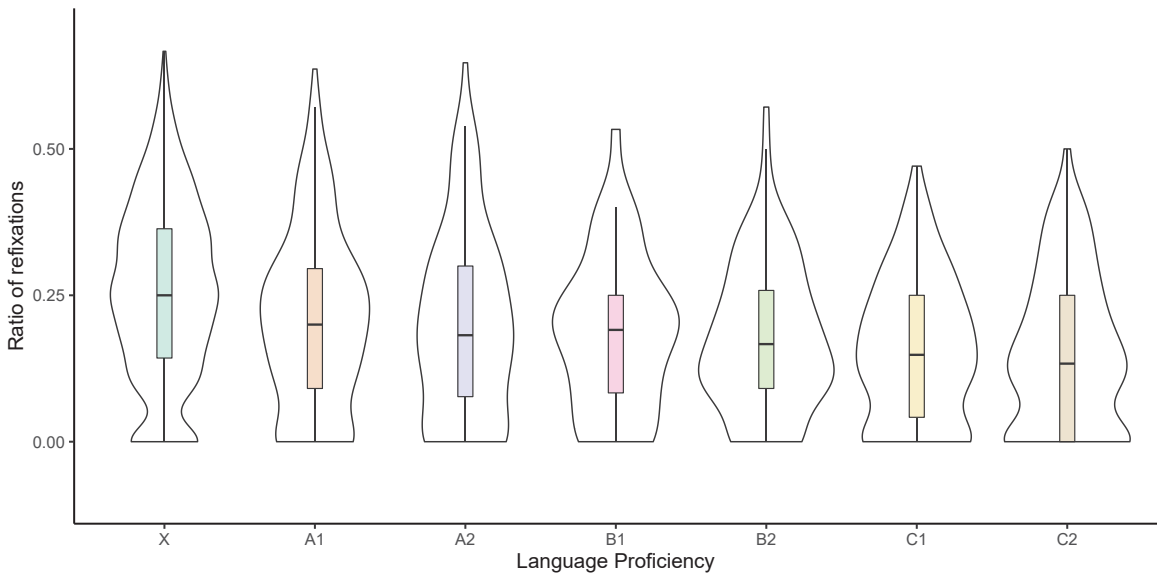


Figure 3.3: Violin plots (width normalized per group) showing the distribution of the refixation ratio (RFR) grouped by proficiency according to CEFR [51]. Density plots depict the sample distribution over the proficiency groups.

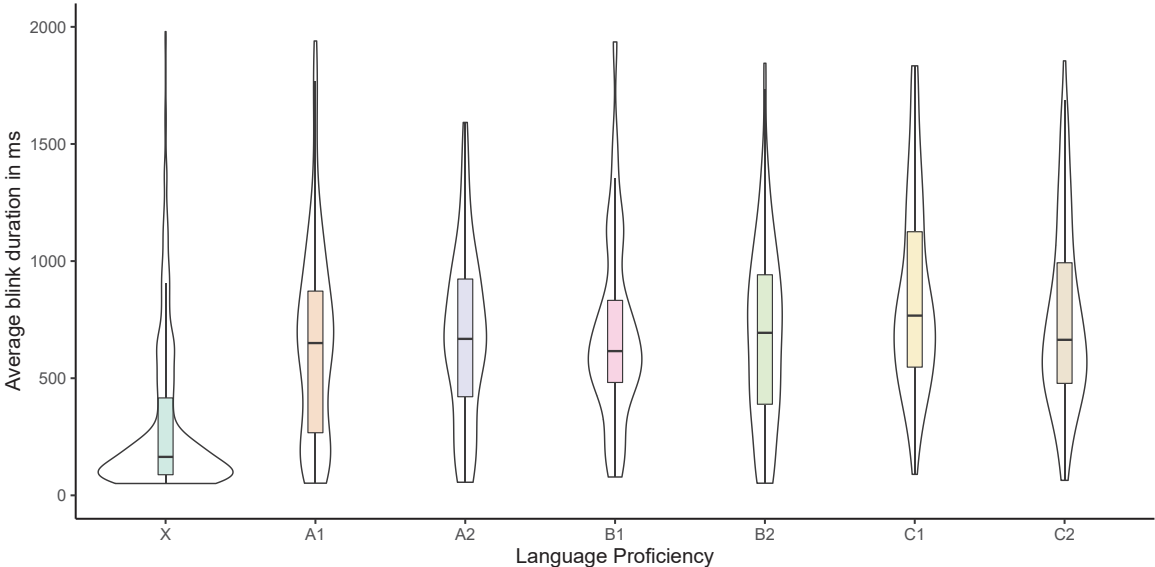


Figure 3.4: Violin plots (width normalized per group) showing the distribution of the average blink duration (ABD) grouped by proficiency according to CEFR [51]. Density plots depict the sample distribution over the proficiency groups.

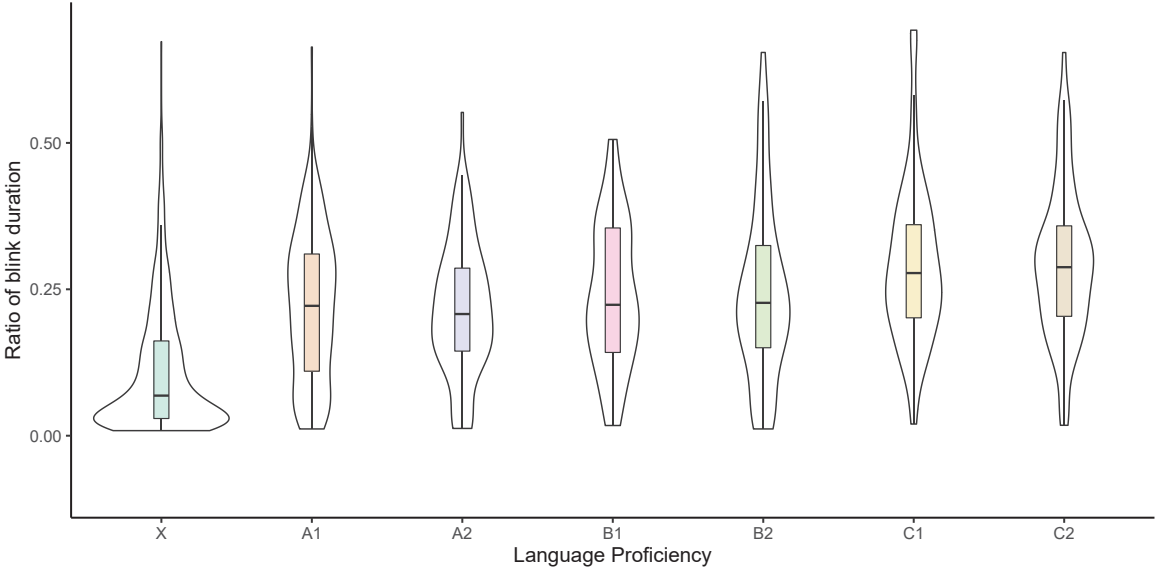


Figure 3.5: Violin plots (width normalized per group) showing the distribution of the total blink ratio (TBR) grouped by proficiency according to CEFR [51]. Density plots depict the sample distribution over the proficiency groups.

Result Interpretation and Implications

Statistics on average fixation duration and re-fixation ratio showed a significant effect of language proficiency. We conclude that increased language proficiency levels will result in a lower AFD and a lower RFR (as t-scores are negative). This confirms **H1** and **H2**. Analogously, our results indicate that language proficiency has a significant effect on average blink duration and total blink ratio as well, confirming **H3** and **H4**. However, standard deviation was quite high for these blink-related metrics, which can possibly hinder proficiency estimation by an interactive system. Especially for proficient levels (A to C), blink duration is quite high compared to values found in literature [120]. This could be due to participants “relaxing” during easy questions, which relates to longer blink duration [120]. Nevertheless, simple blink detection (e.g. via a RGB camera) could be an alternative for systems where eye-tracking is not feasible.

Our statistical analysis indicates that we can provide a positive answer to **RQ1a**. Language proficiency showed a significant effect for all metrics, suggesting the feasibility to employ these metrics in interactive systems to detect language proficiency. In other words, our experiment shows that the eye gaze metrics we chose do carry information about language comprehension.

We further investigated the discriminatory strength of those metrics for interactive systems, tailoring to **RQ1b**. Here, we determined technical requirements for proficiency-aware systems, by implementing classifiers for language proficiency detection based on the analyzed gaze metrics.

Classifying Language Proficiency

In the previous section, we evaluated the statistic relevance of certain metrics extracted from our eye-tracking data. AFD, ABD, TBR and RFR showed promising results in discerning proficiency levels. Yet, to realize systems that support real-time detection of language proficiency it is necessary to find a discriminative feature set that defines certain proficiency levels. Ideally, we want the margin between neighboring levels as large as possible and choose these levels based on the application scenario. In this section, we provide answers to our second research question (**RQ1b**) and evaluate the needed technical requirements for detecting language proficiency in proficiency-aware systems.

We highlight specific factors that need to be considered when implementing such an adaptive system. From an HCI perspective, we want to maximize the user experience when interacting with the system such as providing assistance when needed and only when needed. Furthermore, an interactive system should be able to provide feedback in a reasonable time span. Ideally, it should not take longer than a few seconds for the system to make a decision. From a technical standpoint, we might need to consider limitations that may arise with the application scenario, such as low-cost hardware that will result in noisy data. As such we tested our findings on an artificially downsampled dataset.

Baseline Classifier

Based on our previous results, we decided to train a binary classifier that predicts whether people are not proficient or proficient (levels A to C) in the displayed languages. This choice conforms with our aspired application scenario in two ways. Firstly, public displays such as timetables exhibits “predictable information” [51] that can be read by a basic language user (A level proficiency). Thus, the choice to include all proficiency levels into one class is reasonable. Secondly, overeager and unpredictable assistance systems tend to be rejected by the users [88]. To achieve a good user experience, minimizing false proficiency classification for proficient users should therefore be a primary goal.

We constructed a dataset containing the following features from our eye-tracking data as described in the previous sections: average fixation duration (AFD), refixation ratio (RFR), average blink duration (ABD) and total blink ratio (TBR). We used the same maximum observation time as before (4.5 s). After balancing the class distribution⁷, we evaluated the dataset using 10-fold cross validation on four different common classifiers⁸. All classifiers were executed with standard parameters. An overview on achieved accuracy is given in Table 3.3.

	J48	SVM	NN	BayesNet
Accuracy	78.3%	64.9%	78.8%	77.6%
F-Measure	78.3%	64.3%	78.8%	77.6%

Table 3.3: Accuracies and F-Measures for selected classifiers.

While most classifiers exhibited a similar average accuracy, it is vital in our case to report the false positive rate (FPR) of each class. As mentioned before, our primary goal was to keep the ratio of falsely classifying proficient users (classified as non-proficient) to a minimum, while our secondary goal was to maximize true positive rate (TPR) of either class. Table 3.4 shows more detailed descriptive statistics on the decision tree J48, including false and true positive rate for both classes. This allows us to spot irregularities that may arise between classification accuracy of the two classes.

True positive rate was higher for the class of proficient language users (class P), yet this came with a higher false positive rate as well. In other words, in 24.6% of cases we wrongly classify a non-proficient user (class NP) as being proficient. Of course this is inconvenient for the user as he will not be provided with the needed assistance. However, our primary

⁷ Assuming equal amounts of non-proficient and proficient users.

⁸ J48: Java derivative of C4.5 [240]

SVM: LibLINEAR package. L2-loss (dual form) [77]

Neural network: MultiLayer Perceptron [104]

BayesNet: Bayesian network [104]

Class	TPR	FPR	Precision	Recall
NP	75.4%	18.8%	80.0%	75.4%
P	81.2%	24.6%	76.8%	81.2%

Table 3.4: Classifier statistics on J48 given proficient users (class P) and non-proficient users (class NP).

goal was to keep the false positive rate of the class of non-proficient users to a minimum. In the standard configuration J48 misclassified 18.8% of all instances as non-proficient when they were in fact instances of proficient users. Hence, the system would attempt to provide assistance although the user would be able to read the given language, which is undesirable and has a negative impact on user experience.

Figure 3.6 illustrates the top three levels of the resulting tree model⁹. Average blink duration is used as a first separation step. After just two tests (on ABD and AFD) almost 32%¹⁰ of instances are correctly assigned to the non-proficient class (3.6% are wrongly classified in this step). The built tree exhibits several of these heavy leaves that carry most of the instances. This indicates that pruning may be a meaningful strategy.

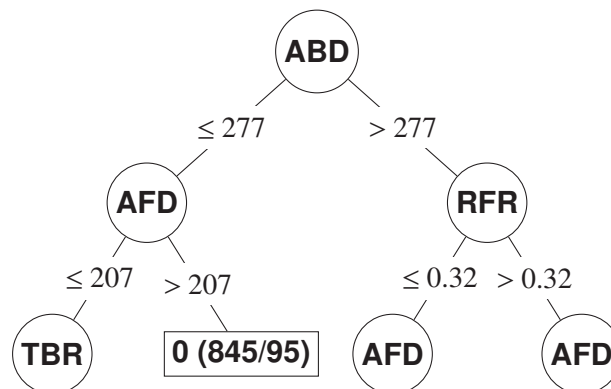


Figure 3.6: Top three levels of the built decision tree using J48 with standard parameters. ABD and AFD are given in milliseconds; TBR and RFR are given as ratios. Leaf nodes (rectangles) indicate predicted class (0 for NP; 1 for P) and show (correctly classified/misclassified) instances. Values are rounded for visual clarity.

A more heavily pruned tree is shown in Figure 3.7, basing its decision on only three remaining features (ABD, RFR and AFD), while still achieving 77.64% accuracy. Additionally, true and false positive rate of each class do not change by more than two percentage points, indicating that heavy pruning is indeed a valid option. This avoids overfitting to the training set and may generalize better for unseen data.

⁹ Using J48 and standard parameters.

¹⁰ 845 out of a total of 2643 instances.

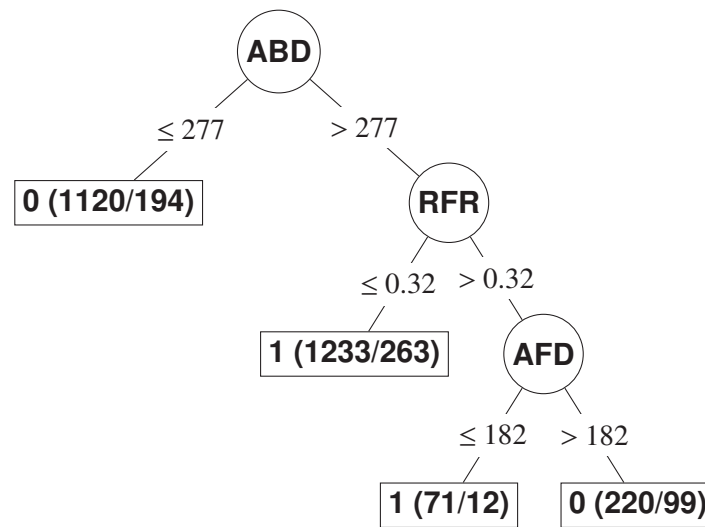


Figure 3.7: Complete decision tree using J48 and heavy pruning. ABD and AFD are given in milliseconds; RFR is given as ratio. Leaf nodes (rectangles) indicate predicted class (0 for NP; 1 for P) and show (correctly classified/misclassified) instances. Values are rounded for visual clarity.

Cost-Sensitive Classifier

To further reduce the false positive rate of the non-proficient class, we introduced a different cost function for the classifier. We penalized classifying proficient users as non-proficient more severely. This is a compromise between maximizing overall classification accuracy (results shown in Table 3.3 and Table 3.4) and minimizing the false positive rate of one class. We decided that not more than one out of ten users should be wrongly classified as non-proficient. To achieve this goal, misclassifying a proficient user had to be three times more expensive than misclassifying a non-proficient user. Table 3.5 shows the results after applying such a function.

Class	TPR	FPR	Precision	Recall
NP	58.9%	9.90%	85.6%	58.9%
P	90.1%	41.1%	77.1%	74.5%

Table 3.5: Classifier statistics on cost-sensitive J48 (standard parameters) given proficient users (class P) and non-proficient users (class NP).

Overall accuracy dropped only slightly to 74.5%, yet the gap between the true positive rate of both classes increased. While TPR of the proficient class increased, the TPR of the non-proficient class decreased to 58.9%. It was now more expensive to classify as non-proficient, hence it was safer to classify as proficient. The same applied for the false positive rates. We successfully pushed the FPR of the non-proficient class to below 10%, yet this also

meant that the FPR of the proficient class increased and fewer users were to get the needed assistance.

Class Distribution

Generally, it is important to adjust the cost function based on the scenario and think about how to penalize wrong classifications. Besides the cost function, the distribution of instances into respective classes is vital. In a different scenario, such as displaying advertisements on public displays, it is more favorable to emphasize highly proficient users. Since using puns in advertisements to spark the interest of potential customers is a common technique, aiming for a binary classifier that separates C-level language users from other users is more beneficial. If the users do not understand the display pun due to their proficiency level being too low, it seems reasonable to combine them with the non-proficient users. Table 3.6 shows classification results for a grouping into the classes highly-proficient HP (C1 and C2) and less-proficient LP (X, A1, A2, B1 and B2). Overall accuracy was 79.3%. The false positive rate for both classes was approximately 21%. Again, we can ask ourselves whether misclassifying one class is more severe than the other and adjust a cost function as we see fit.

Class	TPR	FPR	Precision	Recall
LP	79.2%	20.6%	79.3%	79.2%
HP	79.4%	20.8%	79.2%	79.4%

Table 3.6: Classifier statistics on J48 with different class distribution: highly-proficient (class HP: C1 and C2) and less-proficient (class LP: X, A1, A2, B1 and B2) users. Levels according to [51].

Recording Duration

As outlined previously, fast responsiveness and interactivity are vital for an adaptive information display. Hence, we evaluated a minimum recording duration that allowed for a sensible classification of non-proficient users. For this purpose, we limited the maximum observation time for our dataset in one second intervals from one to five seconds¹¹. Figure 3.8 shows accuracy as well as true and false positive rate over time.

The classifier metrics did not change much when reducing the maximum observation time to only three seconds. After that, we saw a strong increase in the false positive rate of class NP. Further investigation revealed that the classes exhibit a strong overlap, thus making accurate classification difficult. We believe that this is due to the reduced sample size, hence allowing noise in the recording to have a larger influence.

¹¹ Before then our dataset used a maximum observation time of 4.5 seconds.

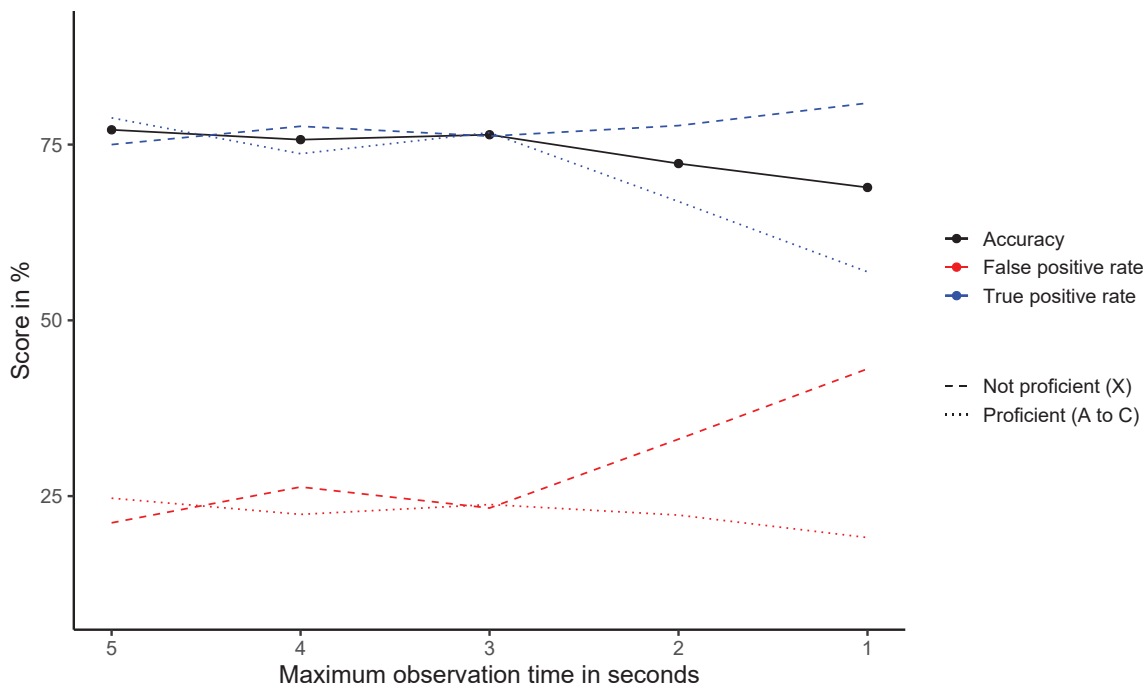


Figure 3.8: Accuracy, true and false positive rate of proficient (class P: A1 to C2) and non-proficient (class NP: X) class given different maximum observation times of recorded gaze data.

Sampling Rate

In this study, we used a high-quality eye-tracking device with a sampling rate of 250 Hz . However, eye-trackers for commercial product integration can be expected to have lower sampling rates. To evaluate the viability of our approach for its stated purpose, we down-sampled¹² our recording data to the following sampling rates 125 Hz , 62.5 Hz and 31.25 Hz , which approximated the sampling rates of low-cost eye-tracking devices.

Reducing sampling rate introduces noise in the event (e.g. fixation, saccades) detection of eye-movement behavior. Here, we showed the effects for the average fixation duration. A classification accuracy of 64.7% was achieved from the original sampled data (i.e., 250 Hz). Figure 3.9 illustrates how classification accuracy, as well as true and false positive rate for each class, varied when decreasing the sampling rate.

Sampling at a rate of 125 Hz had only minor influence on the classification results. Further downsampling drastically reduced true positive rate of class NP (FPR of class P increases), yet the false positive rate of the non-proficient class stayed constant. Hence, at a sampling rate of 62.5 Hz detecting a non-proficient speaker correctly was less likely. For the lowest sampling rate 31.25 Hz , FPR for the non-proficient class strongly increased. At this rate, a sample was obtained every 32 ms , which was about the mean difference that we found between the two classes (NP and P) at high sampling rates. Thus, temporal downsampling

¹²By omitting every n-th sample.

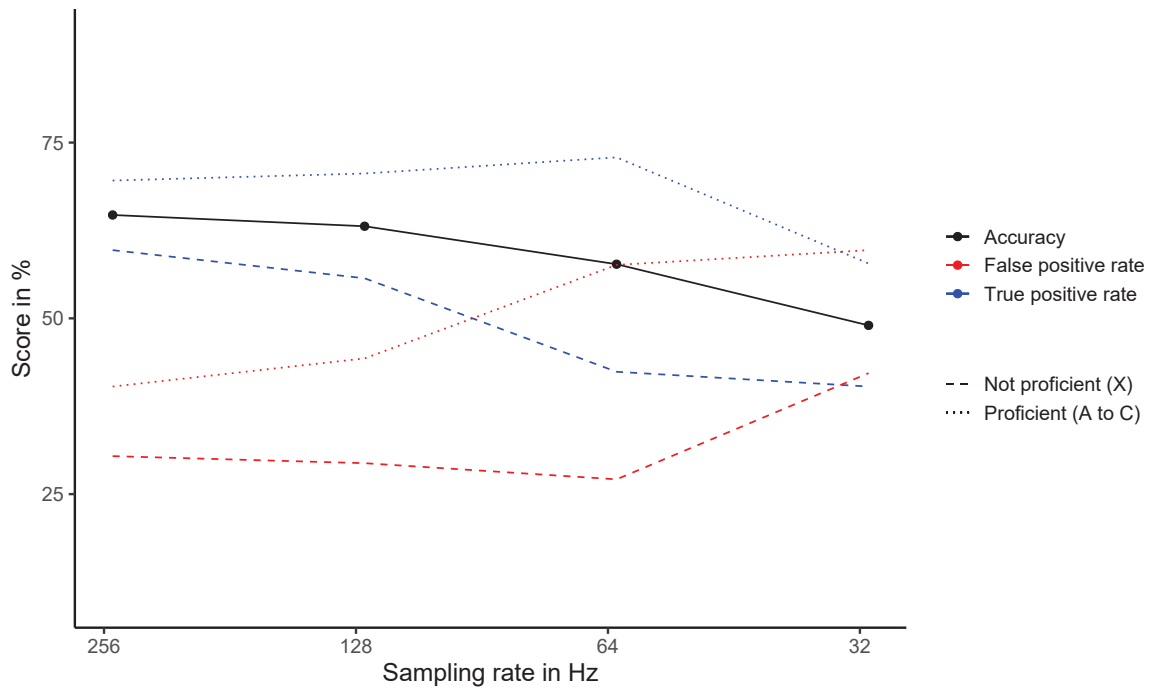


Figure 3.9: Accuracy, true and false positive rate of proficient (class P: A1 to C2) and non-proficient (class NP: X) class given different sampling rates of the underlying gaze dataset (logarithmic scale).

rendered information between categories less discriminable and sensible classification was not possible at this resolution.

3.2.3 Discussion

Through the initial statistical analysis and our investigation of how to build classifiers using our data set, we have shown that rapid automated detection of language comprehension is possible. The data we extracted from the user’s gaze enables us to reliably predict their proficiency in a given language. It needs to be noted that our proposed classifiers operate on a coarse level. Their primary purpose is to determine whether the user is able to understand the interface presented in a given language. More sophisticated methods and larger gaze pattern samples are still required to estimate exact proficiency levels, as shown in previous work [199, 335].

We can conclude that the answer to **RQ1b** is positive — it is possible to build a reliable automated system that will promptly detect whether a user is able to comprehend the content presented in a given interface. An emergent question is how these systems can be used in practice and what technical prerequisites are needed for deployment. Below, we further elaborate on **RQ1b** by providing a set of technical guidelines for implementing proficiency-aware systems that want to make use of gaze data.

Adhere to Required Granularity Levels of Proficiency

Adaptation and proper specification of provided and expected proficiency levels is essential, as outlined in Chapter 2. In this analysis, we depicted how to tailor to the needs of different application scenarios by adapting the cost function of the classifier as well as changing the class distribution. The application scenario highly influences the classification strategy, as we have shown when grouping different proficiency levels for binary classification. Fine-grained detection down to the proficiency levels of the CEFR [51] did not deliver satisfying accuracy. Cost functions can be introduced to better fit the use case scenario. Here, a compromise has to be made between a more eager assistance system and a more rigorous one. Depending on the application scenario, one might opt for either direction. Besides adjusting the cost function, changing the class distribution is also a possibility, e.g. when considering advertisement banners instead of public displays in an airport. Hence, determining whether a user is able to use an interface in a given language from gaze properties is entirely possible. Achievable accuracy depends on the application scenario and the selected proficiency levels. If fine-grained detection and high accuracy is needed, e.g. in an language education software that observes the learning process of a student, other sources of context should be used in conjunction, while eye-tracking can provide an auxiliary input. Consequently, a proficiency-aware system should consider additional probabilistic priors. In this regard, we feel that modifications to our current classification approach (e.g., by adding a "location prior"¹³) holds promise for subsequent work.

Consider Temporality of Proficiency Detection

For ad hoc, gaze-based proficiency-aware systems that require fast and immediate detection without relying on past gaze data from the user (cf. Section 2.4), we recommend a minimum interaction period of three seconds and a sampling rate not less than 100 Hz, while lower rates might still be applicable for specific scenarios. In our study, we evaluated how long and how often a system should record eye movements of interacting users. Higher resolution or longer recording duration only slightly improved classification results. These findings confirm the feasibility for ad hoc proficiency-aware systems based on gaze metrics, as "detection time" is short enough for user interaction and sampling rate is not too high for low-cost eye-tracking devices.

Robust Gaze Features for Proficiency-Aware Systems

While more sophisticated algorithms (e.g. scanpath analysis) could potentially results in a higher detection accuracy for proficiency level, this use case present robust gaze features based on metrics derived from fixation and blink data. Making use of exact gaze position would require a per-user calibrated eye-tracking device. Here, new methods on implicit calibration relying on predicted gaze targets [152] or smooth pursuit as shown in TextPursuits

¹³Emphasizing predominant languages at the system's deployment location.

[158] are applicable. If access to fully calibrated gaze data is available, more elaborate comparing methods are possible, such as directly comparing scanpaths [79, 136] or clustering approaches [259].

Detecting blinks and fixations — as used in our metrics — is feasible using only relative gaze data, thus omitting a per-user calibration of the eye tracker. Further, the low sample size of collected gaze data during user interaction favors simple machine learning algorithms. Moreover, if gaze behavior is predictable based on screen content, constraints on calibration accuracy are less severe. Hence, it is well suited to be used in public displays where "walk-up" interaction is required [211]. Additionally, it enables the usage of less accurate and cheap eye-tracking devices as exact gaze position is not necessary.

Limitations

In our study, we used a set of languages that reflected the possible linguistic abilities of our participant base. We mostly selected languages that are spoken in Europe. However, we believe that as long as reading direction and alphabet are kept the same, the results of our study are reproducible with other languages. Even reading direction might have only a minor influence [245]. Statistics on whether the alphabet was a confounding factor (e.g. Arabic, Greek) have shown that there was no significant difference on the four metrics discussed previously. Nevertheless, this was a subjective result based on the subset of our used languages.

Another limiting factor was the uneven distribution of proficiency levels. Naturally, we had a lot of data samples for non-proficient users, but missed out on proficient levels. B1-level proficiency was especially rare among the participants.

Based on the application scenario, participants might react differently. For example, when on a stressful trip figuring out their next connection on an information display versus reading questions during a calm lab study. This will likely introduce another noise factor that influences the resulting data. The nature of this noise and its impact in an in-the-wild system is to be evaluated. Consequently, before deployment, extensive pre-testing needs to be conducted for a particular interface to determine what comprehension levels are required in a given interaction scenario. Here, proficiency-aware systems can make use of the user's context (e.g., location priors) as outlined previously.

3.2.4 Conclusion

In this use case, we examined the perception of different languages and their influence on a person's eye movements. Contrary to common approaches in reading comprehension, which focused on post-hoc evaluations, we evaluated the feasibility of real-time language proficiency detection suitable for language-aware interfaces.

In our study, we presented 150 questions in different languages to our participants and analyzed the effect of language proficiency on certain gaze characteristics. We found that proficiency had a significant effect on the average fixation duration and refixation ratio as

well as average blink duration and total blink ratio measured over a maximum timespan of 4.5 seconds. To effectively utilize these results in real-time adaptive information displays, we proposed a classification scheme that realizes language proficiency detection and recommended technical guidelines on recording duration and sampling rate of the eye-tracking device, contributing to **RQ1**.

While this use case focused on an ad hoc user interaction and relatively simple visual stimuli, our next two use cases will focus on prolonged interaction and complex tasks.

3.3 Use Case II: Assessment of Programming Proficiency using Gaze Properties

Due to the increasing demand for coding skills in academia and industry, evaluating the programming proficiency of a possible employee or student is a vital step towards acceptance. Yet, objective models to evaluate programming proficiency have scarcely been explored. In academic settings, we rely on assessment scores as an indicator of an individual's ability to code. In this use case, we describe a study using eye-tracking to measure programming proficiency while considering typical constraints that such a setup induces. This includes the necessity of a readily available setup that is robust and reliable regardless of the environment.

The application of using eye-tracking to evaluate programming proficiency in school settings has broad implications. Through this, teachers can develop an insightful understanding of each student's progress. This enables educators to foster specific learning modules, provide additional support, or modify teaching plans according to the overall progress. Moreover, eye gaze data can be recorded in a non-intrusive way, obviating the need for body-worn devices. Eye-tracking can, therefore, be conducted pervasively by deploying a remote eye tracker together with a display. The continuous retrieval and processing of eye gaze can be evaluated more frequently during the exercises themselves rather than having single exams at the end of courses. Furthermore, user interfaces can provide implicit assistance by using proficiency as a metric. Individual abilities can then be specifically fostered to expand the skill set of the user.

In this use case, we investigate source code as an intermediate textual representation. Compared to the previous use case, understanding the structure and the sequence of text elements is now of importance. Consequently, the complex task makes it harder to accurately assess proficiency (cf. Section 2.4).

3.3.1 Methodology

Research has identified self-reported scores [78, 162] and past performance [5, 34] as the best indicators of programming proficiency. In this use case, we introduce eye movements

as an additional factor. An extensive overview of eye-tracking usage in the domain of computer programming is provided by Obaidallah et al. [219]. In terms of programming proficiency, researchers have, for example, leveraged eye tracking in combination with customized source code visualizations to test programming proficiency [19] and investigated gaze pattern during a programming task [220].

In this work, we have opted to evaluate participants' program comprehension. Again, we favor robust gaze metrics in our analysis as they are well-suited for proficiency-aware systems. We further investigate occurring gaze patterns and their relation to the participants' programming proficiency. Based on **RQ1** as reiterated below, we formulate one related research question tailored towards this use case:

RQ1

What are suitable metrics inferred from a user's gaze for proficiency estimation?

RQ1c: Can we determine a user's programming proficiency from gaze properties?

For the study, we adapted twelve programs from online beginner courses and lectures¹⁴. We chose to focus on Java as a programming language as it is widespread and commonly learned by programming beginners¹⁵. To determine the difficulty of our programs, five experts with an academic background in computer science¹⁶ rated each program on a three-point scale. The average score determined the program's difficulty. An overview of the used programs depicted by a short description in ascending difficulty order can be seen in Table 3.7.

Procedure

Participants first signed a consent form and provided their demographic data. After calibrating the eye tracker (SMI Red 250), participants were shown a simple training task (sum calculation) to familiarize themselves with the task procedure. The actual experiment consisted of twelve programs (cf. Table 3.7) of varying difficulty with a break in the middle. The order of the programs was randomized for each study. Participants were asked to enter the behavior of each program on the following screen. The sequence for each trial was a blank screen, the program itself, and a text entry box. Only the eye-tracking data for viewing the program was submitted for analysis. Each study lasted a maximum of 60 minutes. Participants were compensated with 10 Euros. Ethical approval for this study was obtained from the Ethics Committee at the University of Constance.

¹⁴<https://www.tu-chemnitz.de/informatik/ST/research/material/fMRI/index.php>

¹⁵www.tiobe.com/tiobe-index/

¹⁶Academic researchers holding a degree in computer science.

Program description	Difficulty
Faculty calculation	1.2
Element-wise array comparison	1.4
Element-wise counting of equal characters	1.4
Checking for containing substring	1.4
Doubling and square root calculation	1.4
Calculation of the powers of two	1.4
Extraction of unique elements	1.6
Array reversal	1.6
Prefix sum calculation	1.8
Binary search	2.0
Greatest common divisor	2.4
Exponential calculation	2.6

Table 3.7: Example programs used in the study given the calculated difficulty.

Participants

We recruited 20 participants from the University of Stuttgart through mailing lists for our study. All participants were students and reported different programming abilities. The data of 14 participants (4 female, 10 male; Age: $\bar{x} = 21.4y$, $s = 1.93y$) was submitted for further analysis.

3.3.2 Results

We measured the assessed performance by grading the participants' answers by three experts¹⁶ using a three-point scale: wrong (0 points), small errors (0.5 points), correct (1 point). A small error was counted when the answer pointed towards the correct one but lacked explanation details. Hence, a maximum of 12 points could be achieved. Our data contained scores between 5 and 11 points ($\bar{x} = 8.0$, $s = 1.7$).

We analyzed eye movements related to fixations and saccades described in past work [134]. However, we were unable to relate these measurements to programming proficiency. This proves the connection to programming proficiency to be more intricate, hence, difficult to grasp using a linear relationship.

The assessed performance showed no significant differences given program difficulty, which we aggregated into two levels based on the score: easy (below 1.5) and hard (above 1.5).

Further, for each trial, we averaged fixation duration up to a maximum of two minutes¹⁷. We have investigated the average fixation duration during the easy and difficult tasks for

¹⁷ Average answer time was normally distributed with $\bar{x} = 126s$, $s = 67s$.

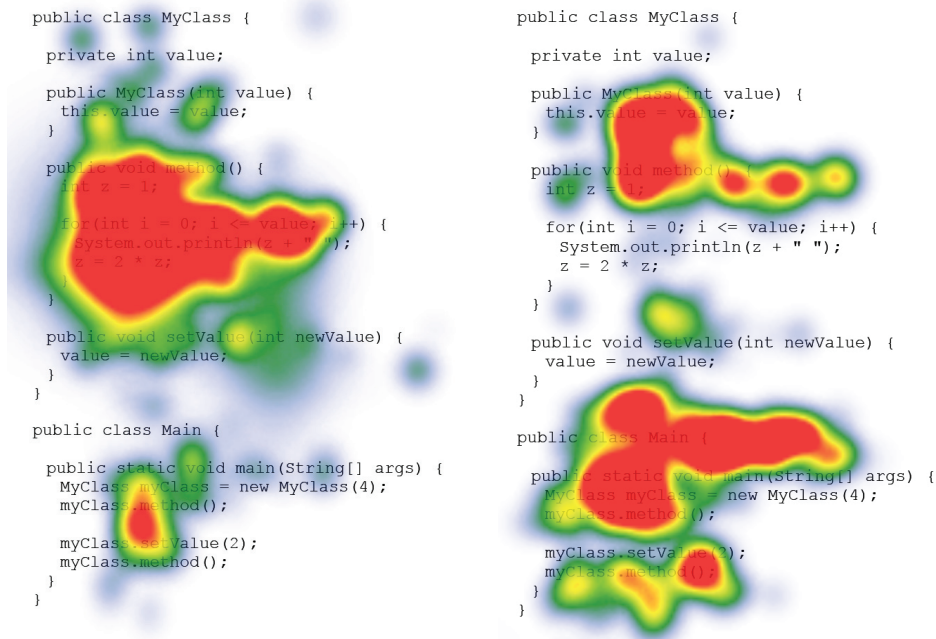


Figure 3.10: The heatmaps of two exemplary participants showcasing different eye movement patterns when solving the same programming task. The choice for a heatmap visualization and omission of AOIs boxes was done for visual clarity.

statistical significance. A one-way ANOVA showed no significant difference between the easy and difficult programming task complexities.

We examined several areas of interest (AOIs), such as method declarations, loops, main classes, and method bodies. However, we did not find a significant relationship between our eye-tracking metrics and the respective AOIs. We also reviewed participants' strategies when looking through each program. Here, we examined their eye movement sequence using the AOIs for each program and found that people of different skill levels used a variety of strategies. In other words, there was no distinct strategy — in terms of eye movements — when solving the task for highly proficient participants. Participants tended to either use a holistic approach by getting an overview first and then looking at individual methods and code fragments. The other main strategy that we identified revolves around on-demand gazes at specific methods without scanning the whole program first. An example of this can be seen in Figure 3.10, where two participants showcase different eye movement patterns solving the same programming task. The pattern is not correlated to their reported proficiency.

3.3.3 Conclusion

In this use case, we evaluated the suitability of eye movement metrics as an indicator of programming proficiency (**RQ1c**). Our investigation has shown several difficulties in deciding on an objective measurement for programming language proficiency that is consistent across

all users. Visualizing the eye-tracking data can help to identify weak problem-solving strategies that may not be obvious based on the participants' answers. In this case, eye-tracking is used as an auxiliary modality within the evaluation process.

Our results also show that common eye-tracking metrics are not suitable to reliably identify a person's programming proficiency. The same strategy used for analyzing reading patterns (cf. [245, 247]) is not suitable for evaluating more complex text visualization, such as structured source code. Moreover, we could not identify distinct problem-solving strategies for either proficient or less proficient participants. The employed strategy is not driven by one's proficiency but rather by personal preferences and experience. However, we believe this provides a vital aspect of education. By visualizing solving strategies, educators are able to identify mistakes — and their source — more efficiently and provide tailored feedback. Even during the instruction process, optimal strategies can be visualized and conveyed. As such, proficiency-aware systems tailored for programming proficiency can play to their strength when employed for the teacher rather than the student.

Through this use case, it is clear that further exploration into measuring programming proficiency is necessary and that there is great potential in this area. An objective and readily-available assessment of one's abilities is essential for quality education. We will pick up this direction in the next use case in the form of a tutoring system for piano players. In contrast to this use case, proficiency estimation is done on a user-dependent basis, warranting more accurate results.

3.4 Use Case III: Gaze as a Predictor for the Proficiency of Piano Players

Playing the piano is a challenging task and mastering new pieces requires extensive practice. Sight-reading score notation is a complex transcription task including aspects of perceptual, cognitive, and motoric processes. Consequently, the user's gaze plays a vital role in assessing the information provided. Gaze-based proficiency-aware systems can utilize this gaze information to provide user-tailored estimations of proficiency accordingly.

We complete our investigation of gaze metrics for proficiency-aware systems with this third use case. We combine insights from the previous use cases in a person-dependent piano tutoring system. Additionally, this task introduces physical task elements and combines them with the cognitive effort necessary to read score notation. Here, we focus on a long-term adaptation towards the user in a complex task domain (cf. Section 2.4).

3.4.1 Methodology

Reading music scores is generally more complex and thus more demanding than text reading [246]. The complex notation, the cognitive transfer to motor commands, and the player's

necessary movements make this a highly complex task [246, 291]. While there is a dictated pattern when reading score notation [188, 323], occasional refixations on already examined parts are possible [239]. Yet, similar to eye movements during text reading, familiarity with the given score notation results in fewer and shorter fixations [188, 227].

Based on **RQ1** as reiterated below, we formulate one related research question tailored towards this use case:

RQ1

What are suitable metrics inferred from a user's gaze for proficiency estimation?

RQ1d: Can we determine the proficiency of piano players from gaze properties?

To allow for a tractable investigation of piano proficiency, we opted to associate playing difficulties (as reported by the participants) as a measure of proficiency. In other words, this work aims to predict difficult passages during piano play by leveraging recorded gaze data.

Procedure

After providing informed consent and calibrating the eye tracker, participants played through three music pieces (counter-balanced) for a total of two repetitions. The pieces were chosen from three intermediate difficulty levels¹⁸. To gather ground truth data, participants self-reported difficult passages by highlighting them on the screen immediately after play-through. The study lasted approximately 60 minutes and participants were compensated with 10 Euros per hour. Ethical approval for this study was obtained from the Ethics Committee at the University of Constance. The setup of the study, including a MIDI keyboard (Doepfer LMK2+), a remote Tobii 4C eye tracker, and monitor with score notation, can be seen in Figure 3.11.

Participants

We recruited six participants (4 female, 2 male; Age: $\bar{x} = 29y, s = 14y$) through personal contacts and participation in related studies. All participants were advanced pianists with, on average, $\bar{x} = 20y (s = 16y)$ of experience.

3.4.2 Results and Conclusion

After eye event detection based on an I-DT algorithm [256], we further processed the gaze events with a line detection algorithm. This step ensured that fixations were robustly mapped to their corresponding bar. Its main idea is based on including a "carriage return" detection

¹⁸Based on PianoBookGuide [192], IMSLP Music Library [237], and Wolters [330].



Figure 3.11: Study setup including keyboard, loudspeaker, and monitor with attached eye tracker.

Dependent variable	df	t	V	p
Average fixation count	32	-9.05		< 0.001
Average saccade (X-Dim.)	32	5.11		< 0.001
Average refixation count			8	< 0.001
Average fixation duration			153	< 0.05
Average dwell duration			10	< 0.001

Table 3.8: Statistical results of paired t-tests and Wilcoxon signed-rank tests (bold: significant).

of the collected gaze data. Finally, a mapping of fixations to bars in the score notation was realized through constraining positional and temporal location¹⁹ of each fixation.

Again we have evaluated common gaze metrics and tested whether there is a significant difference for difficult passages. All features are aggregated and normalized based on sliding epochs. Based on the normality of data, we either conducted a paired t-test or a Wilcoxon signed-rank test. Results are given in Table 3.8 for a selection of gaze metrics.

Our final classification algorithm combined these features in a user-dependent machine learning algorithm with a simple outlier detection based on bar dwell time. This allowed us to simulate different sensitivity levels for the algorithm. The confusion matrices for each of the three levels are depicted in Figure 3.12.

Here the lowest sensitivity level (far right in Figure 3.12) is optimized for few false positives²⁰. However, this setting misses a considerable number of difficult parts; only 29% are

¹⁹A fixation might be assigned a different bar than its x/y coordinates based on its temporal data.

²⁰Classifying a passage as difficult when it is not.

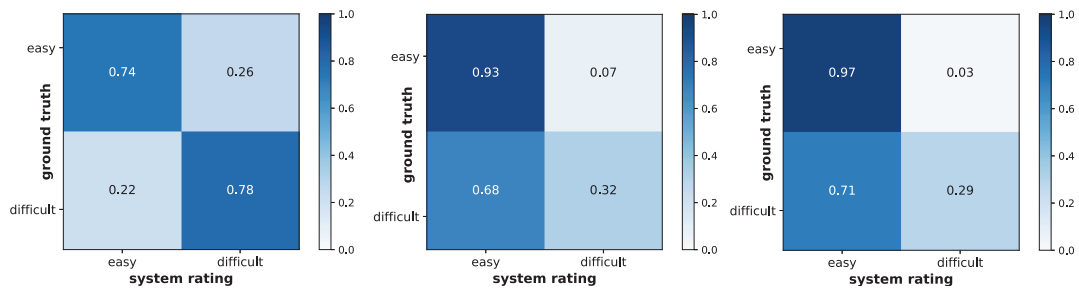


Figure 3.12: Confusion matrices depicting the accuracy of the final prediction algorithm for difficult passages grouped by three sensitivity levels (from left to right: highest to lowest). There is a total of $N = 1230$ samples with 386 difficult parts.

recognized correctly. A higher sensitivity level (far left in Figure 3.12) can classify up to 78% of difficult passages correctly, although this level also increases the false positive rate to 26%.

Our findings have shown that gaze can be a suitable predictor of difficult passages during piano play (**RQ1d**). Difficult regions exhibited significantly different gaze metrics such as higher fixation count and dwell duration. Similar to our first use case (cf. Section 3.2), the performance of a classifier is strongly dependent on the chosen sensitivity level. Nevertheless, a simple algorithm can already detect most piano players’ difficulties, making it a valuable tool in piano tutoring systems.

3.5 Chapter Summary

In this chapter, we have looked at three use cases that varied in task complexity and interaction duration with the user; two of the main dimensions of proficiency-aware systems as outlined in Section 2.4. Our findings have contributed to **RQ1** through technical guidelines, task specifications, and feasibility evaluations. Ultimately, gaze can be a powerful tool for proficiency-aware systems to estimate proficiency but is subject to task constraints that need to be carefully considered.

Achievable accuracy highly depends on the application scenario and the selected proficiency levels. Suppose fine-grained detection and high accuracy are needed, for example, in language education software that observes the learning process. In that case, other sources of context should be used in conjunction, while eye-tracking can provide an auxiliary input.

For coarse-grained detection, gaze can provide robust proficiency estimation across users. Even a short interaction duration allows for at least binary estimation, which might be sufficient for some use case scenarios. Complex tasks, i.e., complex visual stimuli or additional involvement of motoric aspects, justify person-dependent algorithms for accurate proficiency estimation. When a user’s gaze is not solely occupied with information intake for cognitive processes, domain-specific optimizations need to be made (cf. piano play).

Personal strategies and teaching background can potentially govern the top-down analysis of visual stimuli. Here, estimation of proficiency is challenging but identified task-solving strategies could be exploited by domain experts to identify teaching methods (cf. programming proficiency).

Some domains might not offer a straight-forward definition of proficiency, e.g., there is no gold standard or tests usually require extensive examination. However, the research in this chapter has illustrated that there is often a suitable proxy that can be communicated to the user. Ultimately, the user decides how to interpret the given data, a concept that we will pick up again in Part II of this thesis.

In the next chapter, we will focus our attention on another physiological sensing modality: electromyography. Measuring muscle activity can potentially give unique insights for users about their own movements and how to improve the execution of physical tasks.

Chapter 4

Proficiency From Electromyography

When we interact with the physical world, we predominately use our hands to manipulate objects and tools at our disposal. Consequently, proficiency-aware systems should be able to recognize, assess and support manual motor tasks, covering a wide range of application areas from everyday handiwork over fitness exercises to playing an instrument.

This aspect of proficiency-aware systems is not novel. The idea of sensor-based assistance systems to monitor the correct execution of fine motor tasks is prevalent in many areas. Such systems evaluate task execution and can provide assistance if necessary. Essential for any automatic feedback is the correct task recognition by the sensor-based assistance system. For fine motor tasks, a camera-based object segmentation seems natural, as applied by Johnson et al. [138], detecting faulty hand posture during piano play. Yet major drawbacks of camera-based systems are occlusion and privacy; especially for fine motor tasks, self-occlusion occurs frequently. Furthermore, users might also be hesitant when having their training recorded by cameras.

Another modality to monitor body movements is electromyography, which measures the muscle activity of a user. Individual motor tasks exhibit distinct electrical patterns triggered by muscle contractions that can be detected and analyzed [206]. As such, EMG provides deep insights into the working mechanism of motor tasks and allows informed feedback for users.

This characteristic makes EMG a prime candidate for proficiency estimation in scenarios where motor tasks are prevalent, such as activity recognition. In this domain, other prominent modalities are inertial measurement units (IMUs) [279], capacitive sensing [98], and camera-based approaches [18]. As outlined, camera-based systems suffer from privacy issues and occlusion, especially in dynamic scenarios of finer motor tasks. IMUs, capacitive sensing, and EMG all provide high accuracy and distinctiveness, with IMUs usually requiring the least amount of setup time. Here, capacitive sensing requires object modification, and EMG requires applying electrodes. Yet, IMUs struggle to offer additional bodily insights that are not visible by an observer, such as a supervisor. In this regard, both capacitive

sensing and EMG can provide unique insights — for example, applied pressure — that facilitate an understanding of the user of their own performance and ultimately proficiency. Additionally, EMG inherently focuses explicitly on "user alteration", enabling interaction across devices and objects. We believe this to be an important design requirement to enable seamless integration for proficiency-aware systems.

Consequently, we chose EMG as a sensing modality for physical tasks (cf. Section 1.1) and evaluate its feasibility to detect motor tasks in proficiency-aware systems in this chapter. In particular, we leverage EMG to detect muscle activity that occurs naturally during manual interaction. In two use cases, we investigate technical and task requirements of EMG-based interactive systems, guided by **RQ2** as outlined below. In Part II, we will have a closer look at how we can design appropriate feedback for users to gain bodily insights (Section 5.1).

RQ2

How accurately can we estimate proficiency from a user's muscle activity?

Recognizing Guitar Chords

In this first use case, we focus on fine-grained motor tasks by drawing on a guitar tutoring scenario. Here, we leverage EMG to recognize guitar chords played by the participants. An automated system ascertains whether a currently played chord is correct and estimates the user's proficiency. Similar to our research in Chapter 3 (Proficiency From Gaze), we focus on technical requirements of the employed modality. Additionally, we have a closer look at who (system or user) controls the adaptation (cf. Section 2.4).

Assessing Exercise Form

While our first use case investigates static hand postures and their recognition via EMG, this second use case focuses on assessing a user's exercise form during physical activity. Additionally, we evaluate the validity of low-cost mobile EMG devices. While the first use case makes use of a high-grade medical device to record electromyograms, we envision EMG to become more and more ubiquitous in the future.

This chapter is based on the following publications.

Jakob Karolus, Hendrik Schuff, Thomas Kosch, Paweł W. Woźniak, and Albrecht Schmidt. 2018. EMGuitar: Assisting Guitar Playing with Electromyography. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. ACM, New York, NY, USA.

<https://doi.org/10.1145/3196709.3196803>

Jakob Karolus, Annika Kilian, Thomas Kosch, Albrecht Schmidt and Paweł W. Woźniak. 2020. Hit the Thumb Jack! Using Electromyography to Augment the Piano Keyboard. In *Proceedings of the 2020 Designing Interactive Systems Conference (DIS '20)*. ACM, New York, NY, USA.

<https://doi.org/10.1145/3357236.3395500>

Jakob Karolus, Felix Bachmann, Thomas Kosch, Albrecht Schmidt and Paweł W. Woźniak. 2021. Facilitating Bodily Insights Using Electromyography-Based Biofeedback during Physical Activity. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction (MobileHCI '21)*. ACM, New York, NY, USA.

<https://doi.org/10.1145/3447526.3472027>

4.1 Related Research and Background

Research on the relationship between neural activity and muscle contraction dates back to the early 20th century, where Adrian and Bronk [1] utilized needle electrodes to measure muscle activation. In more recent research, EMG has predominately been used in clinical applications such as prosthetic control [117, 230] and detecting muscle diseases [14]. Due to the availability of affordable recording equipment, HCI has taken an interest in EMG as an input modality in the last decade, mainly focusing on auxiliary input methods relying on specific hand and limb gestures to trigger muscle activity. In this section, we will introduce the physical background for EMG recording necessary throughout this thesis and highlight research projects conducted in recent years.

Recording Muscle Activity via EMG

The contraction of individual muscles allows us to direct our extremities and move our bodies. The basic element for this process is the motor unit, consisting of the motor neurons and their associated muscle fibers. To control muscles, the brain sends electric signals to the motor neurons in the spinal chord. These are responsible for innervating the fixed set of their muscle fibers by sending action potentials, activating the fibers via depolarization and thus contracting the muscle. The electric signals originating from depolarization along the muscle fiber generate an electric field which can be measured via Electromyography. [186, 206]

Yet, since each motor unit controls a varying number of muscle fibers, the measured motor unit action potential (MUAP) is a spatial-temporal superposition of the contributions of individual action potentials [186]. This becomes particularly obvious when working with surface EMG (sEMG), where sensor electrodes are placed on the participant's skin. The generated electric signal from a specific muscle fiber needs to penetrate several layers of tissue to reach the sensor electrodes, degrading and blurring - similar to a low-pass filtering effect [186] - the signal in the process. Hence, it is difficult to pinpoint single muscle activity, yet this technique is far more suitable for HCI than invasive needle EMG, where the electrodes are injected directly into the muscle. [206]

EMG signals are measured as the potential between a ground electrode and a sensing electrode. Signal properties are influenced - among others - by anatomy such as the size of the muscle fiber, the distance to the recording site as well as recording equipment. Amplitude values range from $50\mu V$ to $5mV$ [29] warranting sensitive equipment. Yet, this makes EMG inherently noisy as the electrodes pick up electrical activity from other sources, e.g. secondary muscle movement, skin movement and environmental sources such as power line noise. Hence, the raw signal needs to be cleaned beforehand. We will discuss the details of applied filters and preprocessing steps in the respective use cases. [186, 206]

EMG in human-computer interaction

For usage in HCI, it is important to be able to distinguish between different signal signatures. In 1988, Heftner et al. [113, 114] explored autoregressive modelling to recognize EMG signatures. In their work, they mention common shortcomings and issues when working with EMG that hold true today, for example subject dependency and the need for training the muscle movements required for activation.

Facial electromyography

Approaches to implicit recognition examine the detection of emotions by recording facial muscles. Vrana already showed significant differences for EMG signals (and heart rate) in different emotional states such as joy, anger and disgust [320]. In a more recent work, Mandryk et al. [191] utilized EMG among galvanic skin response (GSR) and heart rate to objectively evaluate the emotional state of the user in play environments. Similarly, Perusquía-Hernández et al. [232] have shown that EMG provides deeper insights into facial movements than computer vision, revealing covert behavior. Facial EMG has been used for direct input as well. In TongueSee [341], Zhang et al. have shown the detection of six tongue gestures recorded by electrodes on chin and throat with 94% accuracy. Furthermore, Bourland et al. showed that it is possible to recognize different spoken words using facial electrodes [27]. These can even be silently mouthed, allowing for targeted verbal communication using beamforming speakers.

Intimate Interaction

Costanza et al. exploited an important property of isometric muscle activation, that is continuous activation but no visible movement. Due to the discreet nature of the interaction, participants were able to navigate an audio menu using muscle contraction of the biceps. Observers were not able to detect any movements when the participants were wearing short and long sleeves [49, 50].

Gesture Recognition

When using a limited number of electrodes the produced EMG signal can be quite ambiguous due to multiple muscle fibers contributing to the signal. Hence, research has been using different approaches to counteract this issue. One solution is employing multi-modal approaches combining EMG and other sensors, such as force sensors [203] as well as inertial sensors [109, 228] available in the commercial Myo-Armband¹.

To reliably detect a reasonably sized set of gestures, numerous EMG sensing channels need to be deployed. In their work, Saponas et al. [261] employed a circular array of 10 electrodes (yielding 8 channels) around the forearm to distinguish gestures in four different sets with up to 95% accuracy. Yet they also report a high variance in detection accuracy for the different gesture sets and across participants. Given the casually placed sensors and usage of off-the-shelf machine learning, they conclude the huge potential for muCIs (muscle-computer interfaces) with purpose-built hardware and more sophisticated machine learning methods [261]. In their follow-up work [262], they additionally evaluated the approach in a more natural environment, demonstrating gesture classification even when the hand was already occupied, e.g. while carrying a bag or a mug.

A more extreme approach - regarding the number of sensing channels - was applied by Amma et al. [7]. They managed to distinguish 27 gestures using a high-density grid of 192 electrodes attached around the forearm with up to 90% within-session accuracy. After compensating for electrode shift, the authors reported 75% cross-session accuracy. Their method essentially detected characteristic areas of muscle activity (low or high) and shifted the recorded data, although the technique might be difficult to apply on an inter-person level [7].

An inherent challenge of EMG-based interfaces is engineering user-independent classification algorithms. Works by Kerber et al. [156] and Huang et al. [127] showcased possible solutions to this issue. While the first one relied on a robust classification approach, the latter work leveraged implicit calibration to collect data samples without interrupting the user. Consequently, advances for electromyographic sensing with low-cost, mobile devices are essential, such as follow-up work from Saponas et al. [263] and most recently printable electrodes for rapid fabrication of on-skin interfaces [217].

¹ <https://www.myo.com/>. Product has been discontinued

EMG in Musician-Instrument Interaction

For music creation, biosignal characteristics are used to map music characteristics such as pitch and volume. In a work from 1990, Knapp and Lusted introduced Biomuse [163] a MIDI-controller for musical composition using biosignals. Arslan et al. showed a similar system with bio-orchestra [10].

Since music composition involves many different parameters, researchers have adopted the multimodal approach to enrich interaction. Tanaka and Knapp presented a method to control music devices using EMG and relative positioning as complementary input channels [303], while Donnarumma et al. showcased the combination of EMG and mechanomyography² (MMG), enabling continuous activation through isometric muscle contraction recorded by EMG and triggers through MMG on contraction onset and outset [64].

4.2 Use Case I: EMGuitar - Recognizing Fine-Grained Motor Tasks

Mastering fine motor tasks takes years of time-consuming practice. Commonly, expensive guidance by experts is essential for success. Current computer-based training aids usually use vision-based assistance systems, which suffer from occlusion when monitoring manual tasks. In the domain of music education, researchers have already experimented with different modalities for feedback and assistance. In guitAR [185], Löchtefeld et al. showed an augmented reality approach to depict correct finger position on a guitar fretboard, while MusicJacket [139] employed a vibrotactile jacket to inform users of bad posture during violin play.

The scope of this use case focuses on the feasibility of using EMG to detect fine motor tasks and thus evaluate necessary hardware requirements in an exemplary guitar tutor scenario. We report person-dependent classification of seven guitar chords utilizing a ring of electrodes around the forearm (eight channels) with an F1 score of up to 89%. Based on these results, we present the proficiency-aware system EMGuitar, an interactive guitar tutoring system, which we have evaluated in a study for user acceptance and recognition accuracy.

Regarding the presented dimension of proficiency-aware systems (cf. Section 2.4), we look at the frequency of interaction between user and system as well as the aspect of agency by evaluating the system in an automatic and user-controlled condition. We further contribute technical constraints regarding electrode count and calibration methods and define the domain of tasks suitable for EMG-based proficiency-aware systems.

² The sound produced by the oscillation of the muscle tissue when it extends and contracts [64].

4.2.1 Methodology

Previous research has shown that finger movements and specific gestures can be detected effectively using EMG under different recording scenarios (10 electrodes setup [261] vs high-density grid with 192 electrodes [7]). We investigate if EMG is not only feasible as an explicit input method by invoking defined gestures, but can be used implicitly as well. More specifically, if EMG can be leveraged to recognize domain-specific movements and augment a system's understanding of the user's actions, ultimately leading to an estimation of their proficiency. In this scenario, users do not have to learn the intermediate representation to interact with the system. They are free to execute their main task and the system is able to provide assistance and evaluate their work.

Our investigation in this use case is two-fold, addressing (1) technical feasibility as well as (2) technical and task requirements of EMG-enabled proficiency-aware systems. This allows us not only to report on the suitability of EMG as a proficiency proxy for physical tasks, but also identify limitations with regard to the task context. We operationalize this investigation with two research questions derived from **RQ2**:

RQ2

How accurately can we estimate proficiency from a user's muscle activity?

RQ2a: Can a system distinguish between a given set of complex hand and finger postures using EMG?

We investigate the feasibility of EMG to detect domain-specific motion on a fine-grained level, i.e. individual finger movements and hand postures. For this purpose, we draw on musician-instrument interaction and aim to recognize played chords during a guitar playing session. We believe that playing chords on the guitar is a representative case for many fine motor tasks, because it mainly consists of isometric muscle activation - continuous muscle activity without visible movement - which is recognizable with EMG [206, 303]. Individual finger movement is the result of multiple muscles working together [271], hence, we utilize a ring of electrodes around the forearm to capture the majority of the involved muscle activity.

The change between two chords often results in a complete re-positioning of hand and fingers. As such, different postures do not have a common axis of change, possibly limiting application areas for EMG-based proficiency-aware systems. To address this challenge, we formulate our second research question:

RQ2b: What are technical and task requirements to recognize complex hand and finger postures with EMG?

To this end, we evaluate the requirements of the proficiency-aware system, including necessary detection accuracy and electrode configuration for reliable feedback and task requirements when interacting with the system.

For the purpose of addressing our research questions, we first evaluate the recognition of guitar chords in a constrained setup and apply the lessons learned to our guitar tutoring system EMGuitar before evaluating its performance in a tutoring scenario.

4.2.2 Feasibility Study - Recognizing Guitar Chords via EMG

In a preliminary study, we first evaluated different feature combinations and algorithmic parameters often used in related literature. We tasked participants to play a sequence of guitar chords while we were recording their muscle activity.

Apparatus

We used the BrainVision actiCHamp EEG recorder³ with active surface electrodes to perform EMG sensing. In total, we applied ten electrodes around the participants' left upper forearms. This included one ground (GND) and one reference (REF) electrode, minimizing noise levels. Furthermore, active shielding reduced interference to electrical effects and artifacts due to cable movement. The remaining eight data channels were sampled at 500 Hz , as related work [207] states that most power of the signal is within 5 Hz and 250 Hz . While the actiCHamp system is mainly used for electroencephalography, it can be used to record the electric signals from muscle activity as well.

The electrodes were attached in a ring-like fashion around the arm using adhesive foil rings as shown in Figure 4.1. Each ring contains five electrodes (four data electrodes plus either GND or REF), GND and REF where placed on the inner side of the forearm respectively. Afterwards, each electrode was injected with conductive gel and checked for impedance. While early research suggests placing the electrodes on the muscle belly for best and robust recordings, we follow the approach depicted in the work from Saponas et al. [261]. Contrary to their setup we spaced the electrodes equally around the whole forearm, since it will not rest on a table when pressing down guitar chords. This setup allows to implicitly capture the relative electrode location and helps to pinpoint local muscle activity. Prior to the start of the experiment, electrodes were injected with conductive gel and their impedance was checked.

The recorded data was transmitted via USB to the Recording PC (cf. Figure 4.2) and redirected as a data stream into the local network. On the other side, the Stimulus PC (cf. Figure 4.2) was presenting chord patterns on a 27 inch LCD monitor, while simultaneously sending a marker stream into the network. Both streams were locally time synchronized on the Stimulus PC using the lab streaming layer (LSL) [170] framework and saved to its hard drive. A schematic depiction of the complete setup is shown in Figure 4.2.

³ <http://brainvision.com/actichamp.html>



Figure 4.1: Electrode configuration with two rings around the forearm each consisting of five electrodes.

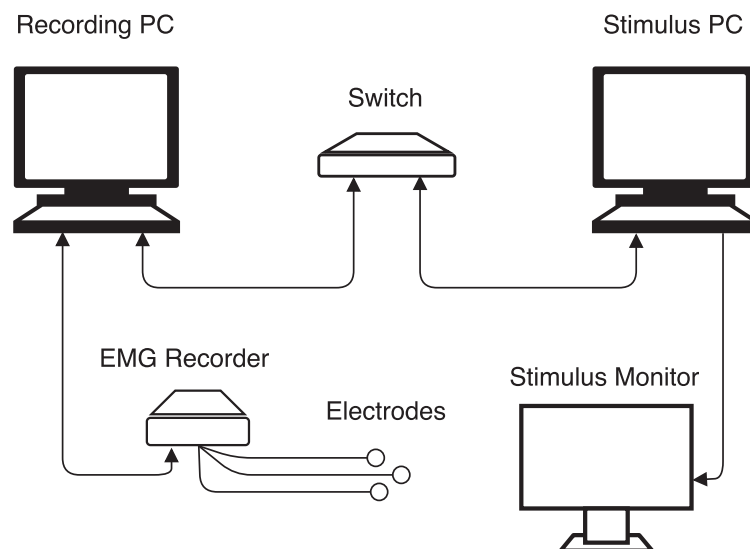


Figure 4.2: Schematic depiction of our apparatus.

Participants

We collected data of three male participants with a mean age of $\bar{x} = 23.3\text{y}$ ($s = 2.1\text{y}$) from the University of Stuttgart through word of mouth. All participants had previously played the guitar as a hobby on a regular basis and were able to play the sequence without flaws.

Procedure

We introduced the prospective participants to our study and gave them time to familiarize themselves with the provided guitar. After providing informed consent, we attached the electrode rings as described in the previous section and fixated the electrode cables with tape if necessary. Subsequently, we asked the participants to play the following sequence

of chords: ⁴, *Am*, *C*, -, *G*, *D*, -, *Em*, *F*, -, *Fm* at a speed of 100 beats per minute (BPM) until they felt comfortable playing it. Each chord was played for one bar (four beat strumming pattern). This selection allows a player to play a wide variety of songs. We played a 100 *BPM* metronome throughout the study to help the participants stay in rhythm. Recording time was five minutes for each subject and total study time did not exceed 20 minutes including setup time. Ethical approval for this study was obtained from the Ethics Committee at the University of Constance.

Data Preparation and Feature Generation

We applied the following filtering steps to clean the signal and remove environmental noise [206, 261]:

1. A fifth-order butterworth bandpass filter between 2 Hz and 100 Hz . This step is important to attenuate long-term drifts, the DC offset and high-frequency noise.
2. A fifth-order butterworth bandstop filter to attenuate frequencies between 49 Hz and 51 Hz in order to remove power line interference.

A common analysis for time-series data is the segmentation into epochs [206], therefore we divided the EMG signal into windows of fixed size for feature generation. Hop size was set equal to window size. We expected participants to change chords at the end of each bar distorting the signal due to the muscular activity during the chord changes, so that the new chord would already be pressed down when the following bar started. Hence, we discarded a portion of the signal at the end of each bar. A depiction of this process can be seen in Figure 4.3. Overall, we discarded approximately 500 ms to 750 ms depending on the used window size, leaving us around 1.65 s to 1.9 s of signal that we submitted for feature generation. The total bar length was 2.4 s .

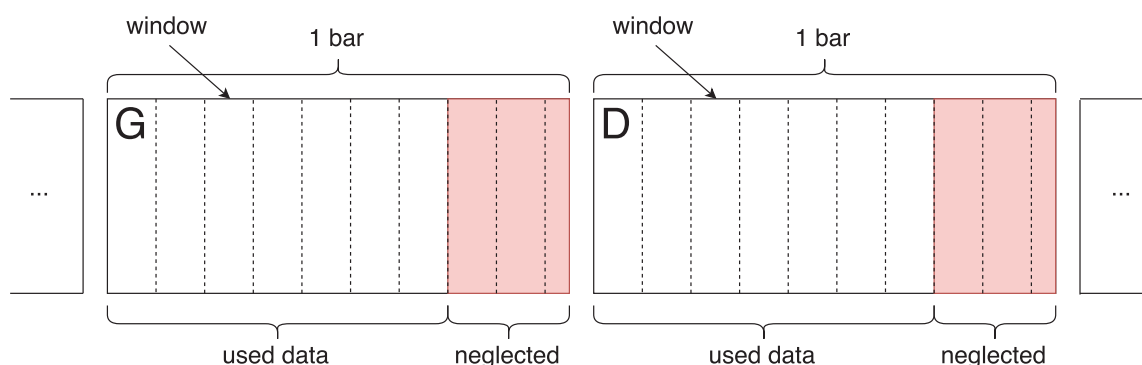


Figure 4.3: Depiction of used and discarded (red) signal parts. One box corresponds to one bar whereas each dotted line marks a new window.

⁴ No chord played.

Based on related work [206, 261, 236], we evaluated the following groups of features:

Root Mean Square (RMS)

The root mean square (RMS) is a commonly leveraged feature when analyzing windowed time-series. For a series $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$, RMS is defined as

$$RMS(\mathbf{x}) = \sqrt{\frac{1}{n}(x_1^2 + x_2^2 + \dots + x_n^2)}.$$

The RMS can be seen as a proxy for the intensity of muscle activity as the amplitude of the EMG signal increases when the muscular activity increases [206]. We computed the RMS for each channel as a separate feature. In addition, we computed the pair-wise ratios between all channels (RMS ratio), yielding a total of 36 values for every window (8 channels plus 28 pair-wise ratios.).

Frequency Energy

A change in muscular activity may not only be reflected in an increased amplitude, but may be present in the signal's spectrum as well. This effect can be attributed to the increased firing rate of the motor units. One possibility to measure this is to take the Fourier-transformed signal and sum up the frequency energy in different bins of frequencies. We divided the spectrum between 0 Hz and 100 Hz into 10 Hz sized bins and use the resulting energy sums as features for each channel (10 in total).

Phase Coherence

Phase coherence indicates the synchronization of motor units. In other words, whether firing happens synchronously (same phase) or asynchronously (different phase). We extracted the mean coherence of a window as well as use pair-wise ratios again, yielding a dimensionality of 36.

Results

We tested multiple combinations of feature sets and will discuss the respective performances in the following.

Train and Test Data Split

We split the derived data set into 80% training data and 20% test data. Since subsequent windows of the EMG signal are correlated (in a probabilistic sense) [206], we did not shuffle the data set before the train test split, as this would have artificially boosted the classification performance [108]. This also matches the real-world approach of calibrating the system before using it. Prior to training our classifier, we normalized (zero mean and unit variance) the data by estimating the mean and standard deviation on the training set. We applied the same transformation to the test set, using mean and standard deviation from the training set.

Feature group	F1 score
Root mean square	0.83
Root mean square ratios	0.83
Frequency bins	0.41
Phase coherences	0.59
RMS and RMS ratios	0.87

Table 4.1: Weighted class-averaged F1 score for given feature groups (averaged over all participants).

Normalization is important as it scales feature values and reduces rounding errors due to limited precision.

Model

We used the training data to train an support vector machine (SVM). To estimate the hyperparameters, we performed a group k-fold cross validation. This type of cross validation matches our data’s constraints as it ensures that in each iteration of the CV, instances of one group, i.e. one bar/ chord, stay together during the k-fold, hence not artificially boosting the classifier’s performance [108]. Throughout our analysis, we used $k = 3$ as higher values induced unacceptably high computation times for fitting the model.

Feature Performance

We analyzed the model’s performance with regard to prediction accuracy and necessary computation time. First, we show how well the model performs using different sets of features. For this analysis, a fixed window size of $250ms$ was used. The results are shown in Table 4.1. We only report F1 scores for visual clarity. Measured precision and recall did not show outliers for any set.

Our results show that the combination of RMS and RMS ratio features outperformed all other tested combinations of features in terms of recognition. We additionally evaluated other feature groups found in related work such as ARIMA and Wavelet transform coefficients [242]. However, those did not yield sufficient F1 scores. For clarity, we omitted these from further analysis.

To get an overview over the strengths and weaknesses of our classifier regarding specific chords, we report the confusion matrices of the models for each participant in Figure 4.4.

Feature Computation Time

In order to find out how well the classification could be embedded into a real-time setting, we measured the processing time of each feature group (averaged over the participants) on

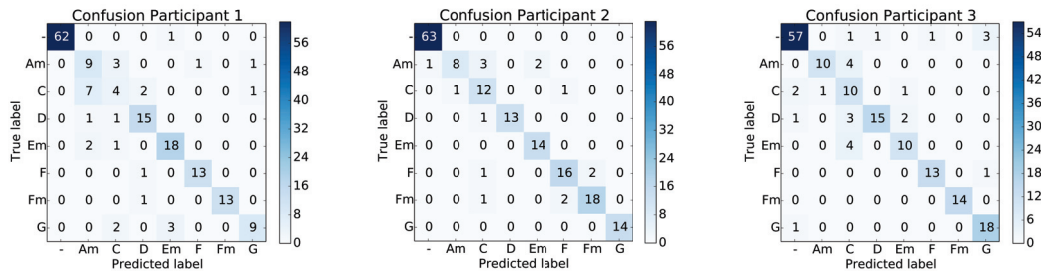


Figure 4.4: Confusion matrices with respect to the classifier’s performance on the test-data. The total number of instances per class varies over the three participants because of the random starting position in the pattern and the train-test split.

Feature group	Averaged processor time
Root mean square	1.32 s
Root mean square ratios	1.38 s
Frequency bins	1.73 s
Phase coherences	7.20 s

Table 4.2: Averaged processor time required to compute the respective feature groups for a five minute recording on a MacBookPro⁵.

a MacBookPro⁵ for a five minute recording. The results are shown in Table 4.2. All tested feature groups can be computed fast enough and utilized for real-time applications.

Window Size

Furthermore, we evaluated the trade-off between classification performance and latency, which depends on the chosen window size. We chose to use the RMS and RMS ratios feature groups, as they reached the highest performance in our test with a windows size of 250ms. The relationship between window size and average F1 score is depicted in Figure 4.5.

Inter-Person Performance

Next, we assess inter-person recognition. We apply a leave-one-subject-out validation on our data. The respective results are listed in Table 4.3.

Number of Electrodes

In this study, we used a total of eight sensing channels to construct features. To evaluate the viability of EMG-sensing devices with a limited number of channels, we composed different subsets with fewer channels and tested the recognition accuracy. We additionally tested six

⁵ i7, 2Ghz, 8GB RAM.

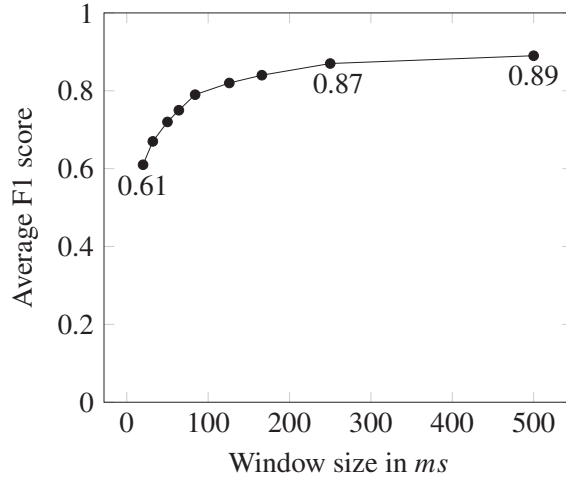


Figure 4.5: Average F1 scores of the classifier given window size. All values are calculated using the RMS and RMS ratios feature groups.

Test set	Accuracy	Precision	Recall	F1 score
Subject 1	0.23	0.14	0.23	0.17
Subject 2	0.18	0.39	0.18	0.12
Subject 3	0.39	0.40	0.39	0.36

Table 4.3: Inter-person classifier performance using the RMS and RMS ratios feature groups. Training was done on the other two subjects' data.

four-channel, ten two-channel and eight one-channel setups. In Figure 4.6, we report the maximum average F1 score of each subset, respectively.

Additionally, we found that evenly spaced electrodes around the forearm in the four-channel setup performed best. This explain the rapid accuracy decrease for the two- and one-channel setup, as they are not capable of encircling the entire forearm.

Summary

Our investigation has shown that recognizing a set of complex hand and finger postures such as guitar chords, using a ring of ten electrodes around the forearm is possible with high accuracy. It needs to be noted that we deliberately chose a simple classification scheme and did not account for accurate electrode positioning. We believe that this gives a lower boundary on recognition accuracy and can be improved upon.

Consequently, we can provide a positive answer to **RQ2a**. We have further identified contributions to **RQ2b** regarding technical requirements. We found that the feature groups of **RMS and their pairwise ratios provide good recognition accuracy** while being feasible for a real-time scenario regarding computation time (**RQ2b**).

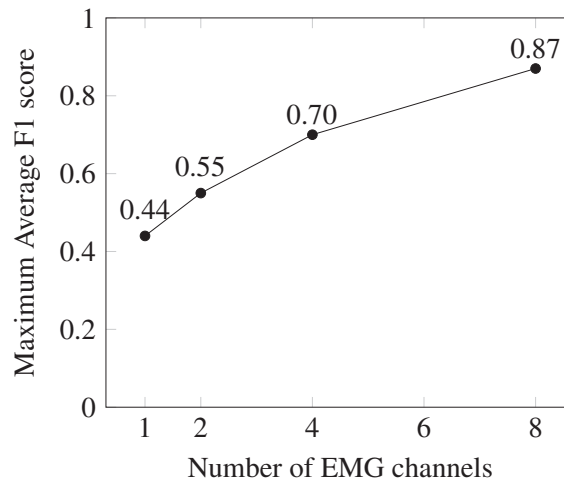


Figure 4.6: Maximum average F1 score for different EMG channel configurations. Features include RMS and their ratios with a 250ms window size.

We found that increasing the window size increases recognition performance up to an F1 score of 0.89 using a 500ms window. However, this induces a trade-off between the response time of the system and detection accuracy. Our results indicate that choosing a **window size of 250ms** is a good trade-off between latency and accuracy (**RQ2b**). Furthermore, we do not recommend using fewer than **eight electrode channels** for detecting fine motor tasks, as recognition accuracy drops for lower numbers. An evenly spaced electrode configuration is to be preferred (**RQ2b**).

When evaluating individual confusion matrices of the participants, we found that the classifier struggled to distinguish between the very similar chords Am and C. Hence, we suggest to **evaluate feature performance for each class separately**. Lastly, we observed that **inter-person classification is difficult** if electrode positioning is not fixed and training data is limited. These task constraints and requirements should be considered for proficiency-aware systems leveraging EMG (**RQ2b**).

4.2.3 Main Study - Evaluating a Guitar Tutoring System

A guitar tutoring system should be able to assess the student's play accuracy in real-time and respond accordingly. If this is technically feasible, how can we deliver adequate feedback that supports the learning process? What are technical and task limitations for such a system? In this study, we want to answer the question whether such a system is able to provide automatic tempo adjustment during the learning phase or if manual adjustment is to be preferred.

To adequately learn a new procedure, e.g., playing a new song, it is essential to first execute the individual steps slowly yet correctly and gradually increase speed when execution

is flawless. We pick up on this paradigm and apply it in a guitar tutoring system. Furthermore, we evaluate the user experience of EMG-based tutoring after we have shown that it is technically feasible.

Scenario

Based on the results from our preliminary study, we developed EMGuitar, a guitar tutoring system that provides tempo adjustment for song playback. EMGuitar decides on a new tempo based on the measured player's proficiency.

Contrary to audio-based systems, EMG-based systems are able to detect *how* someone is playing. While two performances might sound the same, one player might use a different hand posture resulting in effortless play. Especially when learning the guitar, bad hand postures can lead to hand pain diminishing the joy of playing. Hence, the system can advise the player which chords and chord changes need practice.

Our tutoring system detects the player's proficiency by evaluating how accurately a given chord progression was executed. If the player is comfortable with the current song tempo, chord accuracy will be high and the system increases the tempo. If the chosen tempo is too fast, the player will struggle to finger the chords correctly, hence accuracy will diminish. The system reacts by slowing the tempo for the next iteration. During the study, we analyzed how the musicians interact with EMGuitar and which requirements and problems arose when they used the system.

Design

We employed a within-subject design with two conditions; the first with manual tempo adjustments by the participants, the second with the tempo adjusted by the system. We used the same apparatus as in the preliminary study. Preprocessing and feature extraction were identical. As our previous evaluation has shown that the RMS and RMS ratios feature groups performed best, we used these groups with a window size of 250ms.

Since the system is aimed at beginner guitarists, we reduced the set of chords to C, F, Am and G. This set of chords forms the *I-V-vi-IV-progression*, which is heavily used in pop music. By including the barre chord F, we introduced a challenging chord for beginners as well.

To counteract learning effects due to repetitive chord changes, we created two chord sequence patterns (A and B) out of all possible permutations of the above chords. Each chord appeared 12 times in each pattern, while their duration was randomly assigned from one to three beats. A pattern consisted of 20 bars; each containing four beats. For calibration, we tasked the participants to play two bars of each chord twice.

The tempo adjustments were made between 40BPM and 90BPM. Tempi lower than 40BPM were hardly playable, while 90BPM was challenging if chords changed as often as every beat. Based on test runs, we chose the following mapping for the new tempo:

$$T_{new} = \min(90, \max(40, T_{old} + (Acc_{player} - 0.5) \cdot 100)),$$

where T_{old} is the previous tempo and Acc_{player} the estimated chord accuracy of the player.

Since chord changes might happen every beat, we chose to only neglect one epoch ($250ms$) at the end of each chord change. Re-evaluation on our previous data set has shown that accuracy only decreased slightly, average F1 score dropped from 0.87 to 0.84. To further counteract noise due to chord-change artefacts, we set up a majority voting as a post-processing step.

Participants

We recruited eight participants from the University of Stuttgart through mailing lists. The data of seven (1 female, $\bar{x} = 22.6y$, $s = 2.3y$) were used for further analysis⁶. All participants were beginners on the guitar and reported normal sight and hearing. They were able to play the four chords. After the experiment, each participant was paid an allowance of 10 Euros.

Procedure

After introducing the prospective participants to our study, we handed them a detailed study description, stating that they were to play a given chord sequence with the guitar during several iterations. Playback, provided by a metronome, would be adjusted either manually by the participants or automatically by the system. Participants were made aware that adhesive electrodes would be placed on their forearm and their muscle activity would be recorded. After providing informed consent, the participants were asked to complete a demographic questionnaire, polling sex, age, work field, highest educational qualification, their eyesight and general health. The experimenter was present to answer any questions.

Before the actual experiment was started, the experimenter placed electrodes on participants' forearms. They were then given time to familiarize themselves with the provided guitar. Subsequently, we conducted a first calibration at $40BPM$ followed by a first run consisting of five repetitions of either pattern A or B and intermediate tempo adjustments between repetitions. Adjustment were either made manually or automatically. The last repetition was always executed at $90BPM$, ensuring the same end tempo for each participant. After a short break, another calibration was conducted followed by five repetitions using the other experiment condition. A complete overview is depicted in Figure 4.7.

At the end of the experiment, we asked the participants to fill out two identical questionnaires, one targeting the manual condition and one for the automatic condition. The questionnaires contained questions based on the work of Yuksel et al. [338] and their piano tutoring system. These questions were aimed at the learning process and mastery of the piece, as given in Table 4.4. Another questionnaire was specifically tailored to the automatic condition where we asked the participants the questions listed in Table 4.5.

⁶ We believe this to be a meaningful sample size for our formative evaluation since all participant are familiar with the instrument [130].

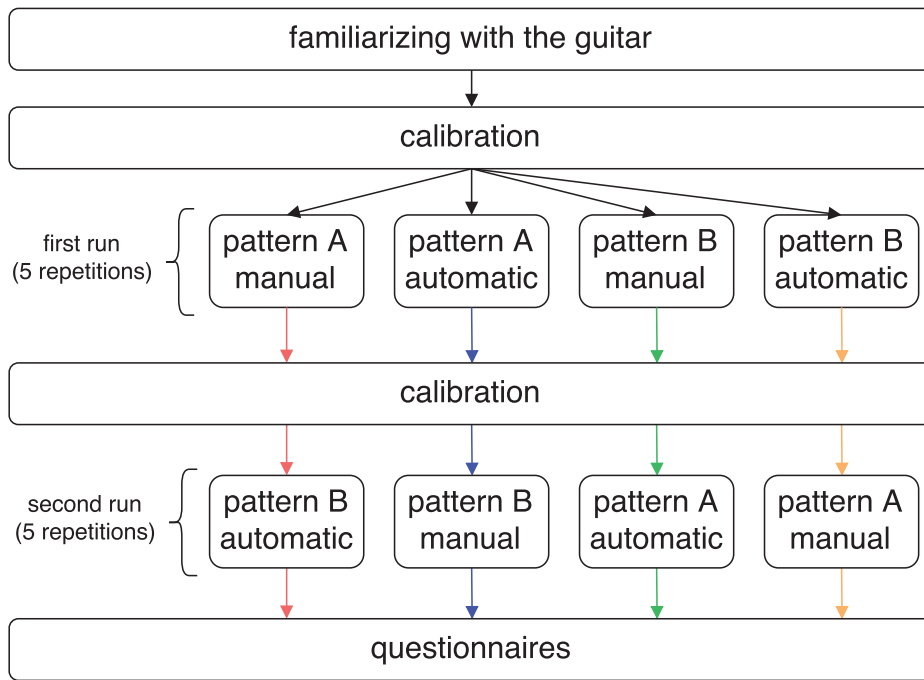


Figure 4.7: Study procedure for the evaluation of EMGuitar.

Learning process and mastery of the piece

- Q1** How well did you feel mastering the piece?
 - Q2** How correct did you feel you played?
 - Q3** How easy was it to learn the piece?
 - Q4** How enjoyable was it to learn the piece?
-

Table 4.4: Questionnaire (7-point Likert) for both conditions (automatic, manual) adapted from Yuksel et al. [338].

The whole study including setup and electrode placement did not exceed 60 minutes. Ethical approval for this study was obtained from the Ethics Committee at the University of Constance.

Results and Discussion

In this section, we will discuss the results from our questionnaires and their implications. We will also revisit challenges that we identified in our preliminary study.

Questionnaires

We compared the answers for the automatic and manual conditions from the first two questionnaires (cf. Table 4.4). A Wilcoxon Signed-ranks test showed that the reported Likert

User experience of EMGuitar

Please rate your perceived accuracy of the system in evaluating your performance.
 How did you perceive the tempo adjustments the system made?
 How challenging were the system's tempo choices?
 Did the electrodes limit you in playing the guitar?
 Would you use the system to learn to play guitar? Why or why not?
 Imagine the system/electrodes could be integrated into your garments. How does this change your perceived usability of the system?
 Further comments

Table 4.5: Questionnaire (5-point Likert and free text) exclusively for the automatic condition.

scores for all four questions were not significantly affected by the condition the subjects were using. The results are illustrated in Figure 4.8.

However, the results of the first two questionnaires (cf. Figure 4.8), indicate that the participants were comfortable with either tempo adjustment. We believe that the automatic system gave the participants a feeling of security as it relieved them from having to choose a new tempo, which was lower on average for the automatic condition⁷. However, answers for "How enjoyable was it to learn the piece?" suggest that the manual condition was more enjoyable for the subjects. We believe, that this can be explained by the low accuracy reported by the system and might have irritated the subjects. Nevertheless, participants stated that the electrodes did not limit them during their guitar play in the third questionnaire.

We analyzed the textual feedback provided by the users in the third questionnaire (cf. Table 4.5) using affinity diagramming with two researchers identifying themes in the qualitative data. We found that users perceived an increase in playing accuracy:

The given tempi helped to challenge me in play more quickly but still accurately.
 (P7)

Further, users were able to realize, reflect upon and rectify possible flaws in their playing style:

Yes, because my own adjustment was probably too high, the system probably is more realistic. (P2)

EMGuitar was also perceived as offering a playful experience and the opportunity to practice alone while still receiving feedback was welcomed by the users:

Yes because it's fun and it has the self-learning aspect into it without having a teacher to keep telling me what to do. (P3)

⁷ Automatic: $\bar{x} = 57$, $s = 18$; Manual: $\bar{x} = 66$, $s = 12$.

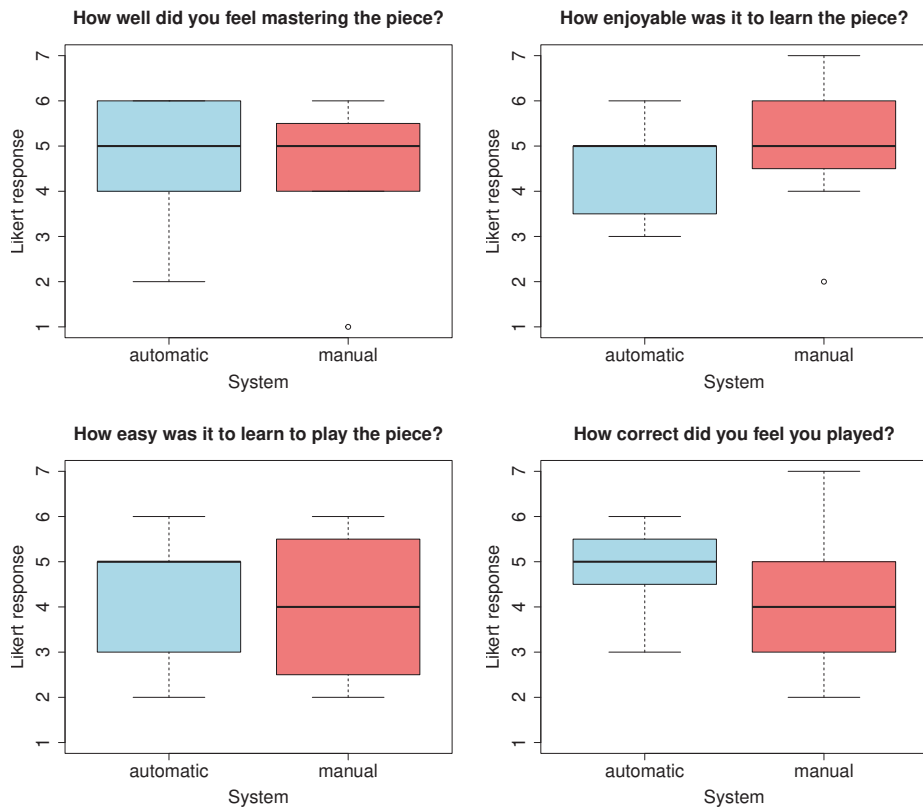


Figure 4.8: Boxplots illustrating the results of the first two questionnaires.

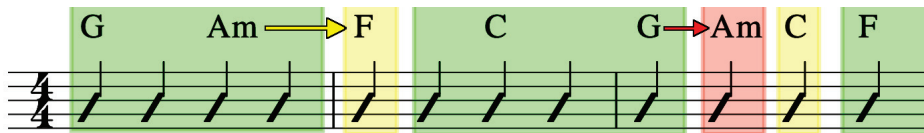


Figure 4.9: Possible feedback visualization for chord grips and chord changes. Colors range from green (flawless play) to red (faulty play).

In the free text comments, the subjects asked for more sophisticated feedback like chord and chord-change highlighting. This matches with our vision of the system and encourages further development. Figure 4.9 shows a possible visual feedback.

Intra-Person Performance

To assess person-dependent performance during the experiment, we utilized the recorded calibration data. We report F1 scores for training and testing on either calibration, respectively, yielding an average F1 score of 0.89. To evaluate the degradation of accuracy due to electrode location shift and sweating, we additionally trained on the first calibration set and

tested on the second, conducted approximately 20 minutes later, only reaching an F1 score of 0.59 ($min = 0.25$, $max = 0.72$).

We have shown that detection accuracy is high when calibrating just before the evaluation. Yet it is also evident that performance can degrade quite fast after half an hour of playing time. Investigating the individual confusion matrices of the participants, we identified the following structural errors:

Electrode location shift and impedance change Over time the electrodes start to shift their location and pick up different signal characteristics. Despite fixating them with tape, even small shifts can lead to a drastic signal change. Furthermore, increased sweat production, e.g. due to nervousness, changes the electrode impedance and the resulting signal as well.

Similar postures Especially between Am and C, we have observed more misclassifications. As both chords are rather similar, this is not surprising. Hence, it can be difficult to distinguish motor tasks that bear close resemblance.

Estimated chord accuracy when playing either pattern A or B ranges between 30% and 70%. While this can be attributed to the player proficiency, quick chord changes at high tempi introduce artefacts in the signal chain. We believe that detecting chord changes can remedy this problem and also provide feedback on flawed changes (as explained in Figure 4.9). For scenarios where this issue is problematic, a possible solution could be to introduce a temporal dependency between consecutive chords, e.g. by using hidden Markov model (HMM) as depicted in Figure 4.10.

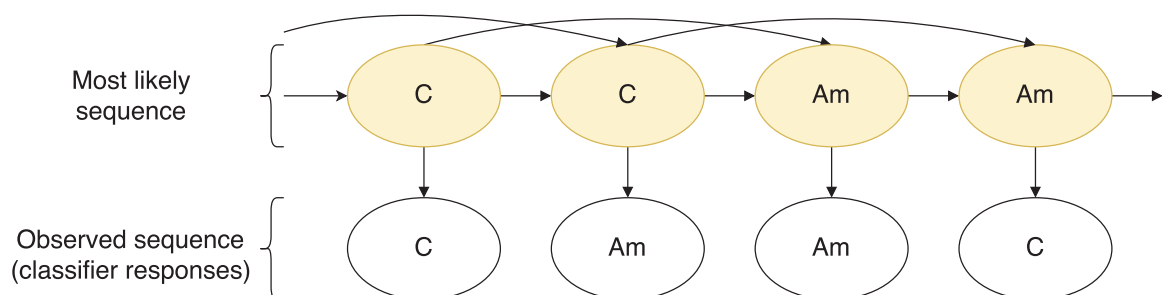


Figure 4.10: Schematic depiction of a second-order hidden Markov model (HMM) for chord estimates. The most-likely sequence is calculated based on the estimated transition and observation probabilities.

Inter-Person Performance

Inter-person performance was evaluated using the same leave-one-subject-out (LOSO) approach. Furthermore, we applied a principal component analysis (PCA) before learning.

The PCA reduces the dimensions of the feature-vector while preserving the maximal possible variance. This technique can help to compensate inaccurate electrode placement across participants and has been applied to EMG signals before [340]. Despite relying on only a maximum of nine principal dimension, the PCA-version performed as well as the original classifier relying on a feature-vector of 36 dimension. While we cannot report any significant improvements regarding classification performance across participants, these findings suggest that dimensionality reduction techniques are effective and suitable for EMG-based detection. We additionally tested the PCA with whitening transform ensuring uncorrelated components. However, this did not increase the performance for any of the participants.

4.2.4 Requirements and Limitations

We have presented EMGuitar as an exemplary tutoring system that recognizes fine motor tasks. Our results show that a high classification accuracy can be achieved using an electrode ring around the forearm. However, performance across participants is not yet satisfactory. To foster further research for EMG-based proficiency-aware systems and provide an answer to our second research question **RQ2b**, we enlist technical constraints and task requirements based on our findings with EMGuitar.

Multiple Sensing Channels Are Essential

We recommend a **minimum number of eight sensing channels** for the detection of fine motor tasks. Best results can be achieved when the electrodes are spaced evenly along the observed extremities. More channels can be beneficial to estimate electrode shift and achieve better cross-session accuracy as demonstrated by Amma et al. [7].

User-Dependent Calibration Is Still Required

We advise for **user-dependent calibration** during the training phase of the system. Our results showed that cross-session accuracy quickly deteriorated under environmental influences. Hence, it is vital to perform prior calibration of the system requiring the need for a human supervisor. A multi-modal approach can remedy this limitation. Other modalities, such as sound for our guitar scenario or camera vision, can help to calibrate for EMG automatically. Iterative approaches to calibration have already been researched [140], concluding that complete re-calibration is preferable. Independent component analysis (ICA), as shown by Naik et al. [212], might be a method to tackle this challenge, revealing the contribution of specific muscle fibers and applying automatic re-calibration, but is subject to preconditions, such as sufficient number of EMG channels.

Accurate User Feedback Needs To Model Transitions

Especially for fine motor tasks comprising temporal sequences, we advise to **model the transitions** as well. We found, that users lacked the appropriate feedback when changing

between chords, such as depicted in Figure 4.9. Similarly, modeling transitions between each motor task makes detection more robust, but tends to be problematic for EMG-based systems, as isotonic⁸ muscle activation occurs. Here, we motivated possible solutions using hidden Markov models as shown in Figure 4.10.

Beneficial Task Characteristics For EMG

The results of our studies as well as related work [262] suggest, that recognizing isometric muscle activity⁹ works best for EMG-based sensing. This means that **motor tasks involving static hand postures** combined with constant pressure, for example, pressing down guitar chords, are suitable as we have shown in our work. Related work suggests that isotonic muscle activity⁸ is more difficult to recognize using EMG [303].

A multi-modal approach to recognize transitions between repetitive isometric states can remedy this limitation, thus allowing recognition of various movements during prolonged activities such as repetitive motions in sports or following a procedure in assembly work. Even playful interaction with controllers for games or musical instruments are conceivable like YouHero [54] utilizing the GuitarHero controller or the Air-Guitar Hero scenario already illustrated by Saponas et al. [262].

Consequently, this makes EMG-based recognition inherently unsuitable to detect creative movements, e.g. during improvised dancing, as appropriate training data cannot be generated. However, a more general model could be researched in the future detecting only distinct signal characteristics of certain movements and rely on the user to provide interpretation as we showcase in Section 5.1.

4.2.5 Conclusion

In this use case, we have showcased the feasibility of detecting domain-specific, fine-grained hand and finger postures using Electromyography in a guitar playing scenario. Contrary to recent approaches to forearm EMG, we did not aim to recognize a pre-defined set of gestures that need to be executed by the user. Instead, we explore EMG as an implicit sensing modality for proficiency-aware systems. This enabled us to recognize domain-specific movements and postures such as guitar chords to support users in learning, and allows for natural interaction with the instrument [316].

In two studies, we evaluated the accuracy of an electrode ring around the forearm to detect different guitar chords and showed that the commonly used root-mean-square (RMS) feature of signal segments and its ratio between sensing channels is still reliable in our scenario. In EMGuitar, we applied these results and implemented a tutoring system that automatically adjusts playback speed of a song to match the player's proficiency. Our system was received

⁸ Elongating and shortening of muscles including visible movement.

⁹ Continuous muscle activity without muscle movement, e.g. when pressing against a wall.

well; neither diminishing ease of use nor being disruptive for the participants. Yet, as we intentionally did not account for highly accurate sensor placements, only within-session classification achieved high accuracy.

Based on our results, we propose technical requirements for EMG-based fine motor task detection including preferable electrode count and the necessity for supervised calibration. We believe that our approach is suitable for a number of physical tasks that follow certain characteristics as outlined in the requirement section.

We envision the use of electrode bands as a natural sensing modality for fine motor task training in proficiency-aware systems, especially since sensing electrodes can be easily integrated in garments and wearables. Non-contact electrodes relying on electric field sensing [311] are even less obtrusive. Such systems will be able to ascertain a user’s proficiency in the main domain task and provide assistance accordingly. Consequently, it is vital to investigate the accuracy of low-cost and mobile devices for EMG. The following use case covers this scenario.

4.3 Use Case II: FitBack - Recognizing Exercise Form

Physical activity is an increasingly important part of our lives. The number of people exercising regularly is on the rise [73, 8], as is the number of interactive systems that accompany exercise every day. However, while regular physical exercise offers mental and physical health benefits, these come at the price of the hard work needed to master the forms and techniques required for a given sport.

As humans, we rely on proprioception for orientation in physical spaces [238]. This sense of relative body orientation allows us to effectively position our body and extremities. When learning new body poses, we try to mimic movements to get a sensation of “how the movements feels” which allows us to create a *motor memory* [171]. The efficiency of this technique highly depends on the task and the individual. Even today, the exact mechanism of consolidation of motor memory and possible distractors is not completely understood [171].

While EMG carries much information about our body physiology, understanding the signal is challenging. To that end, this use case investigates user performance when interacting with muscle activity data. In this lab study, we examine the performance of amateur users when supported via EMG-based biofeedback through our mobile prototype FitBack while performing bicep curls.

4.3.1 Methodology

This use case will focus on the technical aspects of FitBack, such as feasibility of low-cost mobile EMG prototypes in the context of **RQ2**. The work also includes an investigation of design qualities of EMG-based biofeedback that will be discussed in Section 5.1.

We chose an exploratory approach by first building a mobile prototype capable of recording electromyograms via adhesive electrodes. We used this system to explore necessary requirements and resulting technical constraints in a user study with 18 novice to experienced participants.

With regard to **RQ2**, we define a related research question for this use case:

RQ2

How accurately can we estimate proficiency from a user's muscle activity?

RQ2c: Can we effectively use EMG-based biofeedback to support users during physical activity with a low-cost device?

This research question investigates user requirements in the context of technical constraints that arise from mobile electromyograms. Existing research [263] has already demonstrated the feasibility of EMG for recognizing body movements. In this work, we provide a technical evaluation of the signal accuracy and distinctiveness with regard to recognizing the correct form of fitness exercises using a mobile, low-cost EMG device. We believe that a low entry barrier is essential to advance EMG-supported physical activity. This research question is explored in the experimental investigation, where we assess the technical feasibility of FitBack by comparing it with expert and coach reviews. We also assess user experience, perceived workload, and flow experience during the exercises when interacting with FitBack.

This use case focuses exclusively on the technical aspect with regard to **RQ2c** as outlined above, hence only depicting necessary study details. A complete introduction into the study design and rationale is given in Section 5.1.

4.3.2 FitBack

FitBack is an integrated system that records and processes EMG data. The system was developed with requirements for physical exercises in mind, such as mobility, allowing users to freely move around. Different feedback modalities and granularities allow FitBack to cater to a broad audience of users, while a simple but robust detection algorithm is used to recognize exercise repetitions. The following section provides an overview of the implementation of FitBack. Note that FitBack uses the hardware base of EMBody (Section 6.2), but employs a different detection algorithm.

Hardware

FitBack's hardware is based on an ESP32 microcontroller¹⁰, which is a low-cost and low-power system on a chip with integrated WiFi. We measure EMG using a bipolar measurement technique [206] including a reference electrode and two sensing electrodes to minimize

¹⁰<https://www.espressif.com/en/products/socs/esp32>

noise artifacts. Amplification is realized through an existing design¹¹ that we adapted for our purposes. The amplified signal for each channel gets processed by the ESP's Analog Digital Converter (ADC) yielding a 12 bit resolution. Measurements are broadcasted over the network via UDP at a sampling rate of 200 Hz , which is sufficient for the following filtering steps. The whole hardware system (see Figure 4.11) can be powered by a powerbank and fits in a 3D-printed case allowing it to be carried around by the user.

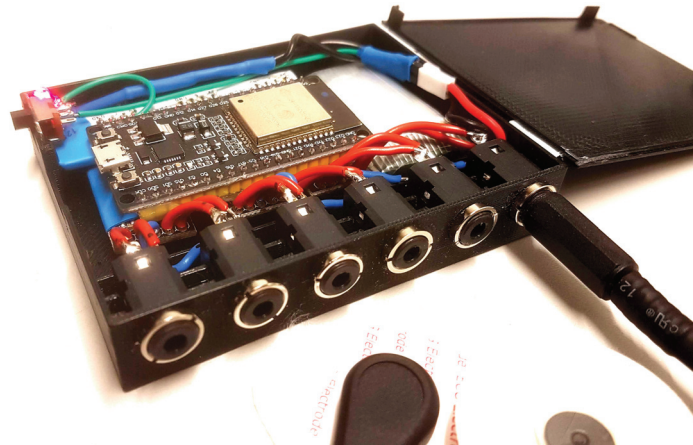


Figure 4.11: FitBack: Hardware prototype to record EMG data.

Software

To process incoming EMG data, FitBack includes an accompanying software application that receives the data samples to generate biofeedback. First, a bandpass filter between 2 Hz and 100 Hz is applied [261], reducing long-term drifts and high-frequency noise. A follow-up bandstop filter between 49 Hz and 51 Hz removes interference from power line noise. Secondly, FitBack calculates epoched root mean square (RMS) features with a window size of 40 samples¹². RMS values can be seen as a proxy for the amplitude of the EMG signal, hence increasing when muscular activity increases [206]. A Savitzky-Golay filter [269] was used to counteract fluctuations that might be confusing to users. After these processing steps, the signal is visualized.

To detect individual exercises, FitBack further cross-correlates the RMS signal with a target signal which is acquired for each participant during the calibration phase before the experiment. A large correlation coefficient indicates an alignment between the target signal and the incoming signal during the experiment. The correlation value is used to determine whether individual trials were correctly performed by aggregating the values for all measured muscle groups.

¹¹ www.github.com/BigCorvus/2-Channel-Biopotential-Amp

¹² Corresponds to 200 ms ; a hop size of 0.5 times the window size was used. Parameters are based on preliminary tests.

4.3.3 Lab Study

In this study, participants were asked to perform biceps curls, a simple, yet challenging exercise to perform accurately. It involves few muscles making it easy to place electrodes. In a mixed design, users were provided with a total of three different visualizations (see Figure 4.12) as within-subject factor. Additionally, a between-subject factor was introduced in the form of auditory cues. Please refer to Section 5.1 for a detailed rationale of the study design and conditions.

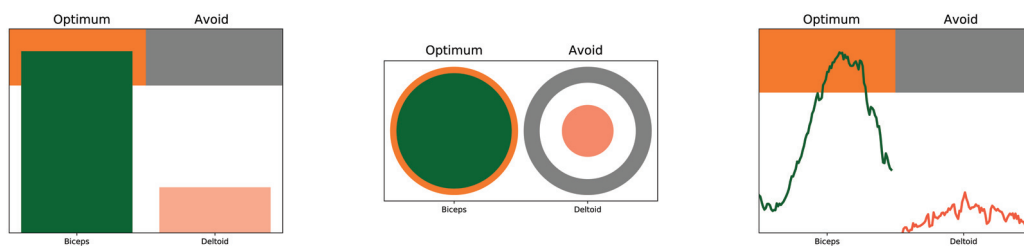


Figure 4.12: FitBack visualizations showing the same EMG data. Abstract representations *Bars* (left) and *Circles* (middle), and the detailed representation *Lines* (right).

Measures

For our data collection, we focused on the following aspects in relation with our research question: (1) system accuracy for performance assessment of the executed bicep curls, (2) usability and perceived workload when exercising with FitBack.

System accuracy

During the experiment, we collected performance assessments by the experiment instructor (a sports professional) who judged each execution as correct or incorrect. Additionally, we logged FitBacks' assessments of the trial. Here, preliminary tests have showed that a correlation value of greater than 0.5 for both the bicep and the deltoid signal indicated correct execution of the exercise. To gather objective ground truth, we conducted three expert reviews of the collected EMG data. The experts are researchers that use EMG-based systems on a daily basis and are familiar with the fitness exercise. They rated each execution of every participant in the study on a 7-item Likert scale. Correct execution was rated highest, while points were deducted for bad form, such as lifting with momentum or using excessive deltoid activation.

Usability and Perceived Workload

We measured the Usability Metric for User Experience (UMUX) [83] to identify FitBack's user experience and detect potential flaws when exercising with it. To assess perceived workload during the exercise, we employed the NASA Task Load Index (NASA-TLX) [111]

in its raw form without the weighting process [110]. We further used the flow experience questionnaire [280] to measure the participants' engagement in the training exercises.

Participants

We recruited 18 participants (four female, 14 male) with a mean age of $\bar{x} = 25.6y$ ($s = 3.24y$) from university mailing lists. The average self-reported general fitness experience¹³ was $\bar{x} = 59.3$ ($s = 22.77$) and $\bar{x} = 42.3$ ($s = 29.81$) for experience in weight training¹³. Participants trained $\bar{x} = 3.2h$ per week on average ($s = 1.54h$). Each participant was compensated with 10 Euros.

Procedure

The experimenter welcomed each participant and explained the intention of the study. After providing informed consent, participants stated their fitness background and demographic data. Next, they received an introduction on how to perform a correct bicep curl and were made aware of common mistakes. Electrodes were placed on the participant's dominant arm on the bicep and deltoid muscles (see Figure 4.13). The experimenter then picked weights that were appropriate to the participant's fitness level. Afterward, the calibration started in which a visualization dictated the rhythm for the bicep curl repetitions. The experimenter verified the correctness of the bicep curls during the calibration phase, both visually and via the EMG data. After successfully calibrating a target signal (cf. Section 4.3.2) for the participant, the experiment started. Participants were given some time to familiarize themselves with each visual condition before starting ten bicep curl repetitions for each condition, yielding a total of 30 repetitions per participant. Note that sound cues were used as a between-subject variable, hence each participant either saw all three visualizations with sound cues or completely without sound cues. Participants filled out the questionnaires for each condition during short breaks. Additionally, the experimenter (acting as coach) silently noted incorrect and correct bicep curls for each repetition. The experiment was concluded by removing the electrodes and conducting the short interview (cf. Section 5.1). In total the experiment lasted approximately one hour. Ethical approval for this study was obtained from the Ethics Committee at the University of Constance.

Results

We report on the statistical analysis of measured usability measures. For each measure, we conducted two-way (*visualization* and *sound*) mixed ANOVAs. If the data deviated from normality, we first aligned rank transformed [328] it. Further, we present an analysis of FitBack's accuracy in detecting correct exercise form.

¹³Measured on a visual analog scale (VAS): 0 to 100.

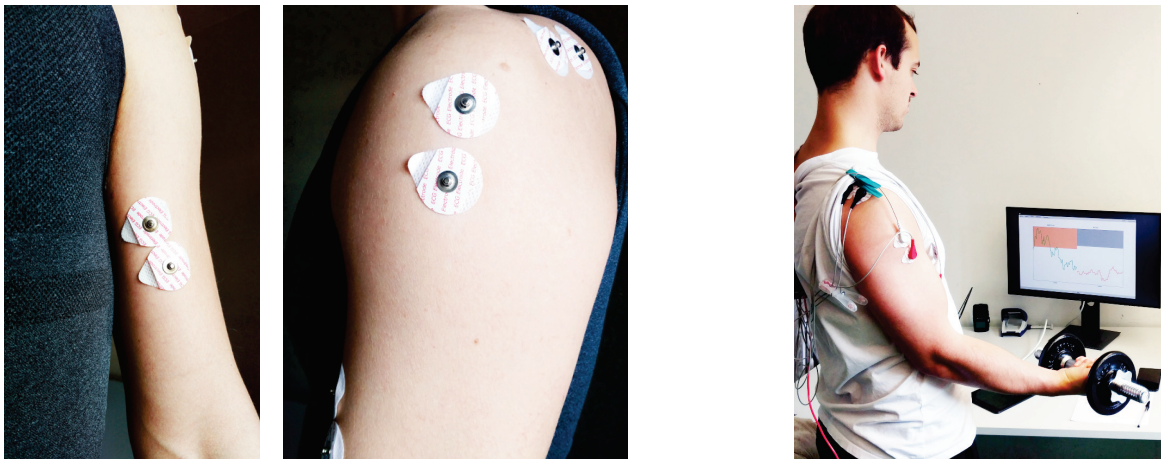


Figure 4.13: Electrode placement on biceps (left) and deltoid (middle) muscles. Reference electrodes can be seen close to the scapula. On the right side: participant during the study.

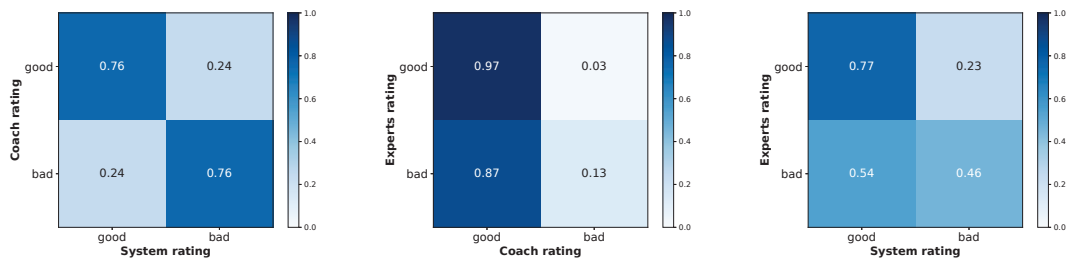


Figure 4.14: Confusion matrices (associated truth labels on the y axis) between coach—system (left), experts—coach (middle), experts—system (right). Note that percentages are based on within-class instances (per row), as class distribution is highly skewed.

System Accuracy

We first investigated the inter-rater agreement of the three experts using the $r_{WG(J)}$ agreement index [135]. A value of $r_{WG(J)} > .99$ confirmed high agreement. We subsequently transformed the averaged expert rating into a binary scale by categorizing all executions rated 4 and higher as correct. Similarly, we transformed the system’s correlation scores into the same binary scale, by aggregating consecutive scores¹⁴ for one execution using the respective maxima, ensuring robust detection. If this maxima was greater than 0.5 the execution was registered as correct. Ultimately, this allowed us to compare all three assessments from the experts, the coach (the experiment instructor) and the system. An overview is provided in Figure 4.14.

On the left side (Figure 4.14), the confusion matrix between the coach’s assessment and the system’s assessment shows a good degree of consensus. It is noteworthy that the class distribution is unbalanced: the GOOD class contains 519 samples, while the BAD class only

¹⁴A correlation score is computed every 72ms.

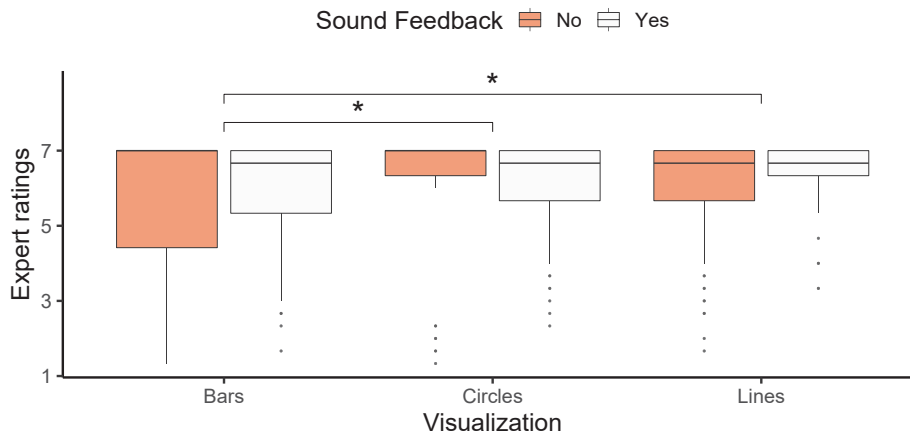


Figure 4.15: Expert ratings given the shown visualization and sound feedback. Significant differences are marked with *.

contains 21¹⁵. However, both the trainer and the system failed to accurately detect bad executions as showcased in the middle and on the right side (Figure 4.14). Using the experts' ratings as truth labels (469 GOOD samples, 71 BAD samples), we compared the coach's and the system's assessment predictions. While the system detected approximately half of the bad executions, the coach only recognized 13% of them. Thus, having access to the actual EMG data (as did the experts), improves judgment of exercise form. While the current algorithm of FitBack is able to detect exercise form to a degree similar to that of a coach, more sophisticated methods, e.g. modeling the EMG curve instead of approximating it with one scalar, are required for accurate automatic assessment.

Exercise Performance

We further tested the performance of the participant with regard to correct exercise form as judged by the experts. Expert ratings given *visualization* and *sound* can be seen in Figure 4.15. We found a significant effect of *visualization* ($F(2,518) = 3.93, p < .05$) and a two-way interaction effect of *visualization* and *sound* ($F(2,518) = 3.80, p < .05$). Post-hoc pairwise comparison using tukey-adjusted p-values showed a significant difference between the visualization *Bars* and *Lines* as well as between *Bars* and *Circles*. We found one significant two-way interaction effect between *Lines* and *Circles* for the factor *sound* (*No - Yes*).

Usability and Perceived Workload

We did not observe any significant differences for either the NASA-TLX, the UMUX nor the flow experience questionnaire. Descriptive statistics for each condition are provided in Table 4.6.

¹⁵In other words: the coach has rated 519 executions as good and 21 as bad.

Visualization	Sound	NASA-TLX		UMUX		Flow experience	
		\bar{x}	s	\bar{x}	s	\bar{x}	s
<i>Bars</i>	No	38.6	17.8	80.1	18.1	44.6	6.06
<i>Bars</i>	Yes	47.3	14.8	81.5	17.3	42.7	4.12
<i>Circles</i>	No	41.3	12.2	80.6	18.0	43.7	4.92
<i>Circles</i>	Yes	47.9	12.0	82.9	12.7	43.1	7.11
<i>Lines</i>	No	43.8	14.4	78.2	21.4	43.4	6.52
<i>Lines</i>	Yes	45.6	14.9	81.9	11.8	42.2	3.19

Table 4.6: NASA-TLX, UMUX and flow experience score per condition.

4.3.4 Conclusion

While usability and perceived workload metrics showed no significant differences for feedback type, our experimental investigation with FitBack demonstrated that low-cost EMG sensing devices are suitable to monitor exercise form. Our quantitative analysis in the study showed that FitBack could perform on a par with a supervising coach in terms of recognizing exercise form (**RQ2c**), though both automatic prediction and the coach perform less accurately than reviews by experts who have access to the EMG data, highlighting its distinctiveness.

As of now, FitBack’s post-hoc analysis is based on correlation coefficients of the respective muscle groups. We believe a more elaborated classification approach, e.g., using multiple features to describe relevant characteristics of the EMG curve, allows one to identify execution more accurately. However, this results in a **trade-off between the amount of customization needed for a specific exercise** (e.g., training the model to detect common errors of biceps curls) **and the capability to generalize feedback for a wide array of fitness exercises** (**RQ2c**).

In this use case, we intentionally refrained from building a sophisticated data processing model to keep the amount of required domain knowledge that informs the model training process to a bare minimum. Building a fully automatic system to supervise specific fitness exercises is outside of the scope of this work. We already know from related work that it is feasible to build accurate detection systems using EMG [261, 7]. In contrast, FitBack is designed to be an aid for amateurs and coaches alike by **presenting data about muscle activity in a comprehensible manner and letting the user reflect on their data**, thus contributing to their own body awareness, as we will highlight in Section 5.1.

4.4 Chapter Summary

In this chapter, we have looked at two use cases that leverage EMG as a modality to estimate a user’s proficiency. Our findings have contributed to **RQ2**, elaborating on technical require-

ments and limitations as well as suitable task domains for EMG-based proficiency-aware systems. We showed that EMG could detect fine-grained motor tasks but requires person-dependent calibration and suffers from high setup costs, such as multi-channel recording systems. While low-cost devices are available, their signal quality only allows recognizing coarse motor tasks.

Most importantly, EMG currently suffers from a lack of interpretability. The signal is complex, and users are unfamiliar with its implications. Consequently, interactive systems suffer from limited user trust in their adaptations (cf. EMGuitar). It is thus essential to introduce users to this — for them — novel technique. Users should become aware of their own muscle activation. Doing so allows for user-driven interpretation of collected EMG data, thus tasking the user with making the connection of their movements to the presented data. We envision that this approach can facilitate a more profound learning process than an "out-of-body" interaction, where a system assesses the user's proficiency.

This leads us to the question of how to design for this kind of proficiency awareness. Especially for complex physiological sensing, there exists a multitude of plausible visualizations. Consequently, the next part of this thesis will introduce and evaluate concepts to facilitate proficiency awareness in interactive systems.

III

FACILITATING PROFICIENCY AWARENESS

Chapter 5

Designing Proficiency Awareness

“What is to be sought in designs for the display of information is the clear portrayal of complexity. Not the complication of the simple; rather . . . the revelation of the complex.”

Edward R. Tufte

Making users aware of their proficiency can potentially be a challenging task. Especially unskilled users tend to overestimate their own expertise [172]. Making matters worse, this user group is generally not aware of their own lack of proficiency. Allowing the system to challenge this belief can potentially lead to abandonment by the user. Consequently, there are open challenges in how to design proficiency awareness in a way that users feel motivated for personal growth while, on the other side, not feel dictated by the system.

The investigation in this chapter focuses on designing proficiency awareness for EMG-based systems as well as in everyday computer-based work. EMG allows for unique bodily insights for users, while our second use case will demonstrate that proficiency awareness can also be realized with simple linguistic metrics in a text production task. We conclude this investigation with an outlook on artificial proficiency and how it can provide added motivation for users. Consequently, the research in this chapter is guided by **RQ3**:

RQ3

What are design criteria for proficiency awareness in interactive systems?

Facilitating Bodily Insights

We first continue our investigation of FitBack, elaborating on the impact of the different visualizations and additionally evaluate the system with sports coaches. Here, we focus our attention on the dimension of feedback types and how to communicate proficiency in a complex task scenario that allows full user control (cf. Section 2.4).

Reflective Feedback in Text Production

Having observed that often simple metrics are sufficient to estimate proficiency, we introduce this use case, where we evaluate opportunities for proficiency feedback in everyday computer work in a simple text production task. Making use of a text readability metric, we investigate the impact of reflective feedback on the users' performance and task approach. While the users are free to act, this system nudges users by judging their work, taking away control (cf. Section 2.4).

A Take on Artificial Proficiency

Finally, we provide a short outlook on EMPiano, an EMG-based system that allows users to utilize musical effects on the piano keyboard regardless of their play proficiency. We highlight how a sneak peek of increasing one's proficiency can be a vital motivator in this transition phase. This use case is unique in a way that it does not communicate the current proficiency of users (cf. Section 2.4), but rather allows users to experience the potential of a higher proficiency by themselves.

This chapter is based on parts of the following publications.

Jakob Karolus, Annika Kilian, Thomas Kosch, Albrecht Schmidt and Paweł W. Woźniak. 2020. Hit the Thumb Jack! Using Electromyography to Augment the Piano Keyboard. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference (DIS '20)*. ACM, New York, NY, USA.

<https://doi.org/10.1145/3357236.3395500>

Annika Kilian, Jakob Karolus, Thomas Kosch, Albrecht Schmidt and Paweł W. Woźniak. 2021. EMPiano: Electromyographic Pitch Control on the Piano Keyboard. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '21 Extended Abstracts)*. ACM, New York, NY, USA.

<https://doi.org/10.1145/3411763.3451556>

EMPiano is openly available on github: <https://github.com/HCUM/empiano>

Jakob Karolus, Felix Bachmann, Thomas Kosch, Albrecht Schmidt and Paweł W. Woźniak. 2021. Facilitating Bodily Insights Using Electromyography-Based Biofeedback during Physical Activity. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction (MobileHCI '21)*. ACM, New York, NY, USA.

<https://doi.org/10.1145/3447526.3472027>

Submitted for review.

Jakob Karolus, Sebastian S. Feger, Albrecht Schmidt and Paweł W. Woźniak. 2021. Designing Proficiency Awareness: Leveraging Reflective Feedback in Text Production.

5.1 Use Case I: Facilitating Bodily Insights Using Electromyography-Based Biofeedback during Physical Activity

Learning how to perform new exercises or mastering different forms of movement tests our abilities in controlling our bodies. Some poses are challenging and the difficulties in reaching the required body awareness to achieve the desired proficiency may lead to dissatisfaction and even abandonment, reducing or losing the accrued health benefits [275] altogether. While coaching experts are widely available, physically active users cannot be constantly accompanied while exercising. Past research in Human-Computer Interaction (HCI) showed that interactive technologies can effectively build body awareness for certain exercise scenarios [314, 160]. Yet, it remains a challenge to develop methods that would enable generic, exercise-agnostic methods of better understanding one's body when exercising. Thus, there is a need for developing ways in which users can monitor their bodies to better understand their sports practice.

Technology has the potential to bridge the gap between how we would like to move and how we can control our bodies. While professional sportspeople use a wide array of sensors to improve their performance [20], these technologies are too complex for everyday users. In this use case, we continue our evaluation of FitBack (Section 4.3) and explore the means for users to become more aware of their muscle activity through electromyography. EMG offers additional insight into muscle physiology to facilitate motor memory consolidation. Users desire a better insight into their bodies [55]. Yet, it remains a challenge to understand how complex physiological data streams, such as EMG, can be effectively used to foster insight.

5.1.1 Related Research

Complementing related work introduced in Chapter 4, this section highlights relevant works in the domain of biofeedback.

Biofeedback

Biofeedback is a type of physiological-driven feedback, where physiological responses of a user are voluntarily measured and communicated with the user at the same time to create a feedback loop [44]. Biofeedback has recently been extensively used in rehabilitation and treatment of disorders such as stroke rehabilitation [52], anxiety [210], or substance abuse [281]. A typical biofeedback session exposes the user to a visualization of a physiological signal that is correlated to a disorder. Users are instructed to proactively act towards a desired signal, where the physiological signal is usually visualized in a simplified form. In contrast to raw signals such as brain activity [165, 57] or heart rate variability [336], simpler visualizations are easier to understand by laymen. Results from previous work showed that

biofeedback has a strong influence over physiological functions and offers users several advantages over traditional treatments, including a visible metric for reflection, progress, and ongoing involvement of the treatment process [231]. For example, biofeedback has been successfully employed for stroke rehabilitation [52], posture corrections [226], treatment of neurological impairments [298], and substance abuse [205]. Apart from medical treatments, biofeedback has shown potential to improve training efficiency, as highlighted in the following section.

In the form of biofeedback, EMG has proven to be effective for behavioral change and rehabilitation [206, 4, 121] as well as strength acquisition [90, 53]. Work by Toader et al. [307] showed that users adapted their exercise form after receiving visual EMG-based biofeedback compared to a control group with no feedback. We extend their work and investigate the influence of different feedback modalities on performance and the users' perceived awareness of their body posture and movements.

Fitness and Feedback

The HCI field has recognized possibilities for interactive artifacts to build increased body awareness and thus contribute to a better experience and skill development in sports. Being more aware of the relative position of one's body has been shown to benefit those practicing yoga [313] or swimming [160]. However, gym exercises, perhaps due to their static nature, received more extensive attention. A number of projects proposed using different sensing modalities for recognizing exercise execution such as inertial measurements [317], pressure sensors [299, 71] or depth cameras [40]. Also, the now discontinued Myo¹ armband could effectively differentiate between various gym exercises [169]. While these systems offered effective ways to count execution or assess the correctness of the motion, they did not enable the user to determine if the correct muscles were used in their exercise. Our work examines the possibilities of direct muscle monitoring and investigates the opportunities of the user possessing that knowledge.

Another strain of HCI work recognized the benefits of helping users understand their own bodies through additional feedback. Turmo Vidal et al. [314] built wearable lights that could be customized to enhance the perception of relative body position, improve exercise technique and facilitate instruction [312]. They showed that additional insight during activities can lead to skill development. Earlier, Hämäläinen [105] suggested using the mirror metaphor, which is particularly suited to gym exercises as fitness clubs usually feature mirrors. Anderson et al. [9] developed an Augmented Reality (AR) mirror system which tutored users in gym exercises. Past work shows many opportunities for possible interventions at the gym and a need for users to understand their bodies better in order to train better. However, it remains an open question what sensing and feedback offers the most benefits to users and how it can be adopted into exercise environments. Our work is different from past research as it specifically investigates what happens when users are aware of their muscle activity and how they can benefit from this knowledge.

¹ <https://support.getmyo.com/hc/en-us>

On the consumer market, readily available products which use EMG-based biofeedback for physical activity, such as Athos², Myontec³, and Mpower⁴ are primarily aimed at sports professionals and competitive athletes. Athos relies on post-hoc reports and Myontec reports on exertion over the course of the whole training session. In FitBack (Section 4.3.2), we rely on displaying live muscle activity allowing users to grasp timings of exercise forms, similar to the activation curves provided in Mpower. All commercial systems rely on pre-configured smart clothing and tailored algorithms, limiting the set of exercises that can be monitored. Lastly, these devices provide very specific feedback, designed for professional athletes and post-exercise analysis with professional coaches. While the technology is available, albeit, for a high price, it remains to be investigated how everyday users can benefit from detailed biofeedback. With FitBack, we use a modular approach where users are able to customize their exercise monitoring to their needs. While this means that it is more difficult to provide specific exercise feedback, our work focuses on user-driven interpretation of the data with the aim to provide deeper bodily insight. Consequently, this work examines if and how EMG-based biofeedback can foster reflection on exercise form to further the users' understanding of their body physiology.

5.1.2 Methodology

EMG in current commercial products offers highly customized feedback about muscle strain, exertion, and training effect. These parameters are primarily relevant to professional athletes. In contrast, this use case investigates if and how EMG data can provide insight into one's own body physiology and how to provide an encouraging—yet challenging—way for users to access knowledge about their muscles. As muscle activity differs significantly among users, no algorithm can offer an effective automated interpretation of the signals that considers the individual context of the exercise. Ultimately, it is the user who interprets signals from their body. Our work explores how interactive technology can facilitate that process of interpretation.

While EMG carries much information about our body physiology, understanding the signal is challenging. In this use case, we report on a mixed-method inquiry concerning first impressions and user attitudes of EMG-based biofeedback, complementing the work mentioned in Section 4.3, where we focused on technical aspects of FitBack.

We take another look at the conducted lab study, focusing on user impressions through questionnaires and interviews. This evaluation is intended to provide a comparison to existing EMG-based feedback systems, providing an abstract visualization of muscle activity. Additionally, users were provided with live data of their muscle activity to see if and how real-time visualization led to insights. To further investigate how to design for rich physiological

² <https://www.liveathos.com/>

³ <https://www.myontec.com/>

⁴ <http://www.mpower-bestrong.com/index.html>

feedback in the context of coaching and supervision, we conducted a set of interviews with experienced professional sportspeople and coaches, during which they experimented with the detailed view of FitBack and considered its possible use in their practice.

Our investigation is informed by the previous research, including technical and design requirements. Commercial systems and research probes established that EMG is an effective method to support fitness exercises. In this work, we evaluate if the access to detailed muscle activity can facilitate an improved understanding of one's own body physiology. With increased insight into how their body reacts to exercise, users will experience more health benefits [198] and develop expertise [175]. To explore this, we replicated existing feedback methods as a baseline but opted for a low-cost mobile prototype, which makes EMG-supported sports exercises accessible for amateurs. We ask how feedback should be conveyed, what is the optimal temporality, and what user audience can benefit from such a system. For the latter, we identified two main audiences, namely novice to experienced self-paced users that engage with the feedback directly, and fitness coaches to gain deeper insights into their student's exercise form.

To this end, we employed a mixed-method inquiry consisting of a user study with 18 novice to experienced participants and interviews with sports professionals and coaches. This allowed us to capture opportunities, requirements, and constraints for using interactive muscle sensing and biofeedback during physical activity.

With regard to **RQ3**, we define a related research question for this use case:

RQ3

What are design criteria for proficiency awareness in interactive systems?

RQ3a: How can we design EMG-based biofeedback to facilitate bodily insights during physical activity?

Here, we look at different feedback modalities (visual and auditory) and granularities for a broader audience (amateurs and coaches). The system should provide easy access to bodily insights both for novices and more experienced practitioners while providing necessary details for experts and coaches to find the perfect form. We compare existing designs with detailed EMG feedback through rankings and evaluate opportunities in our first user study. We additionally conducted semi-structured interviews to gain further insights into how participants in this first study perceived their body awareness. In our second study, we conducted interviews with sports coaches to finalize a suitable design. We focus our analysis on the collected qualitative data and derived themes from the interviews for this research question.

5.1.3 Study I - Experimental Investigation of Biofeedback Modalities

In this first evaluation, we focus on one particular fitness exercise and test different feedback types and granularities using FitBack. Please refer to Section 4.3.2 for a detailed description of the prototype. This section reports on the employed study design, used measures, procedure and participants.

Design

We chose the bicep curl as a reference fitness exercise. This simple exercise is suitable for beginners as well as popular among experienced practitioners [26]. Additionally, it involves only a few selected muscle groups (biceps and deltoid) allowing for easy electrode placement [56]. Despite its simplicity, the bicep curl still leaves a certain margin for error, such as performing curls too fast or using the deltoid to aid in lifting the weight. This significantly reduces the training effect and can lead to injuries [56, 26]. Figure 5.1 shows two examples from filtered EMG signals including the bicep as well as the deltoid. The left side illustrates a correct execution involving exclusively bicep activation, while the right side shows a bad example using the deltoid during the repetition. Focusing on a simple exercise with a clear sequence allowed us to evaluate the suitability of EMG-based biofeedback with a special focus on different needs from a diverse audience (novices vs experienced sportspeople).

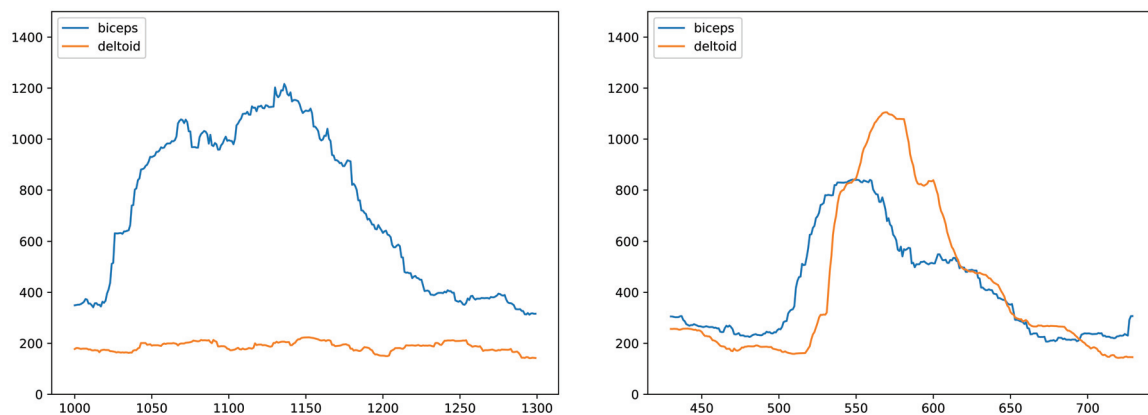


Figure 5.1: Example EMG data for a correct (left) and incorrect (right) bicep curl. The incorrect example clearly shows excessive deltoid use (orange) compared to bicep use (blue). Time (in samples) on the x-axis; muscle activation (as RMS values) on the y-axis.

Design decisions for feedback modality and granularity are informed by past work on EMG-based biofeedback [206], self-reflection [182] and by the design space of personal visualizations [126]: Data, Insight, Context, Interaction. Visual feedback enables detailed and concise feedback and allows users to observe different muscles [181] or different muscle regions [121]. Varying temporal granularity additionally enables users to capture the timing of their muscle activation [4]. However, visual feedback can be problematic where users

are required to visually observe and control their actions. Hence, we introduce auditory feedback as another modality. For both modalities, we implement the most commonly used indices [206]: EMG *amplitude* and *timing* of muscle activation as detailed in the following sections. We employed a mixed design that used the visualizations as within-subject and the sound cues as between-subject factor. All participants were instructed to exercise with each visualization while the presence of the sound cues was varied per participant. We note, that the employed modalities and indices are predominantly used in the introduced commercial products as well. While there are other viable feedback modalities (e.g. vibrotactile), the investigation of this work is primarily focused on visual feedback, due to the high bandwidth constraints of conveying rich EMG data.

Visual Feedback

FitBack implements three different feedback visualizations that change according to the measured muscle activity: *Bars*, *Circles*, and *Lines* as illustrated in Figure 5.2. *Bars* and *Circles* present an abstracted EMG signal to the user based on the EMG *amplitude* (cf. *Status* [182]). For *Bars* the height of each bar grows linearly, while *Circles* realizes linear growth of each circle’s radius, resulting in quadratic growth of the respective area. Both are inspired by a metaphoric approach [76].

In contrast, the *Lines* visualization shows a smoothed version (cf. Section 4.3.2) of the raw EMG signal to the user over time (cf. *History* [182]), similar to one of Mpower’s visualizations. Here, we hypothesize that abstract visualizations can increase the understanding of laymen of the complex signal [47]. Direct representations of the signal, such as the *Lines* representation, provide deeper insights into the training efficiency (cf. *Context* [182]), offering a finer granularity for *amplitude* and *timing*. While this visualization is potentially more difficult to understand for novices, it offers most details and the biggest potential for user-driven interpretation.

All visual representations provide a target zone that needs to be reached during the execution of an exercise (cf. *Goals* [182]), such as high bicep activation marked by an orange area at the far end of the respective visualization. An avoid zone (cf. *Discrepancies* [182]) is implemented for the deltoid in analog fashion.

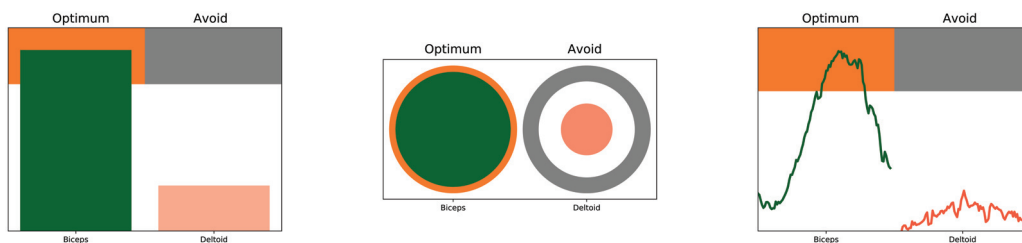


Figure 5.2: FitBack visualizations showing the same EMG data. Abstract representations *Bars* (left) and *Circles* (middle), and the detailed representation *Lines* (right).

Auditory Feedback

In addition to visual feedback, FitBack utilizes auditory feedback to communicate correctly executed repetitions of an exercise. Apart from their use in biofeedback [206], auditory cues have been verified as a suitable modality for successful actions [91], warnings [93], or errors [166] in various research projects. FitBack plays a positive auditory cue when a repetition has been executed correctly (cf. *Goals* [182]), while faulty exercise form (cf. *Discrepancies* [182]) is reported by a negative cue. In our case, the respective sound is played when the measured signal reaches the target or avoid zone.

Measures

For our data collection, we focused on the the impact of feedback type on exercise form through rankings, questionnaires and interviews.

Custom Questionnaire

We asked custom questions which addressed feedback ranking and perception. We asked participants to rate their agreement with the respective statements listed in Table 5.1. All ratings were measured using a visual analog scale (VAS) from 0 to 100.

Ranking and perception of feedback	
Q1	Please rank the feedback.
Q2	The feedback dictated how I performed the exercise.
Q3	I did take notice about the feedback while exercising.
Q4	I used the feedback to improve my exercise performance.
Q5	The feedback did affect my exercise form.

Table 5.1: Additional questions after each feedback type. Q1 from *extremely confusing* to *easy to understand*; Q2-Q5 from *not at all* to *very much*; all visual analog scale (0 to 100).

Interviews

We conducted semi-structured post-hoc interviews with all participants. Interviews took approximately 10 minutes each. We made sure to cover relevant topics including technical elements such as influence of electrodes and perceived latency of the shown feedback. Additionally, we inquired about their perceived confidence in the feedback and had them enumerate some of the advantages and disadvantages of different feedback types from their view, including auditory feedback if applicable. Lastly, we investigated possible application scenarios with FitBack.

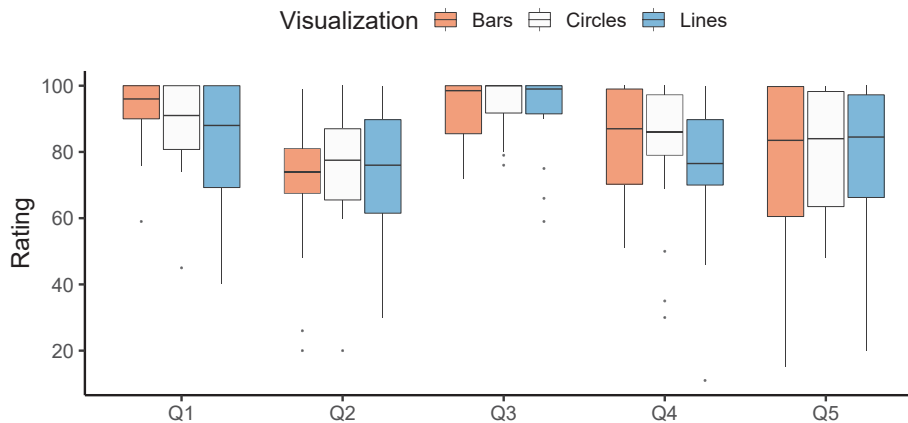


Figure 5.3: Mean scores for Questions Q1-Q5; assessing ranking and perception of feedback (see Table 5.1).

Participants and Procedure

Participants information and procedure of the experiment are given in Section 4.3.3. We kindly ask the reader to refer to this section for more details.

For our first study, we recruited 18 participants (four female, 14 male) with a mean age of $\bar{x} = 25.6y$ ($s = 3.24y$). After attaching electrodes, participants were tasked with repeating the bicep curl exercise with the support of the three visualizations.

Results

We report on the statistical analysis of feedback ranking and perception. For each measure, we conducted two-way (*visualization* and *sound*) mixed ANOVAs. If the data deviated from normality, we first aligned rank transformed [328] it. Further, we include a qualitative analysis based on the interviews.

Feedback Ranking and Perception

Our analysis did not show any significant differences for Q1-Q5. Results on our custom questionnaire polling feedback ranking and perception for individual visualizations are shown in Figure 5.3. Please refer to Table 5.1 for the respective question identifier.

Interviews

All interviews (1:39 hrs of recording) were transcribed verbatim. We opted to conduct a focused analysis based on the pragmatic approach by Blandford et al. [24]. To do so, three researchers open coded a representative 17% of the material. In a discussion, the researchers agreed on an initial coding tree. The rest of the material was evenly split between the three coders and coded separately. A concluding discussion refined the coding tree and surfaced

a code hierarchy with four themes: FEEDBACK TYPE AND GRANULARITY, COGNITIVE EFFORT, GAINING INSIGHTS and USER ACCEPTANCE. The following section presents the content of each theme and associated quotes.

Feedback Type and Granularity Participants commented extensively on advantages and disadvantages of the different feedback types and granularities. All visualizations were understandable for the participants, but there were clear favorites in terms of when and how to use them. For example, the abstract visualizations (*Bars*, *Circles*) were perceived as straightforward and easy to interpret:

It was evident what needed to be done. When the [deltoid] was red, one had to correct one's shoulder movement. (P3)

The detailed *Lines* feedback allowed for a more detailed analysis of the executed exercise. Additionally, participants remarked that *Lines* provided them with a more profound temporal component, as it displayed a history of values that was not present for the abstract visualizations:

If you used the Lines, the feedback was more differentiated. (...) If you wanted to have it very detailed, then you would use the Lines. (P7)

The Lines had a temporal component; one could see the progress until the optimum was reached, even for a longer period. The other two only provided a snapshot. (P14)

Our users expressed the wish to have access to both visualization types (abstract and detailed) as both provide merit in different contexts, such as using abstract feedback for daily workouts but detailed feedback to improve exercise form:

For the daily workout, I would prefer the [abstract visualization], but would switch to Lines occasionally to see what I can improve. (P16)

When asked about the sound feedback, participants commented that they felt it was more discreet than the visualizations allowing them to focus on the exercise. Interestingly, the additional sound feedback provided some participants with a feeling of accomplishment whenever they perceived positive sound feedback:

(...) you really focus on the execution and do not have to look [at the feedback], but you can just perceive the sound, okay, I made it. (P11)

It was a short feeling of accomplishment. That motivated me. (...) that could be integrated well into one's training. (P17)

However, the effectiveness of sound feedback highly depended on the learner type, as some participants preferred the visual feedback, only regarding the sound feedback as a nice addition:

I have not paid attention to it. (...) I have not perceived it consciously. I have only observed the visual feedback. (P9)

Cognitive Effort Participants reported that the level of cognitive effort they needed to invest into understanding and interpreting the feedback varied. Abstract visualizations (*Bars, Circles*) are less cognitively demanding, while the detailed *Lines* feedback was more demanding and required additional concentration, occasionally interfering with exercise execution:

To just do one exercise; the lower fidelity is more pleasant and less cognitively demanding. (P13)

The Lines feedback was more difficult. (...) It was more complex, because there were all those spikes and I did not know exactly what to do without influencing it. (P3)

Interestingly, participants also reported that the feeling of being watched by the system made them concentrate more on the execution of the exercise:

I have concentrated more, because I knew that I was being watched. (P13)

Gaining Insights Participants commented on how FitBack helped them to understand their muscle activity during the exercise and supported them in finding mistakes and incorrect form. It helped them to establish an understanding of their own exertion:

If you are not that familiar with muscles, then the system shows this quite well and you get better awareness for specific muscles, especially those that one has not used. (P15)

By doing so, FitBack facilitated their learning process, making them aware of their movements and supporting correct exercise execution. Participants remarked that it was straightforward to map their movements to feedback provided by FitBack.

Sometimes I would maybe get to the top and I would be at the point where I would normally have started like the second phase, started descending exercise but it hadn't beeped yet so I would have to squeeze for longer. (P1)

FitBack also provided participants with an incentive to focus on their fitness goals. Achieving perfect form is motivating and trackable through the system. Here, participants remarked that FitBack can support users in learning new forms and perfecting them:

If you want to learn a new exercise, or you want to train a specific muscle group that you want to improve. (P10)

User Acceptance A final theme that we identified is user acceptance. One aspect of this theme is trust. The user needs to have trust in the system, believing that it is capable of accurately identifying exercise form. Most participants reported that they trust the accuracy of the system to a high degree. The supervision by the system was regarded as trustworthy, making users more aware of their exercise form. This aspect was especially beneficial during continued workout, where participants lost focus and got tired. The feedback from the system reminded them to stay alert:

As soon as you get tired, it was nice to get a [positive sound feedback] if you have done it correctly or a [negative sound feedback] if not. (P17)

Participants relied on the feedback from the system when correcting their exercise form. They appreciated that mistakes were transparent for them and could be rectified immediately:

If I know that I do something wrong and that this is visualized, I can clearly imagine how to get rid of it. (P9)

On a more technical side, we also investigated whether the induced latency for the abstract feedback posed any issue for participants. Most participants were able to recognize a slight delay, but reported that it did not influence their exercise rhythm.

I only perceived it when I was paying attention to it and tried out when it actually arrives. (...) but it was not perceivable for me during the exercise. (P16)

Given the nature of our system, it is worth investigating to what degree the required setup and placement of electrodes has affected user acceptance. Here, participants reported that the electrodes did not obstruct them. It was noted however, that the required setup needed to be kept minimal for FitBack to be feasible in a daily workout:

I have barely noticed them and they did not restrict me. (P6)

The difficult bit is the placement of the electrodes. I do not know if this effort is worth it. (P11)

However, if the workout is longer or benefits from close monitoring are more evident, e.g. when learning a new exercise, participants were less reluctant to spend the effort required to set up the system:

By all means, a workout is like longer than an hour, and if [you set up the system] it minimizes the risk of injury. (P3)

Another idea mentioned by the interviewees was the integration of electrodes into wearables such as sport shirts and trousers. This would minimize setup time and would make FitBack much more accessible for short daily workouts, also reducing social stigma as remarked by a few participants:

If there could be a wearable that includes the device and cables, e.g. a shirt with electrodes inside, that you can just wear, that would be different. Then you would just need to wear the shirt. (P6)

Summary

Our questionnaire indicated that *Bars* was ranked highest by participants. In our follow-up interviews we found that detailed feedback was more favored by experienced practitioners and in situations where the user wants to perfect exercise form. Abstract feedback, such as *Bars*, was preferred by less experienced users and for daily workout scenarios. In general, all forms of feedback were noticed by the participants and used to adapt their exercise form to a large degree as confirmed by our custom questionnaire, highlighting the viability of EMG-based feedback to facilitate bodily insights.

Interviewees reported that *Lines* required higher cognitive effort, mostly to understand the complex signal. This might have been masked by the fact that users were always monitored by the system, hence implicitly forcing them to concentrate more as stated in the interviews.

5.1.4 Study II - Interviews With Coaches

Our initial study confirmed the feasibility of building an EMG-based exercise assistance system and showed the breadth of the design space. In order to further understand the requirements and constraints involved in using EMG systems for physical activity, we conducted a series of interviews with sports professionals in which participants used FitBack in an open-ended exercise session.

Participants

We recruited four experts who were career sports coaches with different backgrounds and varying levels of experience, see Table 5.2. We recruited the experts through contacts at sports clubs whose members had participated in previous studies conducted by the research team. None of the participants had used EMG systems before, but they had all tried electrical muscle stimulation (EMS) as part of their profession. The interviews took place at locations chosen by the participants. We provided shopping vouchers for the equivalent of 10 Euros for participating in the study.

Procedure

The interview session began by obtaining written consent for participation and recording. Next, the researcher prepared the FitBack system in a location chosen by the participant. A voice recorder and a video camera were used to record the session. We then conducted the initial part of the interview, which concerned the participant's background and experience with EMG/EMS. Afterwards, we asked them to choose a movement form that was particularly difficult for their students and indicate which key muscles were involved in that

Participant ID	Age	Gender	Expertise	Experience
P1	32	M	Personal trainer	8 y
P2	27	F	Yoga	4 y
P3	36	M	Martial arts	15 y
P4	33	M	Martial arts	15 y

Table 5.2: Information about the participants in our study. The sports professionals specialized in different sports and had different levels of experience.



Figure 5.4: Participants during the exploration phase performing different exercises.

movement. Upon choosing the exercise, the researcher attached electrodes to the key muscles. FitBack was started and muscle activity was visible. We then asked the participant to perform the motion correctly and incorrectly. We encouraged them to explore different possible mistakes and to examine how FitBack reacted to changes in movement (see Figure 5.4). We also answered any questions the participant may have had while exploring the system. When they indicated that they were done experimenting with the motion, we conducted the next interview part. These questions explored the participant's interpretation of the EMG signal, differences between signals from different muscles, the suitability of the feedback for everyday coaching and identifying mistakes in movement. We then allowed the participant time to rest and repeated the same procedure for another type of motion. Finally, we conducted an interview where we inquired about the differences in feedback between the two exercises, the possible target user groups for EMG systems, the use of EMG for professional development and requirements for everyday integration. The entire session was recorded on video, and additional voice recording was used when asking interview questions. We also recorded the FitBack screen.

Data analysis

All interviews (2:16 hrs of recording) were transcribed verbatim. Given the volume of the data, we again followed the pragmatic approach to qualitative analysis as recommended by

Blandford et al. [24]. In an initial analysis step, two researchers open-coded one interview to identify key concepts. We then conducted a discussion, which resulted in an initial coding tree. The remaining material was then evenly split between the researchers. Afterwards, all authors refined the coding tree in iterative discussions, finally creating a code hierarchy with four themes. These themes represent the key aspects of using EMG in sport coaching practice discussed by the experts.

Findings

Here, we present the results of the interviews with the sports professionals. We identified four themes in the accounts of using EMG provided by the experts: ASSESSMENT, BODY AWARENESS, EXERCISE FORM and INTERPRETABILITY. Quotes from the interviews are accompanied by the participant ID and, if applicable, the exercise which they were describing.

Assessment This theme describes how the coaches used EMG feedback to assess the correctness of a given body position and how they identified issues with posture. We observed that the experts quickly reflected on EMG data and related the reading to their current movement. Interestingly, low EMG signal was as interesting as high muscle activation, because some exercises required relaxing muscles:

This is very easy to apply when I'm in a static position, which I want to hold. When I'm doing this correctly, I am more or less relaxed. (P4, cadeira)

The experts also described another way to effectively use EMG to assess correctness which was to determine if the right muscles were active during a specific movement. P1 explicitly addressed how specific muscles should work during push-ups.

I can see if I'm activating the deltoid when moving up, as I should. [FitBack] tells me if I'm beginning the movement by raising my shoulder and using the descending part of the trapezius. This tells me if I'm doing it wrong. (P1, push-ups)

Finally, he also reflected on the impact that being able to objectively assess an exercise would have on fitness instruction. Being able to confirm one's professional opinion with objective data was seen as an opportunity for proving the coach's credibility:

If I tell someone that they're using the wrong muscle, but they haven't learnt to control it yet, they will think I'm overdoing it or looking for a problem to solve for the next 20, paid, classes. (...) This gives them the possibility to see I'm right. (P1)

Body Awareness We also observed that the experts remarked extensively on how using EMG potentially led to additional insight about one's body. EMG was perceived as a way to see beyond what they could perceive using their professional insight. P2 remarked that FitBack could be used to determine what muscles were targeted in new exercises, previously

unknown to the coach and their students. One could even verify if publicly available new exercises targeted the muscles which they claimed to target:

You often see those things online: “100 best planks” etc... but you never fully know what these exercises do. With [FitBack], you can see what’s inside, check if [the muscle] is even contracting. (P2)

The form in which the EMG signal was presented affected the experts’ perceptions of their bodies when using FitBack. While our system featured graphs with an adaptive scale on the Y axis, the experts were still curious about the maximum values. Concurrently, they also expressed the need for effective comparisons between exercises:

I’m wondering what would happen if I contracted this at 100%. Would this explode? I would like to zoom this out to see if my deltoid is more active in arm raises or press-ups. (P4)

EMG also allowed the experts to better understand the transitions between different body postures and movement forms. While they knew what muscles were to be active (or inactive) during defined exercises, EMG allowed investigation into what happened in the intermediate stages of the movement. P3 analyzed how his body worked while adjusting his posture in a complicated handstand:

When I changed the alignment of my body, moved my legs, you could ideally see where I contracted the muscle very hard and where it was weaker. (P3, queda de rins)

Exercise Form This theme describes how EMG was used to reflect different exercise styles and different ways of understanding physical activity by coaches and their students. The coaches were eager to speculate on how different types of their class attendees may have perceived the feedback. In contrast to ASSESSMENT, here, the experts described implicit, less defined qualities that can be inferred from EMG data. The need to place EMG electrodes implicitly required reflecting about the key muscles involved in the exercise. This, in turn, made the coaches wonder what other muscles, not explicitly targeted by the exercise, were involved:

Figuring out where to place the electrodes is interesting by itself. If I were to do this myself, I would experiment with this, stick it somewhere, check it out. (P4)

The additional information provided by FitBack facilitated exercising while being aware of the benefits of the exercise. P1 commented extensively on how past philosophies of training promoted maximum exertion, while more modern methods preferred targeting specific muscles and avoiding injury. He saw EMG as a way to facilitate this transition from simple exertion to informed practice:

All the way, just to get super tired and feel pain. If you can't push through the pain, then you're weak. (...) If you make all the exercises from the exercise book, you will exercise all your muscles. But, the modern school of training is not about lifting very heavy bars. You're supposed to slowly learn how to activate specific muscles. I think that most people appreciate that in the long run. (P1)

The coaches provided many comments on the temporal aspect of EMG data. They reflected that timing was differently important in various exercises and requested more control over the time scale. This was most apparent in the comments provided by the martial arts specialists, P3 and P4. They remarked that a change of tempo and thus change of the speed of muscle activation was a key aspect of their practice. P3 asked for more flexibility in the time scale to be able to analyze more dynamic exercises:

It can be used for all the static exercises. I imagined that for dynamic things, it might be different. Capoeira has the full spectrum of exercises and thus paying attention to relaxing muscles is important. (P3)

Interpretability The coaches commented extensively on how they interpreted the output produced by FitBack. They wondered about the form and timing of the feedback and anticipated possible deployment in their exercise classes. The experts anticipated how their students would understand EMG data and how FitBack could benefit the training process. The coaches remarked that EMG in the form presented by FitBack would be too difficult to monitor for larger fitness classes:

It would be tough for classes where we have a lot of people. One coach won't be able to interpret the muscle behavior of many people. But, for personal training, this would be very suitable. (P3)

The experts remarked that EMG had the potential for changing achieving proper form in exercises from a vague pursuit of correctness to a tangible goal. P4 reflected on how a well-known exercise was a challenge when FitBack provided detailed feedback:

It was very stimulating because I was trying to do the exercise better so that it was more visible on the screen, (...) because I like tangible challenges. (P4, primera)

Our experts underlined the importance of associating specific EMG responses with proprioceptive perception, i.e. 'how a movement felt'. As most fitness classes focused on mastering specific forms and limiting the use of incorrect forms, quantifying how far one was from an ideal movement was key to interpreting the EMG data and one's body. One participant saw FitBack as a tool for iteratively developing more precise movements and fine-tuning one's performance. EMG data enabled minuscule changes to posture which would be visible in the measurement data, but hard to perceive with one's senses.

In functional training, where the key goal is developing correct movement patterns, the basic movements are key. Here, [FitBack] can be very useful, guiding you to find the ideal ratio, the perfect movement. You can then copy the movement and master its execution. (P3)

5.1.5 Discussion

Our investigation showed the potential of EMG-based biofeedback to facilitate bodily insights during physical activity (**RQ3a**). In this section, we draw implications and present opportunities for future systems. We conclude with a list of guidelines for designing for EMG-based biofeedback exercises that allows users to increase their own body awareness.

Access to Detail is Beneficial for Practitioners and Coaches Alike

In our analysis, the detailed feedback was favored mostly by experienced sportspeople. Our interviews with professionals revealed that the additional temporal component allowed them to analyze their movements in greater depth and supported them in perfecting their exercise form. Often, minuscule improvements were only visible in the EMG data, allowing **coaches to monitor otherwise invisible muscle activity and curate an appropriate training response**.

Additionally, both studies highlighted that the detailed temporal feedback was especially useful for dynamic exercises where the correct sequencing of muscle movements is crucial. Accordingly, an EMG-based biofeedback system should always provide a feedback view that allows to examine historic data. Future systems can improve the interaction with this view, by providing exercise markers within the view and even a corresponding recording functionality.

Required Setup for EMG-Based Biofeedback Needs to be Minimal

To record electromyograms, FitBack requires the user to attach electrodes on respective muscle groups. This is a major limitation of not only our system but every EMG-based system. While participants reported that the attached electrodes and cables did not hinder their exercise execution, the required time for setting up the system was deemed acceptable only for continued workouts or when one wanted to gain additional insights. For daily workouts, the required effort was rarely deemed acceptable. In this work, we explicitly focused on user requirements and constraints with regard to the suitability of EMG-based biofeedback to increase body awareness during physical activity. We envision that more **biofeedback systems will be integrated into affordable wearable artifacts and clothing** in the near future, enabling easy and fast setup routines as is already possible with commercial products for professionals.

Design Guidelines for EMG-Based Biofeedback

During our work, we identified opportunities, requirements, and constraints of EMG-based biofeedback for physical activity. In this section, we provide a list of design guidelines for EMG-based biofeedback systems to inspire and aid in the design of future EMG systems for sports.

Offer Feedback Customization

The different user groups in our study had differing requirements in terms of feedback. FitBack offers two different feedback modalities (visual and auditory) as well as abstract and detailed representations of the visual feedback. In our evaluation, we found that while the *Bars* visualization was most preferred, our interviews have also shown that access to multiple types allows the system to tailor to a broad audience. There was no feedback form that would not affect exercise form or was ignored by the participants. Novice users preferred abstract feedback that was less cognitively demanding and easy to interpret, cf. COGNITIVE EFFORT. On the other hand, switching to more detailed feedback should always be possible for more experienced sportspeople. In the interviews with coaches, we observed that the understanding of feedback can be different for different movements within the same sport. Thus, **an EMG feedback system for physical activity must offer output that dynamically changes based on the current activity.**

Users exhibited varying preferences with regard to audio feedback. While some participants reported that it helped them keep their focus during a continued workout without having to look at the visual feedback, others strongly preferred visual cues, cf. FEEDBACK TYPE AND GRANULARITY. Thus, **multi-modal feedback systems should provide a choice for the user to prevent sensory overload.**

Allow for Temporal Control

While our experimental study used only immediate feedback, the interviews with professionals revealed that EMG data was meaningful for understanding motion in different time horizons. The experts commented on their need to experiment and establish thresholds of correctness for different exercises, cf. EXERCISE FORM. They also wanted to establish baselines and be able to review past exercises to draw comparisons, similarly to existing commercial products. They recognized the benefits of EMG data both in instantaneous use and over entire exercise session. As a consequence, **future EMG systems for exercise should offer a high degree of control in terms of the time intervals used to aggregate and display EMG data.**

Design for Social Context

Both of our studies showed that participants were eager to consider using FitBack in contexts associated with everyday exercise. What became apparent in the interviews was that these contexts were primarily social and possible social interaction impacted possible interpretations of EMG data. As we observed in the INTERPRETABILITY theme, experts reported

that feedback modalities for individual sessions should be very different from the feedback in group fitness classes. Further, they also noted that the use of EMG could be dependent on a certain fitness philosophy or mental approaches to a given exercise. Participants in both studies reported that they often changed the social contexts of their physical activity. Thus, we recommend that **the granularity of EMG feedback be adjusted to allow for adequate interpretation**. Producing detailed EMG feedback in a context where it cannot be effectively interpreted is likely to cause frustration. Future systems should offer detailed data for individual classes and self-monitoring and ambient and/or summary feedback that would enable coaches in group exercises to monitor the activity of multiple users.

5.1.6 Conclusion

In this use case, we investigated how EMG-based biofeedback can facilitate bodily insight during physical activity. Our two-fold evaluation consisting of a lab study with amateur users and interview sessions with sports coaches highlights the feasibility of this approach. We present implications and design guidelines for future EMG-based proficiency-aware systems, allowing users to gain deeper insights into their own physiology by leveraging user-driven interpretation of bodily signals. When used in a specific exercise context, FitBack can provide data, which help users gain more insight into their exercise. Our work showed that feedback about one's muscle activity and exertion is beneficial for a broad audience, from novices to fitness professionals. FitBack offered insights for all target audiences by providing different views of the recorded EMG data, facilitating proficiency awareness.

As we have observed that it is often sufficient to provide simple feedback of collected data to the user to support proficiency awareness, we investigate these implications for everyday computer work. In the next use case, we utilize a common text production task to evaluate if feedback on the text's readability makes users reconsider their task approach.

5.2 Use Case II: Leveraging Reflective Feedback in Text Production

In this use case, we illustrate the concept of proficiency-aware systems with regard to everyday computing work. To do so, we selected a text production task, which we used to conduct two online studies. In Study I, we establish if feedback on proficiency can improve performance, reflection and user experience in a text production task. Having observed that proficiency feedback offered benefits, we investigate how to design better feedback in Study II.

Our results showed that proficiency feedback enabled users to fulfill task requirements more precisely. While continuous assessment showed the largest improvement in performance, users also reported that this type of feedback interrupted and distracted them during the

text production task. A revision-based feedback approach was perceived as less distracting while still achieving better-than-baseline task performance. We also found that users adapted their approach in solving the task based on the feedback presented, which indicates that they were aware of their proficiency. Adding gamification elements only marginally influenced performance and user experience. The results suggest that subtle interface elements such as progress bars are most suited for proficiency communication when interacting with proficiency-aware systems.

5.2.1 Related Research

We first give an overview of related works on novices to experts transition and how this research area informs our investigation. Additionally, we have a closer look at the impact of user interruptions and reflection as well as requirements and opportunities of gamification in user interface design.

Novice to Expert Transition

An active research area looks at novices to expert transition [45] with regard to alternative expert interfacing options, such as hotkey usage. *Blur* [270] is a system that realizes this switch using subtle notifications, hence making users aware of an alternative interfacing method. However, to have a better outreach, Stuerzlinger et al. [296] argue that adaption needs to be available everywhere. It needs to become ubiquitous and not limited to individual applications. Hence, it is vital to understand what aspects drive the user to customize [12] or wanting to adapt the interface [81]. This implies that we need to investigate *when* the user chooses to adapt their current strategy and *what* makes them rethink their current approach. For our work, we consequently investigate different types of proficiency feedback that vary in temporality and noticeability.

A Take on Feedback: Interruptions and Reflection

While research has actively investigated *how* to construct fruitful [215, 48] and expert-level [337, 184] feedback, the scope of this current work addresses if feedback on task performance metrics can elicit an understanding for the user of their own proficiency in the given task. In this regard, previous research advocates to make the transition to high proficiency easy and fast [202]. A dilemma arises in that interruptions of the user's task, especially if frequent [254], are detrimental [133], yet research argues that without interruptions and associated focus shifts [25], there can be no opening for learning, hence no opportunity to improve proficiency. Malacria et al. [190] have highlighted a similar situation in their work *Skillometers*, where feedback needed to be visible to catch the user's eye while simultaneously being subtle to minimize disruption. Our work expands on this concept by following their solution (manipulating locus of control) and providing proficiency feedback via an ambient display element. Additionally, we purposely interrupt the user (revision-based method) after finishing their task to reflect on their current performance.

Consequently, we conducted a structured exploration of the design constraints involved in building proficiency-aware systems for everyday computer work.

Gamification: Requirements and Opportunities

Recognizing challenges and opportunities around proficiency feedback, distraction, and learning, we explored uses and impacts of game design elements in supporting proficiency feedback communication. Gamification, "the use of game design elements in non-game contexts" [58], has proven to motivate and support learning in formal [297, 13] and informal settings [218]. Gamification is further used to create motivation, engagement, and joy across a wide range of applications, ranging from sports [164] to work environments [222]. Badges, points, and leaderboards are the most common game design elements [107, 282] that can help turn routine and boring tasks into enjoyable experiences [87]. However, recent gamification research emphasizes the need to review and consider game elements beyond those standard ones. The mapping of different game elements and their effects to applications and users has gained increasing attention [309, 223]. This reflects our evolved understanding, according to which applying standard game elements only to drive business goals risks losing positive effects and even alienates users [216].

In order to understand why gamification promises to play a key role in proficiency feedback, it is important to understand key mechanisms and theories as well as differences between the various game design elements. *Flow* [213, 208] is both a process and theory used to inform gamification design. A person finding themselves in a *flow* state is fully immersed in an activity which they consider enjoyable and fulfilling. Up to nine dimensions are commonly described which contribute to a flow experience [106, 208]. Of those dimensions, challenge-skill-balance, clear goals, and feedback are naturally compatible with proficiency feedback. Self-Determination Theory (SDT) [251] is another theory that is concerned with the interplay between extrinsic motivation (e.g. rewards, evaluations, fear of punishment) and intrinsic motivation, i.e. motivation created and sustained within the self by curiosity, interest, or identified values. SDT and its mini-theories have been extensively used to theorize and develop gamification research and applications. In particular, the Basic Psychological Needs Theory (BPNT), one of the six SDT mini theories, has been used to explain and develop gamification designs [315]. The BPNT focuses on three basic psychological needs: *competence*, *autonomy*, and *relatedness*. Environments and applications supporting those needs promote psychological wellbeing and intrinsic motivation.

In contrast, the Organismic Integration Theory (OIT) is concerned with various forms of extrinsic motivation. OIT "recognizes that some behavioral regulations are experienced as relatively alien to the self, or imposed and heteronomous, whereas others can be very much being autonomous and self-endorsed" [252]. OIT refers to a spectrum of internalization of values represented by a task or environment and distinguishes along this continuum between external regulation (i.e. the least self-determined form), introjected regulation, identified regulation, and integrated regulation (i.e. the most self-determined form of extrinsic motivation). We describe in the context of Study II, how we based our selection of game design elements and evaluation metrics closely on the BPNT and OIT.

5.2.2 Methodology

In our research, we want to investigate the impact of proficiency feedback on the user's performance and experience. Ultimately, we ask what are the design criteria for proficiency awareness in interactive systems.

RQ3

What are design criteria for proficiency awareness in interactive systems?

To operationalize these concepts, we formulated three related research questions tailored towards this use case. In short, these questions ask if, when and how proficiency feedback can be effective. We elaborate on the questions in detail below.

RQ3b: Can proficiency feedback facilitate an understanding of one's own proficiency in a writing task?

A proficiency-aware system should assist the user in improving their skill level to facilitate an understanding for proficiency, e.g. feedback on improvements, reasons for failure. It would not directly address task performance, but rather assess the underlying skill set necessary to complete this task.

While input error handling happens to protect the system from wrong input, proficiency feedback would enable users to prevent wrong input in the first place, much like teaching someone how to play a guitar rather than altering the guitar to fit their play style and skill level. We address this research question by analyzing whether users change their task solving strategy after being presented with feedback.

Additionally, we investigate how to appropriately communicate proficiency feedback to the user. Research has postulated that the willingness to customize or adapt one's strategy during task solving might be linked to task proficiency [193], as in power users adapting more willingly because they can grasp its benefits. For this reason, we selected an everyday task of expressing an argument in a submitted proposal to limit the impact of prior proficiency levels.

RQ3c: When should proficiency feedback be presented to users for optimal performance and user experience?

We implemented a continuous feedback option that is updated whenever the user changes the text. While this increases the user's awareness of the functionality, it might distract them from the primary task, yet conversely, breakdowns and associated focus shifts "can be openings for learning" [25]. To strike a balance, we additionally implemented a revision-based method that presented the user with a proficiency evaluation upon completing their initial proposal and let them reflect on their current performance.

RQ3d: Can gamification support proficiency feedback?

Feedback can distract or even annoy users. This is unfortunate, as proficiency feedback should really be perceived as a supportive and useful resource that helps users improve and reach a target goal on their own with confidence and competence. Given the potential of gamification to communicate information, status, and achievements in an enjoyable and motivating manner, we explored the spectrum of game design elements and chose to implement game elements in our work: progress bars, social comparison, and emojis. Our evaluation focused not only on assessing the text production task, but included impact on components of intrinsic and extrinsic motivation, as well as perceived distraction.

Method

We based our investigation on two studies as listed in Table 5.3. The full 3×3 design looks at the independent variables *Feedback Type* and *Gamification Type*, both of which have three levels. Study I first looks at the impact of *Feedback Type* in isolation, while Study II focuses additionally on the gamification part.

	<i>No Gamification</i>	<i>Progress Bar</i>	<i>Emoji</i>
<i>No Feedback</i>	Study I	N/A	N/A
<i>Revision</i>	Study I and II	Study II	Study II
<i>Continuous</i>	Study I and II	Study II	Study II

Table 5.3: Study designs and respective conditions.

We based our scenario in the domain of everyday computing tasks. Participants were asked to express an argument in a proposal using a submission form on a web page. As mentioned above, adapting one’s strategy during task solving might be linked to task proficiency [193], thus we have chosen a simple text production task to limit the impact of prior proficiency levels. Filling out forms on the web is a mundane task and users are familiar with the environment. Further, producing text to express our opinion happens daily, be it in emails, essays or articles. Improving the user experience and secondary benefits (increasing proficiency) for these tasks can thus be contributory on a large scale. Ultimately, broad adoption can foster data-driven algorithms enabling sophisticated recommender systems [200].

We want to highlight that such a proficiency-aware system **does not directly provide feedback on task performance** — writing a good proposal — but rather **assesses and communicates the necessary skill** — here: writing in plain English — of the user to complete the primary task in the first place. Consequently, the system is also not limited to specific texts, but can be employed for any text production task. In our two-study evaluation, we first focus on the temporality of proficiency feedback (**RQ3b**, **RQ3c**) in Study I, before exploring gamification (**RQ3d**) as a means to make feedback more enjoyable and usable in Study II.

For this purpose, we implemented a system that assesses a user’s submitted text proposal in term of readability. Proficiency feedback is provided as an additional display element next

to the form (see Figure 5.5 for an example). To do so, we leverage the Flesch reading-ease score (FRES) [85] which indicates how difficult it is to understand a given text. It is based on average-sentence length (ASL) and average number of syllables per word (ASW). The resulting Flesch reading-ease score can be computed as follows:

$$FRES(ASL, ASW) = 206.835 - 1.015 * (ASL) - 84.6 * (ASW) \quad (5.1)$$

A score between 60 and 70 is interpreted as plain English. Texts with higher scores are easier to read but can be too simplistic. A lower score indicates a more difficult text. The score is widely used for evaluating a text's readability, e.g. the state of Florida requires insurance policies to have a FRES of at least 45 [305].

Your proposal

Your word count: 42

Your current readability score is: **63.7**

Dear Sir or Madam,

my name is John Smith and I have been invited to submit a proposal to make our city more green. It has long been a concern of mine that our lovely town has been pestered by industry pollution.

Too difficult **Too simple**

Only 5% of texts in similar tasks fulfil the requirements better

SEND PROPOSAL

Figure 5.5: Example view of our proficiency-aware web form, showing the progress bar and ranking condition.

In our investigation, the user is tasked with writing a proposal to make the neighborhood more green in plain English (FRES between 60 and 70), which will be presented to the city council; a task taken from a preparation course for the Cambridge English Advanced exam [101]. Participants will be judged by their ability to write a concise proposal in plain English. The complete task description is as follows:

You see this announcement on the notice board of your local council. The Environmental Planning Committee is organising a campaign to make our town more 'green'. You are invited to submit a proposal related to your neighborhood, which will be presented to the city council. Present some factual information about the area, pointing out any relevant environmental issues, and suggest practical measures which individuals and families could take to make the neighborhood more green. Write your proposal in 220–260 words in plain English. Your proposal should be readable by a broad audience.

For both studies, task and procedure remained consistent apart from adapted questionnaires. The following sections detail conditions and measures for Study I and II.

5.2.3 Study I - Temporal Granularity of Proficiency Feedback

In this first evaluation, we addressed the temporal granularity of the provided proficiency feedback (cf. Table 5.3). We specifically looked at *Revision*-based and *Continuous* feedback methods. Together with a *No Feedback* condition, this one-factorial design resulted in three levels.

Revision-Based

After submitting the proposal once, the user was informed about their last readability score. Upon resubmission, the user was able to make changes to achieve a better result if they deemed this necessary. During this revision process, the last score was always visible. Participants were instructed accordingly:

Your proposal will be automatically scored in terms of reading difficulty after submission. You will have the option to revise your proposal afterwards once. You should aim for a score between 60 and 70. Higher scores indicate a too simplistic proposal, while lower scores indicate low readability.

Continuous

In this condition, participants' readability scores were calculated at fixed time intervals (2 seconds) and displayed. The score was always visible next to the text entry field. Figure 5.5 shows the *Continuous* feedback for a gamified condition (cf. Section 5.2.4). Again, participants were instructed accordingly:

Your proposal will be continuously scored in terms of reading difficulty. You should aim for a score between 60 and 70. Higher scores indicate a too simplistic proposal, while lower scores indicate low readability.

No Feedback

As a baseline we added a *No Feedback* condition, where users were not given any feedback on their readability score throughout the whole writing process. Consequently, no additions to the task description were made.

Hypotheses

Based on results from related work, we formulated three hypotheses:

H1a: Continuous proficiency feedback leads to increased task performance.

We hypothesize that the more users are confronted with feedback about their proficiency, the better they can solve a given task. In our case, performance was measured two-fold: (1) being able to submit an adequate proposal and (2) getting one's readability score close to the target zone. We analyzed final readability scores of valid proposals and looked at the change of readability over time.

H2a: Revision-based feedback is less distracting.

We know from related work that constant feedback can be detrimental (cf. Section 5.2.1). Hence, we hypothesize that more subtle and less constant feedback is less distracting for users. We measured this with tailored questions.

H3a: Proficiency feedback alters the task approach of users.

As outlined in Section 5.2.2, a proficiency-aware system should incentivize users to improve their inherent skill set. It is thus important to investigate if users alter their task solving approach to adhere to task constraints. We analyzed this aspect with tailored questions as well as investigating changes in writing behavior (readability over time).

Participants

We recruited a total of $N = 70$ participants⁵ over the Amazon Mechanical Turk Service (MTurk). Participants were reimbursed with 2\$ and promised a 1\$ bonus for an acceptable proposal text⁶. This rate was approved by the institution of the first author (equivalent of 10€/hour). Out of these participants, 19 resided in the European Economic Area, 18 in Canada and 33 in the USA. All participants were informed that study participation was voluntary, that the study could be aborted at any point and that the data would be collected in anonymized form. It took approximately 20 minutes to complete the survey, including writing the proposal text. The average age of the participants was $\bar{x} = 35.4y$ ($s = 8.9y$) with 29% identifying as female, 71% as male. Additionally we asked all participants to provide their self-assessed writing skills [46] (see Table 5.4). After the final submission of the proposal text, we again asked them to evaluate their writing quality based on an adapted scale [137] (see Table 5.5).

The writing assessment tests helped us to assess whether participants exhibited the necessary skills to complete the writing task. We found that participants rated themselves highly proficient in this writing task: $\bar{x} = 13.4$ ($s = 1.8$) for the writing self assessment (max score: 15) and $\bar{x} = 65.4$ ($s = 11.0$) for the writing quality (max score: 80) of their proposal. This confirms that all participants were sufficiently proficient to execute the given writing task. Additionally, the experimenter evaluated each proposal for correctness and adequacy.

⁵ Final number of participants, after rejecting obvious outliers in terms of proposal quality and questionnaire responses.

⁶ Verified post-hoc by the experimenter.

Please rate your skills in writing in the following areas:

Use of Proper Grammar
 Correct Spelling and Punctuation
 Ability to Organize Ideas

Table 5.4: Writing self-assessment questions (adapted from [46]). Five-item Likert scale from *incompetent* to *very competent*.

Please rate your agreement with the following statements:

My proposal correctly follows writing conventions (spelling, punctuation and grammar).
 My proposal is enjoyable and interesting to read.
 My proposal is enjoyable and interesting to read.
 My proposal is organized well (structure, introduction, and conclusion).
 My proposal is persuasive and convincing.
 My proposal demonstrates effective sentence fluency (rhythm, flow, and variety).
 My proposal demonstrates a clear voice (personality and sense of audience).
 My proposal uses effective word choice (precise and vivid wording).

Table 5.5: Writing quality assessment by participants for their proposal text (adapted from [137]). Ten-item Likert scale from *very strongly disagree* to *very strongly agree*.

Procedure

After providing informed consent, participants were asked to provide demographics and assess their writing skills. They were then randomly assigned to one of the three conditions and were given the task to write a proposal for the city council in which they explained their ideas to make the city greener. Depending on condition, they were provided with *No Feedback*, *Revision*-based or *Continuous* feedback of their current readability score. Participants were made aware of these conditions, by telling them that their proposal would be scored and how often this would be the case.

After completing their proposal, participants were asked to fill a raw NASA-TLX [110], assess their writing quality and asked custom questions tailored at their perception of the scoring system, including accuracy and perceived disruptiveness (see Table 5.6). Note that the custom questions were only present for conditions *Revision* and *Continuous*.

Results

For the final dataset, conditions were distributed as follows: *No Feedback*: 24, *Revision*: 24 and *Continuous*: 22 entries. We report our analysis on the following metrics as collected in Study I: task completion time (TCT), NASA-TLX, the final Flesch reading-ease score (FRES) and its deviation from the target zone (60 to 70). Additionally, we take a look at the FRES over time for each condition and analyze in particular whether the *Revision*-based method has prompted participants to alter their proposal. Lastly, our custom questions gave

Perception of the scoring system	
Q1a	I felt that my performance was accurately assessed.
Q2a	I felt pressured by the scoring system.
Q3a	I performed better using the scoring system.
Q4a	The system interrupted me during the task.
Q5a	I could have done the task without the scoring system.

Table 5.6: Additional questions for conditions *Revision* and *Continuous*; from *strongly disagree* to *strongly agree*; all visual analog scale (0 to 100).

insights into the disruptiveness of each feedback method. If not stated otherwise, we conducted one-way ANOVAs to analyze the data. If normality was violated, we first aligned rank transformed [328] the data. All tests (if necessary) were adjusted for multiple comparison using the Tukey method. Effect sizes are given using η^2 (Partial Eta Squared): small ($> .01$), medium ($> .06$), large ($> .14$).

Task Completion Time, NASA-TLX and Final Readability Score

We did not observe any significant difference between the conditions (*No Feedback*, *Revision*, *Continuous*) for task completion time, nor for the NASA-TLX score or the final FRES. Deviation from the target zone (measured in absolute deviation from 65) was also not significant. An overview is provided in Table 5.7.

Condition	TCT in s		NASA-TLX		Final FRES		Final FRES deviation	
	\bar{x}	s	\bar{x}	s	\bar{x}	s	\bar{x}	s
<i>No Feedback</i>	903	746	65.3	21.0	48.8	15.5	12.8	9.0
<i>Revision</i>	973	572	66.2	23.0	50.3	13.0	12.4	9.3
<i>Continuous</i>	1002	761	62.3	19.1	55.6	11.8	7.0	6.4

Table 5.7: TCT, NASA-TLX, final FRES and its deviation from the target zone (60 to 70).

Influence on Readability Score

To further evaluate the influence of the feedback methods on the participants' readability scores, we compared the temporal course of each condition. Since task completion times varied across participants, we rescaled all trials to the median task completion time of 758 s. This allowed us to visually compare the different trials at once and draw conclusions based on the mean and the standard deviation of the readability score over time. An overview of this analysis is shown in Figure 5.6. For each condition, valid⁷ trials are plotted as thin blue lines showing the participant's readability score over time. Further, a red line marks

⁷ A researcher cross-referenced each submitted proposal with the respective time series for the readability score. Trials were omitted if it was evident that the submitted text was just copied into the form.

the average (at any given point in time) over all participants, while the red corridor marks the standard deviation, analogously. Additionally, for *Revision*, black vertical lines show the point in time when the participants started the revision of their text (mean plotted thick). Note that the FRES is highly volatile for short texts during creation. Hence, we omitted the first 100 seconds for these plots.

Figure 5.6 illustrates that participants in the *Continuous* condition narrowed down faster on the target zone. The *No Feedback* condition was worst in this regard. The final deviation over all participants per condition was lowest for *Continuous*. *Revision* and *No Feedback* conditions exhibited the highest variances for the final readability score.

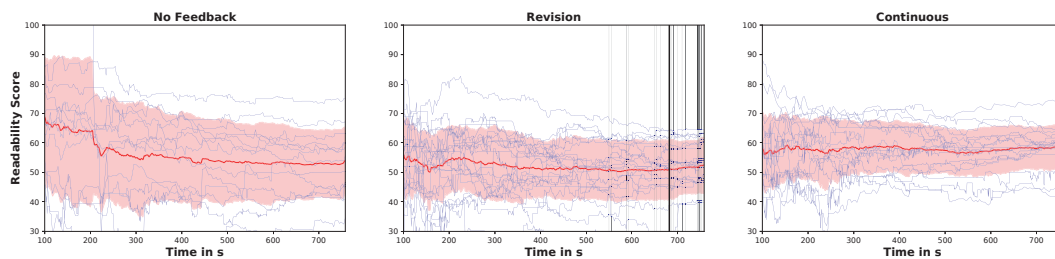


Figure 5.6: Deviation of readability score over time. The red-shaded corridor shows standard deviation over the filtered data (red line). Condition *Revision* (middle) additionally shows timings for the revision prompt (black vertical lines; thick line being the mean). Note that all trials have been rescaled to the median answer time and start at $t = 100s$.

We additionally evaluated whether the revision prompt in the condition *Revision* had an impact on participants' writing behavior. To do so, we fitted a linear model for the collected FRES data after the revision prompt. We fitted the model with the averaged data after the mean revision prompt time (cf. thick black line in Figure 5.6). We then tested against a null model that simulated no change in writing behavior, in other words: no incline. We found a significant difference ($F(1, 41.7) = 346.0, p < .001, \eta^2 = .52$) between the models with a large effect size. This indicates that participants did indeed try to improve their readability score after the first revision.

Custom Questions

The analysis of our custom questions (see Table 5.6) revealed a significant difference (large effect) for Q4a: "The system interrupted me during the task.": $F(1, 44) = 9.92, p < .01, \eta^2 = .18$. All other questions showed no significant differences. Note that for this analysis only the conditions *Revision* and *Continuous* are present. An overview is provided in Figure 5.7, additionally showing slightly higher ratings for Q5a and average ratings for Q3a, indicating that participants believed that they had performed adequately even without the system. Q2a shows that users felt more pressured in the *Continuous* condition, while both conditions were rated as sufficiently accurate (Q1a).

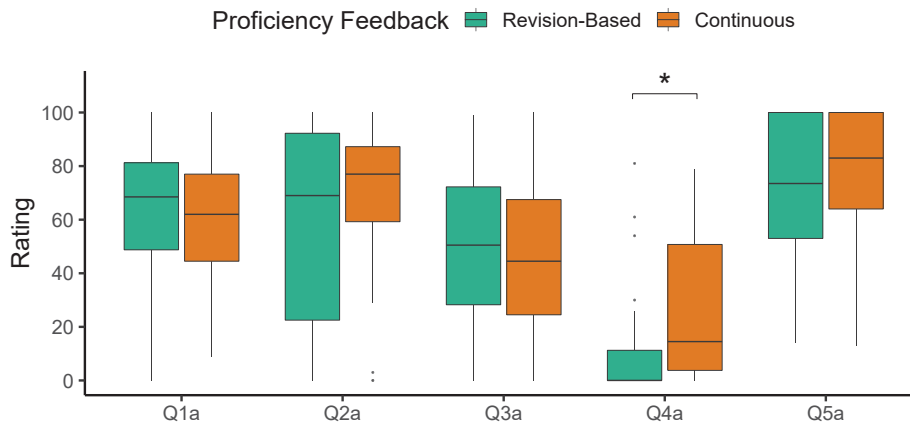


Figure 5.7: Ratings for questions Q1a-Q5a (see Table 5.6). Significant differences are marked with *.

Summary

Our initial evaluation of different proficiency feedback methods has raised some noteworthy facets of proficiency-aware systems. The following summary discusses these first results and highlights aspects that informed the design of Study II.

When users were confronted with a system that continuously assesses their proficiency, it allowed them to reach a target with higher precision, partly confirming **H1a**. This effect was not only present for the final readability score, but also allowed them to narrow down on the target zone more quickly compared to other assessment types (**H3a**). This advantage in performance did not impair task completion times, as there was no significant difference over the different assessment types. Interestingly though, participants in the *Revision* condition tended to narrow down on the target zone as quickly without having had access to their readability score yet. Thus, it remains to be investigated why users in the *No Feedback* condition converged less quickly and were less accurate. One possible effect could be increased extrinsic motivation by participants in *Revision*. Here, participants were informed that they were being assessed by a scoring system. Conversely, participants in *No Feedback* were unaware of this fact and were not put under pressure. Once participants started to revise in *Revision*, a definitive effort by participants could be observed to close in on the target zone, confirming **H3a**.

While participants reported no differences in perceived workload, it was evident from our custom questions that the *Continuous* feedback method caused significantly more disruptions (cf. Q1a) than the *Revision*-based method, confirming **H2a**, though both conditions pressured users to a fair degree (cf. Q2a). The *Revision*-based feedback seemed to cause split reactions from participants in this regard.

Ultimately, *Continuous* feedback offered the best performance in terms of hitting a target readability score at the cost of pressuring users into the need to perform (**H1a**). Here, a revision-based system can produce relief if precision is not essential (**H2a**).

Recognizing that proficiency feedback increases perceived stress and interruption, we aimed to further explore feedback modalities that might remedy those issues. In Study II, we investigated whether gamification elements could make an impact through enjoyable game design components.

5.2.4 Study II - Designing Better Proficiency Feedback Through Gamification

Based on the findings of Study I, we further investigated effects and perceptions around different forms of continuous and revision-based proficiency assessment and feedback. In Study II, we explored feedback modalities in which we expected to mitigate perceived pressure and interruption reported by participants in Study I. In particular, we designed and evaluated contrasting gamification components for both continuous and revision feedback.

Knowledge of gamification mechanisms and theories is key in designing systems that use gamification elements effectively. Often, gamification designs focus only on most common game design elements, including badges, points, and leaderboards [107, 282]. However, a much wider set of game elements has been studied and systematically assessed. Tondello et al. [309] described 49 gameful design elements, classified into eight groups according to users' self-reported preferences. Their systematic characterization emphasizes the value of choosing game elements that are suitable to tasks and users, in order to create sustained engagement.

Given the potential of gamification in creating engagement through different types of feedback, and the design constraints, we based our additional designs and conditions in Study II on an extensive reflection of gamification mechanisms. The taxonomy of gamification elements, described by Robinson and Bellotti [249], provided a good reference. They presented 42 gamification elements coded according to their minimum level (i.e. low, medium, high, and variable) of engagement required. The authors further categorized those elements into six top level categories. We chose to focus on two categories. First, "*Feedback and Status Information*" provided a highly relevant overview of game elements that directly support the purpose of our study. To contrast those rather rational and subtle game design elements, we further chose to explore the design elements which support "*Intrinsic Motivation*", as we wanted to understand the impact of joyful and interesting gamification elements on users' perceptions of additional feedback information. In the end, we implemented two gamification views, shown in Figure 5.8: *Emoji* and *Progress Bar*. Together with a *No Gamification* condition as baseline, this resulted in a two-factorial design including the factors *Feedback Type (Revision, Continuous)* and *Gamification Type*, yielding a total of six conditions (cf. Section 5.2.2). We note that those feedback modalities do not represent a full systematic

exploration of applicable gameful components. Rather, they are designed to contribute an early exploration of how different metaphors impact performance.

Emojis

The left side of Figure 5.8 shows the *Emoji* implementation that indicates compliance with the text production task through five unicode emojis, ranging from sad to happy. This design and implementation is reflected in the *Entertainment* element ("Simple elements can work very quickly in low commitment settings"), as described in the *Intrinsic Motivation* group by Robinson and Bellotti [249].

Progress Bar

The second gamification implementation, shown on right side of Figure 5.8, relates to the element *Graphical Indicators* ("Easy to design, and in fact critical for all gamification."), as described by Robinson and Bellotti in the *Feedback and Status Information* group [249]. We chose to implement a progress bar with a center target range (green zone), as it represents an easy-to-understand visualization of the tasks and provides a form of *progress feedback*, which Tondello et al. [309] referred to as a gameful design element. Having a sense of progress is important in stimulating the basic psychological need *competence*. In addition, we added a statement that places the users' current proposal quality into context with proposals written "in similar tasks". This statement is simulated and based on the current readability score. For example, at a readability score of 49, the statement indicates that "**26%** of texts in similar tasks fulfil the requirements better because **your text is too difficult to read.**", while at a score of 63, the message is more positive: "**Only 5%** of texts in similar tasks fulfil the requirements better". The purpose of this ranking, or text-based leaderboard, is to create a form of perceived competition, also referred to as gameful design element "social comparison" by Tondello et al. [309], nudging participants into better performance, which should work even in those situations where they already hit the inner edges of the green target range on the progress bar. For consistency and simplicity, we refer to this gamified view as *Progress Bar* view.

Hypotheses

We reused our hypotheses from Study I and explicitly describe our hypothesis regarding the use of gamification elements in proficiency feedback (**H4**).

H1b: Continuous feedback leads to increased task performance.

H2b: Revision-based feedback is less distracting.

H3b: Proficiency feedback alters the task approach of users.

H4: Gamification elements alleviate feedback distraction.

Your proposal

Your word count: 42

Your current readability score is: **63.7**

Dear Sir or Madam,

my name is John Smith and I have been invited to submit a proposal to make our city more green. It has long been a concern of mine that our lovely town has been pestered by industry pollution.

Your text meets the requirements.

**Your proposal**

Your word count: 42

Your current readability score is: **63.7**

Dear Sir or Madam,

my name is John Smith and I have been invited to submit a proposal to make our city more green. It has long been a concern of mine that our lovely town has been pestered by industry pollution.]

Too difficult

Too simple



Only 5% of texts in similar tasks fulfill the requirements better

Figure 5.8: Gamification implementation of our proficiency-aware web form. The view on top shows the emoji condition. The bottom view shows the progress bar and social comparison statement. Both gamification views depict the continuous condition.

We know from Study I that proficiency feedback is perceived as pressuring and distracting. We hypothesize that the joyful nature of gamified applications can turn proficiency feedback into an enjoyable experience that is perceived as an enrichment, rather than a distraction. To this end, we implemented two gamified proficiency-aware web views for both the *Revision*-based and *Continuous* feedback conditions, added custom questions to the evaluation, and used several subscales of the Situational Motivation Scale (SIMS) [99] and the Intrinsic Motivation Inventory (IMI) [39], to assess impact on intrinsic motivation and different forms of extrinsic regulation.

Participants

Again, we recruited participants ($N = 147$)⁸ over the Amazon Mechanical Turk Service (MTurk). Reimbursement and information provided to prospective participants was identical to Study I. Out of these participants, 40 resided in the European Economic Area, 42 in Canada, 3 in Australia, 1 in Asia and 61 in the USA. It took approximately 20 minutes to complete the survey, including writing the proposal text. The average age of participants was

⁸ Final number of participants, after rejecting obvious outliers in terms of proposal quality and questionnaire responses.

$\bar{x} = 35.9$ ($s = 9.85$) with 46% identifying as female, 54% as male. Additionally, we asked all participants to provide their self-assessed writing skills [46] again. We decided to omit the self-assessed writing quality of their proposal in Study II, to streamline the procedure.

Similarly to Study I, we found that participants rated themselves highly proficient in this writing task: $\bar{x} = 13.5$ ($s = 1.6$) for the writing self assessment⁹, again confirming that all participants were sufficiently proficient to execute the given writing task. Additionally, the experimenter evaluated each proposal for correctness and adequacy.

Procedure

Following the procedure of Study I, participants were asked to provide their demographics and to assess their writing skills after providing informed consent. They were then randomly assigned to one of the six conditions and were again given the same writing task as in Study I (cf. Section 5.2.3) and likewise informed about their condition.

After completing the writing task, participants completed the SIMS [99] and IMI scales [39]. We included all subscales of SIMS: Intrinsic motivation, Identified regulation, External regulation and Amotivation. For IMI we included three subscales¹⁰: Perceived competence, Effort/Importance and Pressure/Tension. Each subscale was scored on a 7-item Likert scale and included four to six items that were averaged. Note that we omitted the NASA-TLX in Study II as it did not elicit any significant results in Study I. We again concluded with a set of custom questions. Although similar to those in the first study, we focused more on the aspects of interruptions and distractions by the scoring system. We also explicitly tailored to a change in task approach. Table 5.8 shows the final set of questions.

Table 5.8: Additional questions for all conditions; from *strongly disagree* to *strongly agree*; all visual analog scale (0 to 100).

Perception of the scoring system	
Q1b	I felt pressured by the system.
Q2b	I felt that my performance was accurately assessed by the system.
Q3b	I adapted my approach in solving the task due to the system.
Q4b	The system enabled me to complete the task accurately.
Q5b	The system interrupted me during the task.
Q6b	The system helped me to see how well I was doing.
Q7b	The system distracted me during the task.

Results

We report our analysis on the following metrics as collected in Study II: task completion time (TCT), the final Flesch reading-ease score (FRES) and scores from the IMI and SIMS

⁹ Maximum score: 15.

¹⁰The other subscale were redundant with SIMS or not of interest for this study.

scales. Additionally, we again take a look at the FRES over time for each condition and analyze in particular whether the *Revision*-based methods has prompted participants to alter their proposal. Lastly, our custom questions give insights into the perception of the scoring system with regard to each feedback method. Inferential statistics are conducted analogously to Study I. The distribution of data entries over conditions is given in Table 5.9.

Table 5.9: Distribution of collected data entries over all conditions in Study II.

	<i>No Gamification</i>	<i>Progress Bar</i>	<i>Emoji</i>
<i>Revision</i>	25	20	22
<i>Continuous</i>	23	31	26

Task Completion Time and Final Readability Score

We did not observe any significant difference between the conditions for task completion time (TCT) in Study II. TCT was highest for *Continuous x No Gamification* at $\bar{x} = 1280 s$ ($s = 787 s$) and lowest for *Revision x No Gamification* at $\bar{x} = 810 s$ ($s = 496 s$). For the final readability score, we found significant differences with a large effect size for the factor *Feedback Type* ($F(1, 141) = 30.0, p < .001, \eta^2 = .17$). Figure 5.9 shows the final scores for the two factors.

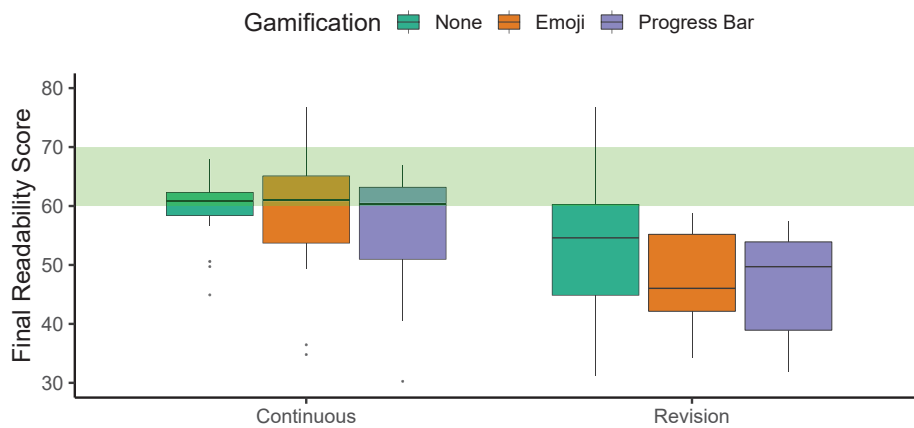


Figure 5.9: Final readability scores given both factors: *Feedback Type* and *Gamification Type*. Scores for *Feedback Type* (*Revision* - *Continuous*) are significantly different. The shaded green area marks the target zone for the readability score (60 to 70).

IMI and SIMS Scales

For IMI [39] and SIMS [99] scales we found significant differences (medium effect) for intrinsic motivation ($F(2, 141) = 5.36, p < .01, \eta^2 = .07$) for *Gamification Type*. Post-hoc pairwise comparison showed a significantly lower score for intrinsic motivation for *Emoji* compared to *No Gamification* and *Progress Bar*. Further, we found significant differences

(small effect) for identified regulation ($F(2, 141) = 4.05, p < .05, \eta^2 = .05$) for *Gamification Type*. Again, post-hoc tests showed a significantly lower score for identified regulation for *Emoji* compared to *No Gamification*. We did not find any significant differences for external regulation and amotivation as well as for any of the IMI subscales (Perceived competence, Effort/Importance, Pressure/Tension) that we employed. An overview categorized by *Gamification Type* can be seen in Figure 5.10.

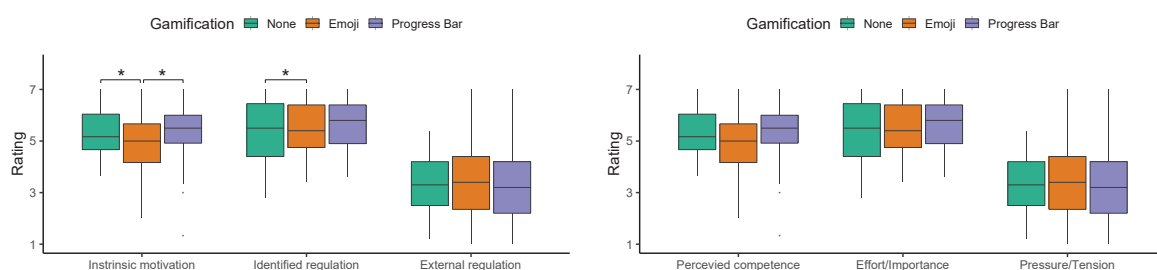


Figure 5.10: SIMS (left) and IMI (right) subscale ratings categorized by *Gamification Type*. Significant differences are marked with *.

Influence on Readability Score

Similar to Study I, we additionally analyzed the readability scores over time. We applied the same data transformation (cf. Section 5.2.3), but adapted the median task completion time to 927 s. An overview of this analysis is shown in Figure 5.11. It illustrates again that participants in the *Continuous* conditions narrowed down faster on the target zone and also exhibited less variance during this process and at the end (cf. Figure 5.9). Compared to Study I, the difference to *Revision*-based conditions is more pronounced. Participants in the *Progress Bar* conditions also narrowed down on the target zone in a more linear fashion. The other conditions exhibited a more ad-hoc adaption.

We could additionally confirm the same effect for *Revision*-based feedback conditions as in Study I. Participants did again try to improve their readability score after the first revision. We performed the same statistical analysis for each revision-based condition and could confirm statistical significance. Further, we fitted a model with aggregated data over all *Revision*-based conditions. Tests against the null model confirmed a significant difference with a large effect size ($F(1, 74.7) = 8298.9, p < .001, \eta^2 = .97$).

Additional Questions

Our analysis revealed significant differences for *Feedback Type* for Q3b: "I adapted my approach in solving the task due to the system." ($F(1, 141) = 5.24, p < .05, \eta^2 = .04$), Q5b: "The system interrupted me during the task." ($F(1, 141) = 13.38, p < .001, \eta^2 = .09$), Q6b: "The system helped me to see how well I was doing." ($F(1, 141) = 10.74, p < .01, \eta^2 = .07$), Q7b: "The system distracted me during the task." ($F(1, 141) = 26.75, p < .001, \eta^2 = .16$) with varying effect sizes. For Q7b we additionally found a significant effect of *Gamification Type*,

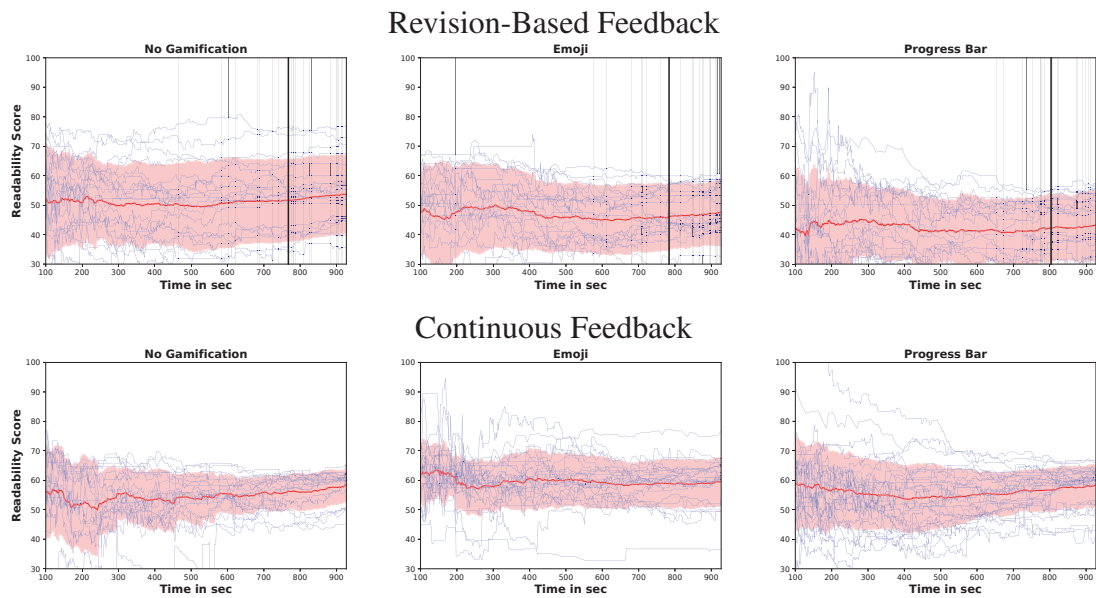


Figure 5.11: Deviation of readability score over time. Red-shaded corridor shows standard deviation over the filtered data (red line). Conditions based on revision feedback (top row) additionally show timings for the revision prompt (black vertical lines; thick line being the mean). Note that all trials have been rescaled to the median answer time and start at $t = 100$ s.

though no post-hoc significances between levels were present. All other questions showed no significant differences. An overview is provided in Figure 5.12 showing high rating (more prominent for *Continuous*) for Q3b, Q4b and Q6b polling the interaction between the participants' task approach and the assistance offered by the system. Scores on Q1b, Q5b and Q7b (pressure and interruptiveness) are more split between the conditions, while Q2b scores just above the midpoint range, polling accuracy of the scoring system.

Summary

Extending our summary from Study I, we further elaborate on additional findings based on the results of Study II.

Similarly to the results of Study I, we observed that *Continuous* feedback allowed participants to reach the target zone with higher precision. The bigger sample size of Study II confirmed that this effect was significant compared to *Revision*-based methods, fully confirming **H1b**. Additionally, Study II provides evidence that *Continuous* methods allowed users to narrow down on the target zone more quickly (**H3b**), showcasing that users adapted their task solving approach in the presence of the scoring system. For *Revision*-based method, we once again confirmed that once participants started their revision, a definitive effort could be observed to close in on the target zone, confirming **H3b**.

Our updated custom questions allowed us to take a closer look at the disruptiveness of the system and how users perceived its assistance. Here, we could confirm that *Continuous*

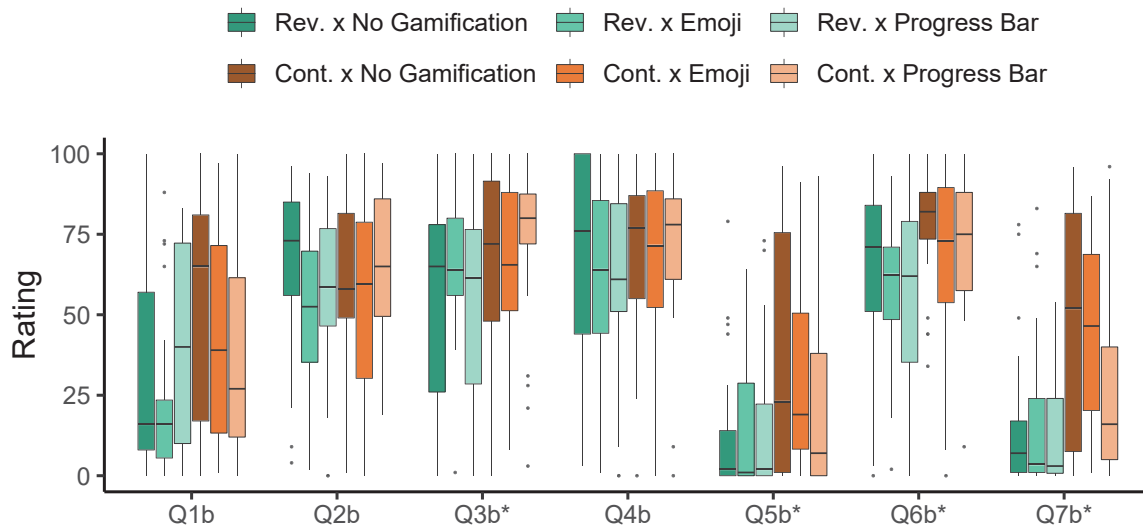


Figure 5.12: Ratings for questions Q1b-Q7b (see Table 5.6). Q3b, Q5b, Q6b and Q7b are significant for *Feedback Type*. Q7b is additionally significant for *Gamification Type*; marked with *.

methods were significantly more distracting and interrupted the user (**H2b**). While both feedback method prompted users to adapt their task solving approach, the impact from the *Revision*-based method was significantly lower (**H3b**).

Our results confirmed, in the *Continuous* feedback conditions, that gamification can alleviate feedback distraction (**H4**). While the participants felt strongly distracted in the *Continuous x No Gamification* condition, they reported less distraction in the *Continuous x Emoji* condition, and even less in *Continuous x Progress Bar*. This shows a valuable benefit of gamified feedback modalities. However, our results related to SIMS and IMI subscale responses indicated that the choice of game design elements significantly impacted users' motivation. Asked about their intrinsic motivation, participants indicated a significantly lower intrinsic motivation for the *Emoji* views, as compared to the *Progress Bar* views and to *No Gamification*. Reasons might include that users felt personally attacked by sad or disappointed emojis and might have seen little interest or value in feedback that was based on metaphors that they are commonly subjected to in social interactions. An additional indication of this was the significant difference in identified regulation between *No Gamification* and the *Emoji* views.

5.2.5 Discussion

Here we reflect on the results of our inquiry and discuss how different forms of feedback affect performance and user experience when interacting with proficiency-aware systems.

Proficiency Feedback Provides Tangible Benefits in Performance

Our investigation showed that users who could use the proficiency-aware features of the web forms were able to fulfill the secondary task requirements (write in plain English) to a higher degree than those who did not receive any feedback. This suggests that proficiency awareness is an effective method in helping users achieve a desired task outcome (**RQ3b**). Interestingly, *Revision*-based methods were almost as accurate as *Continuous* feedback methods in Study I. We hypothesize that this can be attributed to the participants' awareness that they would be rated by a scoring system [66] and the lower sample size. Study II confirmed that *Continuous* was superior in terms of the quality of the generated text (**RQ3c**). Hence, **choosing feedback types is a trade-off between primary task performance and awareness of proficiency** [82]. Further, providing proficiency feedback has not significantly slowed down primary task completion, nor has it increased the perceived workload of users. These results show that **proficiency feedback can lead to tangible benefits in performance quality without negatively affecting effort or speed** (**RQ3c**).

Revision-Based Proficiency Feedback Exhibits Low Interruption Cost

While both feedback methods were able to support users in their task, *Revision*-based methods were perceived by the users as less disruptive. For *Continuous* methods, users reported that they felt interrupted and distracted by the scoring system to a higher degree. Thus, it is important for future designers to recognize the possible negative implications of providing proficiency feedback. **Providing feedback at sparser intervals is likely to reduce disruption** (**RQ3c**). Consequently, when designing a proficiency-aware system, the designer is faced with the challenge of finding the optimal balance for a given task. Frequent or continuous feedback will increase quality, but may frustrate users [133].

Proficiency Feedback Facilitates an Understanding of One's Own Proficiency

Our results suggest that users not only managed to perform better when using proficiency feedback, but they also gained an understanding of their proficiency. We observed that users changed their behavior in reaction to being informed about their proficiency [25]. Additionally, users stated that they had consciously adapted their task solving strategy and that the system helped them in doing so (**RQ3b**), a clear indication of reflection-in-action [278]. Users were aware of their own approach to complete the task and could recognize opportunities to improve their strategy. This fact presents a design opportunity for future proficiency-aware systems, counteracting the "paradox of the active user" [36]. If proficiency can be effectively recognized, **systems can guide users to selecting task completion strategies which are optimized for a given user's proficiency level**.

Proficiency-Related Feedback Promotes Consistency

Further, we observed that proficiency feedback led to less variability in user performance. Participants who used proficiency feedback reached an optimal performance level faster and were able to maintain it with a greater consistency. This effect is more prominent for the

Continuous feedback methods. Users reported that the scoring system helped them to accurately complete the given task. This is especially evident in Study I, where the *No Feedback* group exhibited a large variance for their final readability score. Consequently, we can conclude that **continuous assessment can help to promote and realize consistency among similar tasks (RQ3b)**. Future systems can use this fact to effectively communicate to the user what level of performance is expected and helping the user perform at the desired levels. However, the task in our study was one where the user was asked to perform at a certain, mid-range level. In tasks where exceptional performance would be beneficial, proficiency feedback may guide users to do just enough to obtain a positive assessment [115] and thus potentially hamper excellence.

Gamified Feedback Can Be Effective, but It Is Subject to Design Constraints

Our investigation of gamified feedback elements showcased essential considerations for proficiency-aware systems. While the gamified feedback was not more beneficial in terms of conveying proficiency to users, both gamification elements helped reduce the perceived interruption of the scoring system (**RQ3d**). Compared to *Emoji*, *Progress Bar* provided a less distracting experience, which was almost on the level of *Revision*-based methods. We attribute this to the much more ambient appearance of the *Progress Bar*. The additional information regarding distance to the target zone and overall progress provided a tangible way to keep track of one's progress towards the target zone. This shows that gameful elements can help mitigate the disruptive effect of continuous proficiency assessment. Consequently, **future systems that want to avoid disruptiveness without jeopardizing performance, can explore gamified feedback instead of reducing feedback frequency**. At the same time, meaningful social comparison elements that appeal to the users are likely to stimulate extra effort that goes beyond merely reaching the target zone. A transparent communication of such information across co-workers and peers is likely to impact one's feeling of relatedness; one of the basic psychological needs [285].

Further, using the *Emoji* resulted in a significantly lower intrinsic motivation than the *Progress Bar* and the *No Gamification* condition. In conjunction with the lower score for identified regulation as well as perceived competence, we hypothesize that participants associated the *Emoji* with a form of childish, frivolous feedback. This was an impression that may have contrasted with the serious writing task [286]. Consequently, **if gamified elements are to be used for feedback in proficiency-aware systems, it is a key design consideration to align the feedback form with the content of the task at hand (RQ3d)**. Still, it would be worth to further explore gameful design elements for proficiency feedback that increase intrinsic motivation through interest, curiosity, and play. In this context, it should be noted that recent gamification research focused on mapping the suitability of game design elements to player types [223, 308]. Future work should further explore the impact of player types and the selection of game elements on motivation in proficiency feedback.

Limitations

While proficiency is generally difficult to measure accurately, contributing factors can usually be obtained with ease. Thus, we used a text production task and an associated, established task performance metric (FRES) to illustrate the concept of proficiency-aware systems. Our current work shows that communicating relevant task metrics in a text production task facilitates proficiency awareness. It remains to be seen whether these findings can be replicated for similar tasks. A complete proficiency-aware system should additionally consider more elaborate metrics, such as proposal quality, structure and tone, consequently supporting the user in achieving their task. Work by Kaur et al. [154] can serve as a glimpse into the future where proficiency-aware systems may create action plans for the user. While the scope of the current work addresses the influence of proficiency feedback on user experience and their performance, we firmly believe that providing this framework for the concept of proficiency-aware systems helps identify necessary requirements and constraints to advance further research.

We also recognize that our participant recruitment options were limited due to the epidemic situation at the time of writing. While MTurk allowed us to recruit a large sample of users, it also limited our possibilities in obtaining more information about their experience, e.g. through qualitative post-task assessment. To ensure data quality, we carefully screened submitted results. Further, the dynamics of motivation in crowdwork is a debated topic, e.g. [28, 153]. While our results on motivation presented here may be specific to an MTurk setting, we do note that the scope of this work is not limited to an analysis of motivational aspects only.

5.2.6 Conclusion

In this use case, our investigation focused on different feedback types to communicate proficiency assessments to the user in everyday computer work. We found that being aware of one's proficiency benefits primary task performance and facilitates an understanding of their own expertise for the user. Consequently, users can adjust their approach towards the task and increase their proficiency.

While communicating proficiency is beneficial for task performance and user experience, we also found that feedback needs to be balanced to moderate potential interference with the primary task. We suggest a revision-based approach if precise performance is not essential, as it is less distracting for the user. Similarly, subtle gamification elements like progress bars and associated social comparison complement this method by lowering perceived disruption. We summarized these findings in a set of design implications applicable to proficiency-aware applications, such as choice of feedback type and inclusion of gamification elements in everyday computer-based work environments.

This use case contributes to the understanding of how people reflect on their own skill assessments and, more importantly, how this reflection can be used to encourage users to improve

their proficiency further. In the next use case, we pursue a slightly different approach by introducing the idea of artificial proficiency. We use an EMG-based system to allow piano player to add expressiveness to their play without having to go through lengthy training. Thus, users can get a sneak peek at possible benefits of proficiency improvements.

5.3 Use Case III: EMPiano - A Take on Artificial Proficiency

When interacting with artifacts, we are often limited by the given range of possible interactions provided by the respective interface. Well-designed interfaces clearly communicate desired interactions and rely on expected movements. An example of such an interface is the piano keyboard, which is a well-established interface between the musician and the instrument. However, mastering the piano keyboard requires extensive skill development. Beginner and intermediate players are often limited to basic melodies and there is little room for improvisation. Accessing this aspect of musical play can be challenging as it is usually associated with increased costs, such as higher user workload.

In this use case, we envision the use of artificial proficiency to demonstrate benefits of proficiency increase to user in a tangible way. To so do, we look at expressive capabilities of interaction modalities, such as EMG, that subtly extend the range of interactions, hence allowing for seamless integration. The proposed system EMPiano integrates seamlessly into normal play by allowing musicians to modulate sound pitch using their thumb, supporting musical flow. The system's capabilities allowed users to foster their play creativity without having to worry about limitations of their own play proficiency.

5.3.1 Methodology

Our goal is to leverage electromyography as a secondary input channel, thus increasing the interaction range when playing the piano, providing users with a sense of artificial proficiency. To verify the applicability of EMG in this scenario, we created a prototype that allows the pianist to modulate pitch during play by measuring their muscle activity. We implemented this experience by allowing the user to alter the pitch via a vibrato effect using a thumb wiggle gesture. Based on **RQ3**, as reiterated below, we formulate the following research question **RQ3e** for this use case:

RQ3

What are design criteria for proficiency awareness in interactive systems?

RQ3e: How can we design for artificial proficiency which integrates seamlessly into existing tasks for EMG-based systems?

We ask whether such a gesture can integrate seamlessly into the playing process. It has thus to be within the same domain of muscle movements, but have little overlap with the actual playing motion to prevent false positives. In other words: we wanted to design a gesture that the user would expect, but also one that can be sensed [21]. McPherson et al. [204] investigated possible modulation gestures during piano play that had no to little overlap with existing techniques during play. The so-called space between the note [204] is of vital importance as it can be used to create new musical artifacts.

We research the feasibility of "task language", i.e. finger motions that define the intended interaction¹¹ — in contrast to the position of the fingers on the keys — to allow for player-specific interaction. In other words, we focus on applying EMG to sense the aspects of an object that allow for providing contextual clues for extended interaction. This paradigm has been researched in other areas as well, such as PickRing [329] and work by Theiss et al. [306]. Analogously to the work by McPherson et al. [204], we extend on-the-move interaction through EMG, but do not limit the concept to specific objects, that need to be modified.

5.3.2 EMPiano

In EMPiano, we implemented a thumb wiggle gesture as effect trigger: the user applies force on the piano key while simultaneously tilting his thumb back and forth. This motion then triggers a vibrato-effect (on or off) during play. Tilting direction and applied pressure are determined in a calibration phase and depend on individual preference and playing style. We allowed a variety of slightly different motions across our participants, as preliminary tests found no negative impact on detection accuracy. These tests also showed that using the thumb was most accurate and aligned well with occasions when one might want to add a vibrato effect¹².

EMPiano closely follows the recording steps of EMGuitar. For detailed information on recording protocols and data preprocessing steps, please refer to Section 4.2. Specifically for EMPiano, we set the window size to $150ms$ ¹³. Hop size was set to half the window size. Additionally, we submit RMS (including RMS ratio) features for the last and second-to-last window. Thus, for classifying the current state of the system, we look at a total of $300ms$ worth of EMG samples. Note that classification is performed every $75ms$ (hop size) using a support vector machine with a radial basis function kernel¹⁴. We found that this method ensured stable classification while also maintaining decent latency. No participant reported any perceived activation latency when asked about it in the interviews.

¹¹E.g. playing an octave requires stretching of fingers.

¹²For example: ending a series of notes with a low key.

¹³Corresponds to 75 samples.

¹⁴ $C = 1.0$, $\gamma = scale$ as evaluated a priori. Also see scikit-learn (<https://scikit-learn.org/stable/index.html>).

A computer was connected to an electronic piano keyboard via MIDI providing piano sound output via Ableton Live¹⁵ through external speakers in all conditions. Additionally, this computer connected to the available EMG stream for classification purposes. In our *Electromyography* condition, sound throughput was altered using MIDI signals to emulate a vibrato effect. The audio processing pipeline exhibits a latency of only a few milliseconds.

5.3.3 Evaluation

We evaluated EMPiano in an experiment consisting of two different tasks. The first task *Notesheet* required the user to play a simple piece from notes and modulate at priorly defined positions, introducing the participants to our system and its operating principles. The second task *Improvisation* then allowed the participants to fully explore the prototype and incorporate its features into their own playing style. They were free to improvise at will or play pieces from memory or from notes.

To allow for comparison, we introduced an auditory baseline that offered the same expressive range albeit using a different modality. Hence, we employed a within-subject design with one independent variable *Modality* for each task. The two conditions differed in their way of triggering the modulation. The *Pitch Wheel (PW)* condition required the participants to modulate pitch via the pitch wheel, manually creating a vibrato effect. The pitch wheel represents a standard control mechanism. The second condition, *Electromyography (EMG)*, relied on the calibrated trigger gesture to activate an automatic vibrato effect. Conditions were counter-balanced within each task.

We constructed a set of questions about the perceived accuracy and intuitiveness of the activation movements and the respective sound modulation as shown in Table 5.10. All answers were recorded on a visual analog scale ranging from 0 to 100. This questionnaire was administered for both conditions and both tasks. In a concluding interview, we asked the participants about their experience when using the two different conditions for sound modulation. Specifically, we asked them to contrast the feeling of integration for both conditions and how either had impacted their play style when improvising. Each interview took approximately 10 minutes.

Participants

We recruited twelve participants (9 male, 3 female; Age $\bar{x} = 23.2y$, $s = 3.0y$) from the University of Munich via mailing lists. Although subjective play proficiency varied ($\bar{x} = 32.0$, $s = 19.6y$)¹⁶, all participants were able to play a piece from the written notes and reported an average of $\bar{x} = 11.2h/month$ ($s = 11.2h/month$) spent playing the piano. After the study, each participant was reimbursed with 10 euros.

¹⁵<https://www.ableton.com/en/>

¹⁶Visual analog scale from 0 to 100.

 Questions assessing perceived accuracy and intuitiveness

- A1. Please rate your perceived accuracy of the system in recognizing your augmentation.
 - A2. How intuitive was the gesture movement?
 - A3. How intuitive was the sound modulation?
 - A4. How intuitive was the combination between the gesture movement and the sound modulation?
-

Table 5.10: Additional questions for both conditions (*PW*, *EMG*) and both tasks (*Notesheet*, *Improvisation*); all visual analog scale (0 to 100).

Procedure

After providing informed consent, the participants completed a demographics questionnaire and the experimenter placed the electrodes on the right forearm. The first task *Notesheet* required participants to play a piece from a sheet of music. The songs included marked notes throughout the piece where the player was required to use a vibrato effect. As required by the *EMG* condition, calibration for the trigger gestures was done directly while playing the chosen song twice in different registers, lasting no more than two minutes. This calibration was always executed first. Depending on condition order, the vibrato effect had to be applied using the pitch wheel or the *EMG*-based trigger gesture. Each condition was completed by filling our custom questionnaire (Questions A1-A4).

Task *Improvisation* allowed the participant to experiment with both conditions during a free improvisation session. Participants were given approximately five minutes per condition. The semi-structured interview concluded the experiment. The whole experiment, including electrode placement, did not exceed one hour. Ethical approval for this study was obtained from the Ethics Committee at the University of Constance.

5.3.4 Results

We report on the quantitative analysis of our custom questionnaire and the qualitative analysis of the interviews.

Custom questionnaire

Results on gesture movement and sound modulation are illustrated in Figure 5.13 on a visual analog scale (0 to 100). Please refer to Table 5.10 for the respective question identifier. After confirming normality of the data, a paired sample t-test showed a significant effect of *Modality* for Question 3 ("How intuitive was the sound modulation?") and Question 4 ("How intuitive was the combination between the gesture movement and the sound modulation?"): $t(8) = 3.00, p < .05, r = .72$ and $t(8) = 3.01, p < .05, r = .73$, respectively.

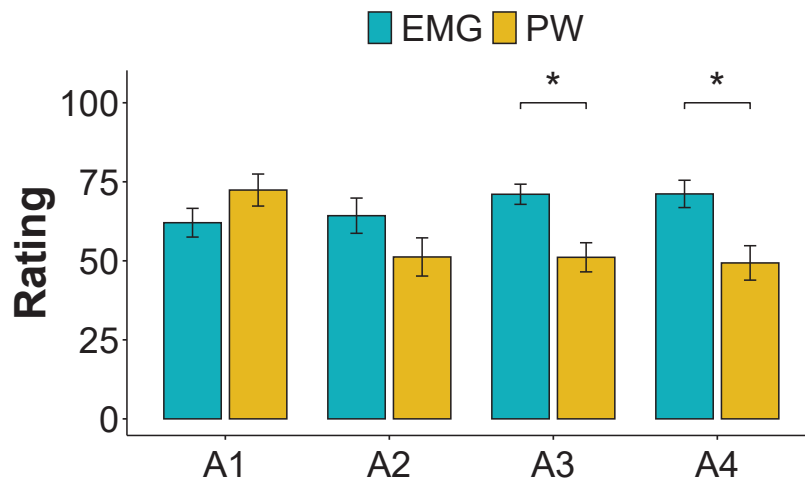


Figure 5.13: Mean scores for Question A1-A4; assessing perceived accuracy and intuitiveness of both conditions (see Table 5.10). Error bars show standard error. Significant differences between conditions are marked with *.

Interviews

All interviews were recorded (total duration 1 : 41 h) and transcribed verbatim. We opted for the pragmatic approach to thematic analysis as described by Blandford et al. [24]. Our analysis resulted in four themes: COMFORT, CONTROL, CREATIVITY and IMMERSION. In the context of this use case, we will only focus on CONTROL and IMMERSION.

Control

Another important aspect for participants was the perceived control over the modulation. For most participants using the pitch wheel, modulation felt cumbersome and it was difficult to achieve a desirable vibrato effect:

On the other hand, it [pitch wheel] is also difficult. Especially that light vibrato is not that easy to achieve. Reproducing this, I experienced it as much more taxing than using the gesture [EMG]. (P8)

While using our system enabled players to stay in control, they also reported on the lack of fine-grained control when needed:

Using EMG, what I really like is that I could be focused on what I was doing, and because this [referring to the pitch wheel] will be very far from when I'm hitting the piano. I could focus on what I was playing. I could just do some kind of thing on the vibrato like you said, but I wasn't able to precisely control the modulation. (P2)

Being aware of the nature of the EMG-based modulation, participants formulated new ideas on how to control the interaction with the piano keyboard using muscle activation:

For example, when I wiggle my ears, I can trigger an arpeggio. (P9)

Relating to the previous theme COMFORT, the participants associated being able to use more fingers for activating the modulation with increased control:

Yes, I think it would help me because there are a lot of tracks in which you may want to use modulation on a specific note, but on this note you're not supposed to use your thumb. Yes, I think it would be great to be able to do it with other fingers. (P2)

Immersion

The last theme resulting from our analysis is IMMERSION. Being immersed is one of the main drivers for getting into a flow state during play. Participants reported a limited and disturbed feeling of immersion when using the pitch wheel:

My left hand was always occupied with that thing [the pitch wheel]. I do not know of any pieces that would allow me to switch in between. (P6)

Using the EMG-based modulation supported many players in achieving a flow-like state and immersing themselves in their improvisation:

For sure. I had the feeling to get into a flow state very quickly, because the other method [pitch wheel] interrupted me motion-wise. That always pulled me out of the flow. I also had the feeling that I struggled to stay on tempo and that pulled me out. With the other system [EMG], I did not have this feeling. I experienced a more flow-like feeling. (P8)

Participants commented how flow could be achieved. Firstly, the EMG-based system allowed for an immediate and non-disruptive integration into the participants' play style:

It was simply very intuitive. If I wanted to use it, it just worked. (P11)

Secondly, due to the seamless integration, the EMG-based system enabled participants to play freely:

And I think, when I incorporated it into my play style, it enabled me to play freely. This method was much better than the other one [pitch wheel], because it was much more effort for me to reach. (P8)

5.3.5 Conclusion

With regard to **RQ3e**, we looked at design requirements that need to be fulfilled to achieve seamless integration of artificial proficiency for EMG-based systems. One important advantage compared to the pitch wheel modulation is the fact that the utilized gesture activation

was rated much higher by players in terms of intuitiveness. For them, the connection between the movement and the resulting sound modulation was straightforward and immediate to apply, facilitating seamless integration.

Further, the increased range is realized without added complexity through seamless integration allowing for good control using already established "task language". This is additionally supported by the increased feeling of immersion that participants reported for the EMG-based system. This was evident when looking at the pitch wheel condition, which was often criticized for its high complexity. Yet, more advanced players appreciated the greater control over the sound modulation when using the pitch wheel. Our observations in this experiment provide an extensive answer for **RQ3e**. Thus, design for EMG-based interaction should balance the amount of control offered to the user.

This use case aimed at validating whether EMG can support seamless integration into existing systems to extend the range of interaction, facilitating a sense of artificial proficiency. Our successful integration of EMG as a secondary modality into an already well-established motor task showed that EMG does have the potential to enable additional interactions with tangible artifacts without altering the existing artifact or introducing new ones.

5.4 Chapter Summary

Designing proficiency awareness is challenging if there is no agreed-upon definition of proficiency in the associated task domain (**RQ3**). Nevertheless, in this chapter, we have illustrated that there are always proxies available that can be used to support users and make them aware of their own proficiency.

In particular, this chapter has shown that users are able to make elaborate connections to their own data, be it physiological signals (cf. Section 5.1) or linguistic metrics (cf. Section 5.2). This highlights an important advantage in terms of scaling for proficiency-aware systems. Often it is sufficient to allow for *some* information to communicate back to the user (**RQ3**). There is no need for complicated algorithms that assist users in reaching maximum task performance. Merely making the "invisible" visible (cf. EMG) is often sufficient.

Further, this implicitly allows user to make their own decision — whether to invest resources or not — ensuring a sense of agency for users. Here, our second use case (cf. Section 5.2) has shown that continuous feedback can potentially rob users of this agency for computer-based work.

In our final use case, we additionally looked at an approach that would *tease* a higher proficiency to users. By foreshadowing what investments in proficiency improvement can offer to users, we highlighted how proficiency-aware systems could make use of artificial proficiency as an additional motivator for users (**RQ3**). Interestingly, this support could readily be available to users until they themselves a high enough proficiency level.

While this chapter has provided insights into how to design proficiency awareness in selected task domains, we argue that, especially for proficiency-aware systems that want to use physiological sensing, high entry barriers often hinder further research and development. We will address this issue in the next chapter complementing Part II: "Facilitating Proficiency Awareness" of this thesis with a toolkit approach to implementing proficiency-aware systems.

Chapter 6

A Toolkit Approach to Implementing Proficiency-Aware Systems

User interfaces which use physiological phenomena as input are an established topic in HCI. Given the amount of measurable biological signals [274] which we emit, there is still potential for new interfaces. Yet, the possibilities for non-experts to explore user input modalities for interactive systems are limited. Most often, expensive equipment and prior training is required to operate systems correctly. Hence, accessing physiological input technologies is often cumbersome and requires expert knowledge. Universally available hardware and toolkits are a first step in this direction, as they allow non-experts to familiarize themselves with the technology and explore its capabilities without having to commit to expensive hardware and training.

Especially in the domain of physiological computing, reducing the need for extensive expertise is essential. Fairclough [75] called for easy data acquisition and abstraction from technical details. As a consequence, such toolkits need to be mobile and effectively support the creative process. Last but not least, toolkits are to "empower new audience" [179]. In this chapter, we explicitly focus on how to include a wide range of professions as potential user groups to advance research for proficiency-aware systems. Consequently, our work is guided by an initial set of requirements motivated by Ledo et al. [179]'s definition of a toolkit:

"[Toolkits are] generative platforms designed to create new interactive artifacts, provide easy access to complex algorithms, enable fast prototyping of software and hardware interfaces, and/or enable creative exploration of design spaces." [179]

Consequently, this chapter focuses on how we can support a wide audience in creating proficiency-aware systems. It is essential for the acceptance of novel modalities, such as EMG, that not only trained professionals are able to exploit these, but that we also make them accessible for engineers, developers, designer, and laymen.

We first introduce EMBody, a data-centric toolkit for EMG-based prototyping and experimentation. While this thesis has investigated gaze as a modality as well, we identified the available support by manufacturers of eye trackers to be already on a very accessible level for newcomers¹. Thus, we decided to focus on creating a toolkit that allows a wide audience to experiment with EMG and prototype EMG-based interactive systems.

The research in this chapter is guided by **RQ4**:

RQ4

How can we support a wide audience in realizing proficiency-aware systems?

This chapter is based on the following publication.

Jakob Karolus, Francisco Kiss, Caroline Eckerth, Nicolas Viot, Felix Bachmann, Albrecht Schmidt, and Paweł W. Woźniak. 2021. EMBody: A Data-Centric Toolkit for EMG-Based Interface Prototyping and Experimentation. In *Proceedings of the ACM on Human-Computer Interaction*, 5(EICS). ACM, New York, NY, USA.

<https://doi.org/10.1145/3457142>

EMBody is openly available on github: <https://github.com/HCUM/embody>

6.1 Related Research

The increasing availability of toolkits allows users to rapidly prototype interaction ideas. Most commonly, toolkits help ease certain steps during this process [179], from ideation and interaction design to signal acquisition and processing to higher level output generation, e.g. by means of machine learning. In the following, we reflect on prominent toolkits within the HCI domain, their purposes and architecture and take a closer look at physiological computing toolkits and electromyographic sensing.

6.1.1 Toolkits

Toolkits lower the entrance barrier for specific stages during the creation process of applications and artifacts [179]. Specialized toolkits, such as Makers' Marks [268], Sauron [267], Pineal [178], ShapeMe [326] and RetroFab [244] support technical users in working with aesthetics and form factors. The aforementioned tools allow novice users to create 3D forms or enclosures, which is a process which usually involves extensive knowledge and iterations using 3D software. By abstracting from this process, e.g. through shortcuts like embedding smartwatches as computation unit [178] or automated processes that convert physical

¹ E.g., Tobii Eye Tracker 5: <https://gaming.tobii.com/product/eye-tracker-5/>.

changes to digital representations [326], these toolkits enable users to focus on designing applications and benefit from advances in sensor technology and 3D modeling.

Other toolkits specifically address these engineering challenges, such as making sense of data (EagleSense [332], SoD-Toolkit [284]) assisting less technically adept users to interpret sensor data and high-level input for their applications. Taking care of data synchronization and filtering is cumbersome and often requires expertise, especially as environments contain more and more sensors every day. A designer for a location-aware application is only interested in a person's exact location and orientation (within a room). How this information is calculated is secondary and not relevant for the application. Here, toolkits such as EagleSense [332] abstract from the technical complexity allowing fast prototyping.

Developing ubiquitous artifacts often involves devices that are interconnected, e.g. a sensor and an actuator. Cross-device communication can be cumbersome and is often abstracted with the help of protocols and toolkits. Examples include toolkits for web-based applications (XDStudio [214], Panelrama [333]), tangible artifacts (Calder [180], ToyVision [194], reacTIVision [142]) or most commonly: wearable devices (WatchConnect [124], Weave [43], Interactex [103], WDK [102]). These toolkits showcase the importance of cross-device compatibility and properly defined interfaces to allow for robust communication among the devices.

6.1.2 Pipeline Architectures

A common approach to toolkit architecture is the use of a pipeline-based structure. This provides a conceptual workflow for diverse user groups and clearly communicates the toolkit's application domain. Toolkits might only provide one — but integral — step of this pipeline, such as enabling laymen to work with electric muscle stimulation [233], paper electronics [243], fostering data engagement [125] or simply enabling rapid prototyping of electronics [322].

This concept of separating individual steps of the creation process allows for a data-centric view during development. Steps can be parallelized to increase efficiency and enable outsourcing of complex processing to domain experts. Not being endemic to the prototyping and toolkit domain, the concept of pipelining can be found in other areas as well, such as visualizations [89], fabrication [276], debugging [294] and media [290].

6.1.3 Physical Computing Toolkits

Past work contributed tools for interaction designers who were not familiar with the employed sensing technologies. These toolkits [266, 128, 98, 304, 295] enable designers to realize their ideas by abstracting underlying sensor complexities and reducing the need for extensive expertise, which is often the most significant barrier to entry in using physiological computing [75].

Other toolkits [95, 195, 196, 319, 102] focus on a more developer-centered approach, e.g. by providing access to low-level hardware using high-level programming languages. A key aspect that unites all these past efforts is the focus on an iterative, design-centered approach [112] to create interactive systems.

6.2 EMBody - A Data-Centric Toolkit for EMG-Based Interface Prototyping and Experimentation

At the end of the previous century, Paul Dourish envisioned how physically and socially enacted encounters with technology will transform our everyday lives [65]. Later, Dag Svanæs [300] proposed expanding the concept of embodiment to, *inter alia*, embedded perception, i.e., extending one's senses and awareness through technology. Embodied interaction involves movement—a process where users implicitly contract and relax their muscles to move their bodies in particular patterns. In this section, we investigate the means of easily allowing HCI researchers to understand our muscles in more detail in order to design for embodied interaction.

However, building EMG systems from scratch is difficult and requires expertise from numerous domains such as sensing technology, signal processing, machine learning and interaction design. Here, off-the-shelf electrical sensors, e.g. shields² and sensors (MyoWare³, BioVolt⁴) for microcontrollers, can only provide the first step of this process. Users need to take care of data processing, calibration and model training. Hence, a significant amount of time needs to be invested, adversely affecting work on designing for EMG-Based interfaces. This constitutes an obstacle for a broader application of EMG-based interaction. Consequently, there is a need for new EMG tools for HCI prototyping. An EMG toolkit for HCI should offer opportunities for interaction designers, researchers, and engineers to engage with EMG-based sensing on different levels of technical complexity.

Based on **RQ4**, we thus formulate the research question **RQ4a** that specifically addresses requirements for EMG-related toolkits:

RQ4

How can we support a wide audience in realizing proficiency-aware systems?

RQ4a: What are design requirements of a toolkit for EMG-based prototyping?

² <https://www.olimex.com/Products/Duino/Shields/SIELD-EKG-EMG/open-source-hardware>

³ <http://www.advancertechnologies.com/p/myoware.html>

⁴ https://infusionsystems.com/catalog/product_info.php/products_id/198

We draw our initial design goals from past toolkit research and take especially physiological sensing into account (Section 6.1). Based on those past works, we derive four requirements for EMG-based prototyping and experimentation. In this work, we iteratively refined, addressed and evaluated the requirements in two workshops (Section 6.2.7), through developing sample applications and conducting expert interviews (Section 6.2.8).

The following section describes the final set of requirements for EMG-based prototyping and experimentation. We detail the corresponding exploration-centered workflow of EMBody as well its software and hardware components, highlighting possible extension points for technically skilled users. All resources needed to build, use and modify EMBody are open-source and are available on github⁵. Finally, a closer look at the different workflows possible with EMBody is provided through two exemplary use case scenarios in Section 6.2.6.

6.2.1 Requirements

Throughout the development process of EMBody, we identified four main requirements for EMG-based prototyping and experimentation. We specifically address the needs of a wider range of professions, who could potentially use EMG, for our toolkit.

Mobility

Mobility is a key aspect to facilitate prototyping and exploration not only in constrained lab environment, but to allow for in-the-field exploration. EMBody offers a low-power and light-weight apparatus which is mobile and can be carried by the user. This property enables straightforward in-situ exploration of interaction scenarios ensuring high external validity. The need for highly mobile EMG-based interfaces was exemplified by past work in HCI, which advocated using EMG for interactions on the go [50].

Data Acquisition

Data Acquisition often produces technical difficulties. Off-the-shelf sensor products rarely provide an abstraction layer for this process. Most commonly data has to be directly read via the analog port of a microcontroller. To facilitate rapid prototyping, data acquisition needs to be reliable and possible from a variety of devices. In the final version of EMBody, data acquisition is moderated using the UDP protocol over an existing WiFi connection. This allows experts to use custom hardware to communicate with EMBody's software application, by adhering to the protocol (Section 6.2.3). The standard hardware already provides a walk-through for users to connect to existing WiFi connections for initial setup. As such, EMBody provides a clear interface between its hardware and software components (see Section 6.2.3 and Section 6.2.4) as well as the user's application. The software component connects to the UDP stream of the EMG hardware (or any off-the-shelf electrical potential sensor on the

⁵ <https://github.com/HCUM/embody>.

market), processes and interprets the EMG data and provides a high-level gesture stream via UDP for the user's application.

EMBody's hardware provides up to six sensing channels without further modifications. Off-the-shelf products often need to be extended, requiring additional electronics development⁶. For EMBody, selecting individual channels is handled by the software application.

Abstraction Layer

To tailor for novices in EMG sensing, the toolkit needs to protect this user group from unnecessary technical details including signal processing and data interpretation. For this purpose, EMBody realizes an exploration-centered workflow (Section 6.2.2) that guides the user through a total of four steps from an initial idea to a final working prototype. Apart from deciding on the electrode configuration (location and channels), which is supported through a manual⁷, the user does not need to have any expertise in signal processing and data interpretation.

EMBody features a predefined data processing pipeline including a set of filters and algorithms that take care of data interpretation, allowing the user to focus on exploration and interaction design. By providing abstraction were required but allowing for customization were needed, EMBody facilitates a design-centered approach [112] to create interactive systems. If desired, users of EMBody can prototype EMG-based applications in proficiency-aware systems without the need to ever know anything about EMG but that it measures muscle activity.

Modular Structure

Experts want to have a fine level of control over the data processing pipeline. Consequently, EMBody offers gradual levels of fidelity through its modular structure. Every part of the workflow (Section 6.2.2) can be customized and adjusted to the user's needs. For an elaborate experiment, researchers might want to adjust the data pipeline by interchanging the classification algorithm or calculating different features. The accompanying software offers convenient extension points for this purpose, exposing various stages throughout the processing pipeline.

EMBody employs this concept and defines an exploration-centered workflow for novice and amateur users when designing for EMG-based interaction. Its modular data processing pipeline allows the expert user to customize EMBody to a high degree while maintaining a low entry threshold for novices.

⁶ E.g. stacking multiple shields for the Olimex board or connecting multiple MyoWare sensors.

⁷ Available at <https://github.com/HCUM/embody/tree/master/manual>.



Figure 6.1: Workflow when designing with EMBody.

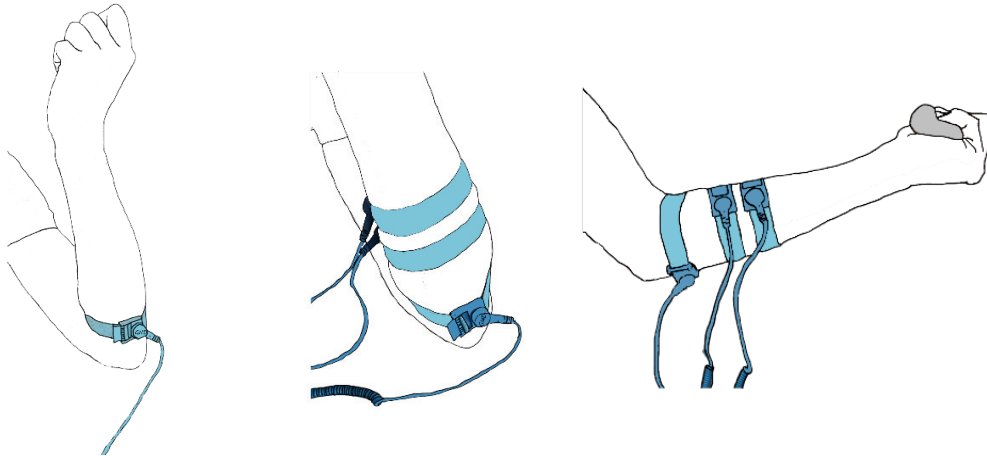


Figure 6.2: EMBody works with a variety of electrodes. Here, they are affixed with straps to the forearm. First, the ground electrode is placed (left) on a location with little muscle fiber, e.g. close to the elbow. Afterwards, the sensing electrodes are placed on the muscle belly (right).

6.2.2 An Exploration-Centered Workflow

EMBody uses an exploration-centered workflow (Figure 6.1) guiding users from a first idea to a final prototype. This allows users to readily start exploring suitable interaction scenarios without the need for further configuration. The data-centric pipeline provides different views of the same EMG data, such as raw and filtered data, as well as generated features and final predictions. The following section introduces this workflow in detail while highlighting extension points for expert users.

Connect Electrodes and Select Channels

The standard firmware on EMBody’s hardware offers a captive portal when connecting to its Wi-Fi network. The portal allows users to configure their preferred connection settings. Once configured, the prototype readily sends recorded data via the network. The system can be powered by any portable power source, such as a small powerbank, allowing continuous operation for multiple days.⁸ Data is transmitted wirelessly without the need for additional cables apart from the electrode connections. Setting up the prototype and placing electrodes (Figure 6.2) is described in the enclosed manual⁷.

⁸ See Section 6.2.3 for performance details.

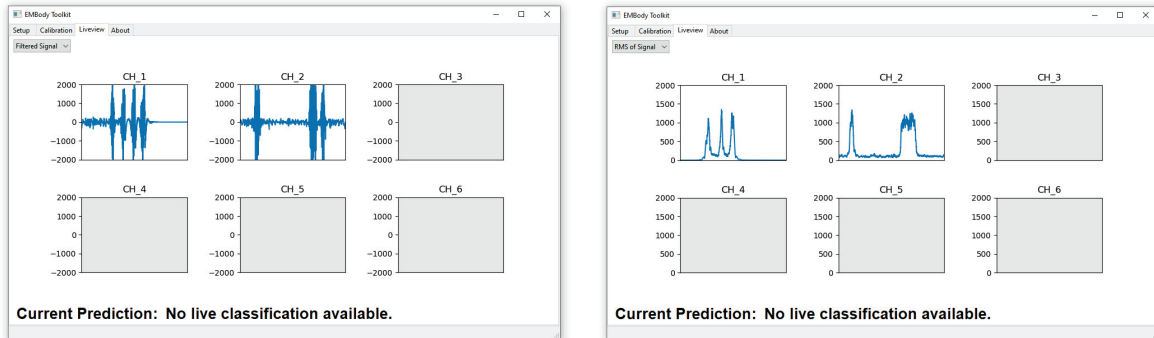


Figure 6.3: Checking the EMG signal in the live view using different views of the EMG data. Filtered signal on the left; generated RMS features on the right. See Section 6.2.4 for algorithm details.

Our toolkit allows recording up to six channels with a standard sampling rate of 250Hz and provides sensing and recognition data via a UDP stream broadcast over the network. This enables users to track several muscles (groups) at once and recognize complex motor tasks. For technically skilled users, the firmware can be adjusted to their needs, offering a much higher sampling rate. Network capabilities are the bounding factor. Additionally, the software implements sanity checks on the received data, such as estimating sampling rate and tracking package loss during critical operations and informs the user about possible ways to solve these issues.

The EMBody live view (Figure 6.3) allows simultaneous tracking of up to six channels and helps the user identify faulty connections. This also allows refining electrode placement when signal quality is low. The manual provides a set of guidelines on how to place electrodes to minimize noise. This first step helps users familiarize themselves with the EMG signal and discover how it reacts to their movements.

Define Gestures and Calibrate

Once electrode placement is completed, users provide a set of movements that they wish to recognize. EMBody will guide the user through this calibration process by instructing the user to perform the respective movements while collecting sample data for each movement (Figure 6.4). Additionally, EMBody verifies that the sampling rate is sufficient for further filtering steps and monitors potential package loss. If irregularities are detected, the user is advised to repeat the calibration, check for connection issues or redo the electrode setup.

During calibration, EMBody synchronizes the specified movements (the calibration labels) and incoming EMG data samples. EMBody collects more data samples for the *NULL_CLASS*, allowing the user to present motions that should not be recognized. This increases the robustness of the classifier. After completion, the recorded data is filtered and saved. An updated overview over all collected calibration data (duration per label) is displayed. For post-hoc analysis, EMBody offers an export function.

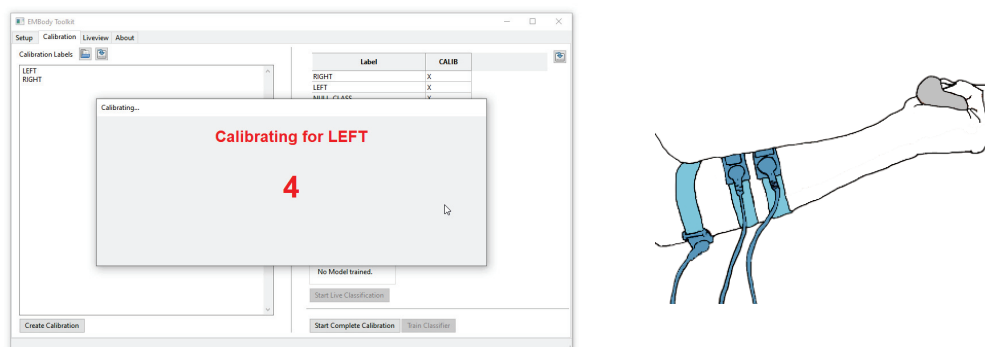


Figure 6.4: EMBody calibrating for a gesture labeled *LEFT*. Corresponding gesture by the user (left arm) on the right.

Train Model and Start Live Classifier

After completing a full calibration, the user is able to train a classification model using the provided discriminative method using a support vector machine (SVM). Internally, EMBody calculates Root-mean-square-based (RMS) features after filtering the data (cf. Chapter 4). Depending on the selected amount of channels, pair-wise ratios between channels are calculated. This approach provides an indication of relative locality for the classifier. After generating the appropriate features, the software trains an SVM and evaluates the model using 10-fold cross validation on the calibration data.

While the significance of this metric is limited to the recorded calibration data, it supports the user in assessing whether the calibration movements are sufficiently distinct with respect to their recorded muscular activity. Low values⁹ indicate that the chosen gestures are too similar. Thus, the user is advised to modify their gestures (backtrack in the workflow) or to provide additional sensing channels which may help distinguish the gestures (restart with electrode setup). This way, impractical electrode configurations and gesture sets can be identified quickly.

The standard feature generation and classification method of EMBody works well for short explicit EMG-based input. To recognize longer movements¹⁰, other machine learning approaches, e.g. regression or correlation-based methods might be better suited. The modular structure of EMBody allows users to substitute and extend the classification module, providing their own training and prediction routines. If the user decides to use the method as implemented, no knowledge about classifying EMG data is needed to use EMBody.

After starting live classification, EMBody switches to the live view (Figure 6.5) and continuously processes the incoming EMG data, generates respective features and provides a new prediction¹¹ every 80ms.

⁹ Values lower than 80% might already be impractical for some applications.

¹⁰ More than several seconds.

¹¹ Based on the provided gesture set.



Figure 6.5: Live view with live classification. Note that the classification is also streamed via UDP.

In the context of proficiency-aware systems, this abstraction allows user to experiment with EMG as a proxy for proficiency. For example, using the standard workflow to first calibrate guitar chords (cf. Section 4.2) and then use the resulting predictions in a tutoring systems or using correlation-based methods for sustained movements (cf. Section 4.3).

Connect to UDP Stream

Live classification is displayed within the software and additionally provided as a network stream (UDP or LSL [170]). Whenever a new prediction is available, EMBody broadcasts the appropriate gesture label onto the network, which can then be accessed by any other application in the network.

6.2.3 Hardware

Our toolkit includes a versatile hardware system, capable of measuring up to six EMG signals in parallel and delivering them wirelessly over WiFi (Figure 6.6). The device senses the muscular activity primarily using dry electrodes, but is capable of working with other electrode types. Compatible electrodes must facilitate a bipolar measurement technique: apart from one reference electrode, two sensing electrodes are used to minimize the impact of noise artifacts. These electrical signals are individually processed by an analog instrumental amplifier and quantified by an Analog-Digital Converter (ADC). Finally, a microcontroller packages the data into UDP packets and transmits them using a WiFi antenna.

We based our design on existing circuits, aiming for compactness, wearability, and flexibility, while ensuring low noise levels and adequate data output. The instrumental amplifiers are an adaptation of an existing design¹², based on the INA2321¹³, a low-power and low-cost

¹²<https://github.com/BigCorvus/2-Channel-Biopotential-Amp>

¹³<http://www.ti.com/product/INA2321>

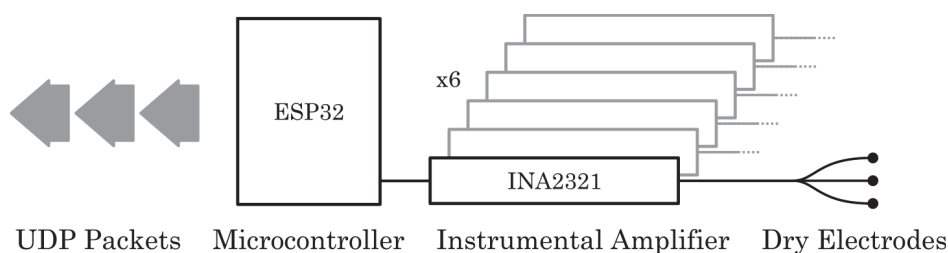


Figure 6.6: Diagram of the EMG system: The system senses muscular activity with electrodes. These signals are amplified and sent in UDP packets over WiFi.

CMOS amplifier. We used a board based on the ESP32 microcontroller, which is a low-cost and low-power *system on a chip* with integrated WiFi. Given the pin layout and the usage of WiFi, the ESP32 offers a total of six remaining ADC channels with a 12 bit resolution each, thus converting the output signals of the instrumental amplifiers to integer values from zero to 4095. Power consumption for the ESP32 with active radio transmission is approximately 240mA ¹⁴, given maximum signal strength. Using a 5200mAh powerbank yields up to 22 h of continuous operation (the power consumption of the amplifier is negligible).

During operation, the microcontroller polls all six channels and packs the measurements into frames of six values including a timestamp and broadcasts them to the connected network¹⁵. For EMG purposes, a sampling rate of 250Hz is sufficient in most cases¹⁶. We provide all firmware files on github¹⁷, allowing users to customize it to their needs.

To ensure portability and versatility, the system is mounted in a 3D-printed case with a wallet-like form factor and powered via the ESP32's USB port. Both powering the device and connecting it to a computer can be done with a micro-USB cable. Further, the inputs of the instrumental amplifiers are connected to 3.5mm stereo audio jacks. This simplifies the management of the electrode cables and allows to adapt their number to specific requirements. Figure 6.7 depicts the hardware inside the case.

The complete system can be built from our custom circuit design schematics and layouts which are available via the github repository. Additionally, we provide assembly instructions, including a parts list. The available firmware and 3D design files complement the hardware resources.

¹⁴https://www.espressif.com/sites/default/files/documentation/esp32_datasheet_en.pdf

¹⁵Communication protocol: "timestamp;CH_x;CH_x;..."

¹⁶Proposed filters by related work [261] make higher sampling rates unnecessary when using the standard pipeline of EMBody (see Section 6.2.4 for details).

¹⁷<https://github.com/HCUM/embody>

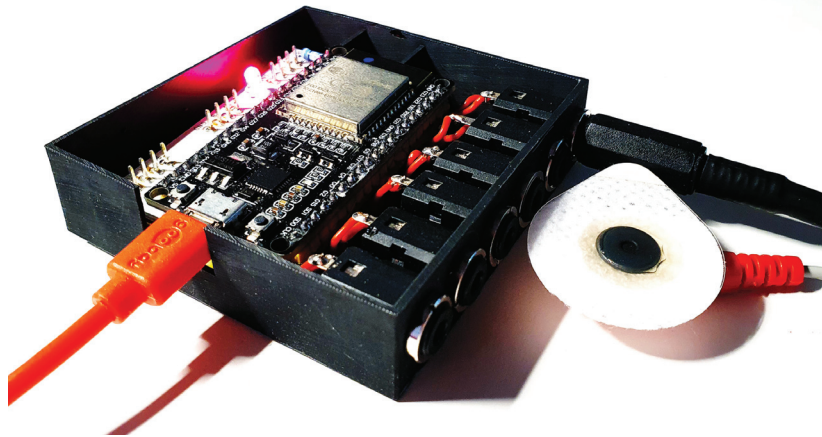


Figure 6.7: EMBody’s hardware prototype showing the six audio jacks on the front to connect electrodes and the microcontroller in the middle. The three amplifiers (each supporting two channels) are placed underneath the microcontroller. A lid (removed in this picture) is also provided.

6.2.4 Software

The EMBody software is a PC application developed in Python on Windows. It does not use native libraries, allowing it to be run on macOS and Linux as well. It follows a modular structure governed by EMBody’s workflow. While it has been developed to be used with EMBody’s hardware prototype in mind, the software can be used with any kind of sensor that uses the communication protocol¹⁵. New research probes such as PhysioSkin [217] and PolySense [122] are promising alternatives that are potentially compatible with our system. The complete source code is open-source and can be readily extended, both in terms of additional GUI elements (making use of EMBody’s stream handling) and logic components (extending the data processing pipeline). The following section describes key components and highlights possibilities for extensions.

Data Processing Pipeline

Data filtering, processing and feature generation is encapsulated in EMBody’s `ClassificationManager`. The following methods are of particular interest. Figure 6.8 shows the call hierarchy and information flow during the calibration and live classification phases.

onRawCalibrationDataAvailable

After EMBody finishes a calibration run, it calls the method `onRawCalibrationDataAvailable` and passes all recorded samples and associated labels (see the source code for details). This method implements preprocessing steps (using `preprocessData`) and populates the internal data structure.

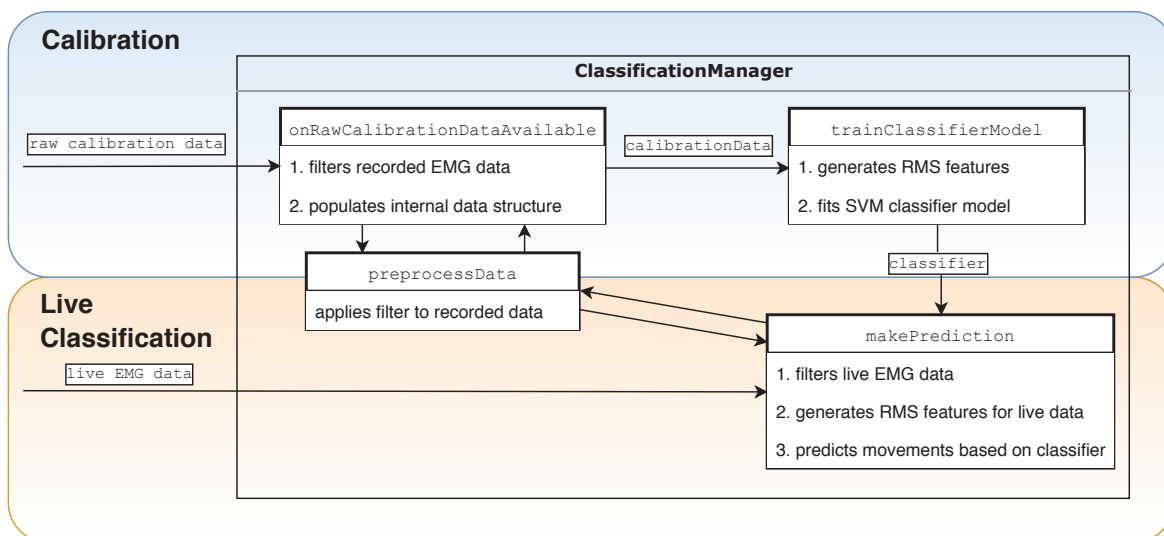


Figure 6.8: Classification pipeline within `ClassificationManager`. Note the different flows for the calibration and live classification phases.

preprocessData

Closely following related work [206, 261], this method applies a bandpass filter between 2 Hz and 100 Hz , attenuating long-term drifts, the DC offset and high-frequency noise as well as a bandstop filter between 49 Hz and 51 Hz in order to remove power line interference. The method returns a dataframe linking data samples to their respective class, i.e. the calibration labels. Additionally, data is grouped per calibration run¹⁸.

trainClassifierModel

Implementing standard EMG features (cf. Chapter 4) for HCI applications, this method provides epoched RMS features and their pair-wise ratios between channels. These values can be interpreted as a proxy for the intensity of muscle activity as the amplitude of the EMG signal increases when the muscular activity increases [206]. RMS is calculated using a convolutional approach and defined as

$$x_{RMS} = \sqrt{\frac{1}{n}(x_1^2 + x_2^2 + \dots + x_n^2)}.$$

One important parameter for calculating RMS-based features is the window size n . It represents a trade-off between classification accuracy and latency, i.e. the time between acquiring EMG data and its prediction. Small windows allow for little latency, but are problematic when recognizing longer-lasting movements. Preliminary experiments confirmed that setting the window size to $n = 20$ yielded a good trade-off. Given a standard sampling rate of

¹⁸The `NULL_CASS` may be recorded multiple times during calibration. Grouping ensures that those samples are processed separately.

250 Hz, this corresponds to a classification latency of 80 ms. These are the standard values in EMBody which can be modified when required.

EMBody implements a support vector classification with a radial basis function kernel¹⁹ using scikit-learn²⁰. Features are scaled to unit variance and zero mean before fitting a model. A subsequent 10-fold cross validation provides a first indication of the model accuracy.

Overwriting the `trainClassifierModel` method allows the user to specify their own feature generation pipeline (cf. [236] for an overview). They can provide custom classification methods. When using scikit-learn, the user may still make use of the live prediction methods. Implementing other libraries, e.g. a correlation analysis, requires the user to also adapt `makePrediction`.

makePrediction

During live classification, the previously trained model is used to predict calibration labels for incoming data. Whenever a new prediction is requested²¹, this method performs the pre-processing and feature generation steps and provides the respective predictions. A voting (mode-based) ensures a robust prediction, hence yielding one prediction per call. Alternatively, a list of predictions is also provided. Users may choose to work on the raw prediction data directly, or accumulate incoming data, e.g. for activity recognition, by overwriting this method. Listing 6.1 provides an excerpt highlighting key steps and possible extensions points.

```
def makePrediction(self, data):
    ...
    #applying filtering steps (Data is in df), change here for individual filters
    apply_bandpass_filter(df, 2.0, self.currentSamplingRate / 2.0 - 1.0, self.currentSamplingRate)
    apply_bandstop_filter(df, 49.0, 51.0, self.currentSamplingRate)

    #constructing data matrix
    X = pd.DataFrame()
    for column in df.columns:
        #add custom features here
        X['rms' + str(column)] = rms_convolution(df[column], self.windowSize)

    #automatically add pairwise ratios of all features
    addPairwiseRatios(X)

    #predict based on generated features in X using pre-trained classifier clf
    try:
        X = self.scaler.transform(X[X.columns])
        prediction = self.clf.predict(X)

        #EMBody uses voted predictions by default (over 80ms of data)
        #return prediction to get a result for each sample within data
        voted_prediction = mode(prediction)[0][0]
        self.currentPrediction = str(voted_prediction)
        return self.currentPrediction, prediction
    except ValueError:
        self.currentPrediction = None
        return None, []
```

Listing 6.1: Excerpt of `makePrediction` highlighting key steps.

¹⁹ $C = 1.0$, $\gamma = scale$

²⁰<https://scikit-learn.org/stable/index.html>

²¹By default, every window size (80 ms).

Stream Handling

In EMBody, the class `StreamHandler` handles the incoming UDP stream from the EMG device as well as the outgoing stream of predictions. Together with the GUI elements, this class implements an observer pattern to inform and update the GUI elements. To that end, `StreamHandler` implements `StreamEventCreator`, allowing it to trigger stream events when required to inform appropriate views. Users who extend this class, or `StreamEventCreator`, are encouraged to deliver GUI updates by notifying their observers.

GUI Elements

Similarly, existing GUI elements receive updates by listening to incoming stream events, by implementing `StreamEventListener`. EMBody provides the following functional views²²:

- **Setup:** Connecting to the prototype, checking sampling rate and selecting channels.
- **Calibration:** Specifying (save/load) calibration labels, running calibrations and training classifier models; exporting complete calibration data (filtered).
- **Live view:** Live feed of incoming EMG data grouped by channel, additional live predictions if classifier is available.

A possible extension is including a new view, showing a live auto-correlation of the signal for different lag sizes. For this purpose, one would want to connect to the live feed of EMG data (Live view) and subclass `StreamHandler` accordingly to provide custom stream events, after processing the incoming data. A new view may listen for events and plot data when prompted. A view-less listener may simply save processed data to a file.

6.2.5 Technical Limitations

EMBody is custom-built and tailored for prototyping purposes. As such, it has not been designed as a precise measuring unit, but to support EMG-based interaction in prototypes. The toolkit is a trade-off between signal accuracy and accessibility for non-expert practitioners. Consequently, we note some important limitations of EMBody.

First, EMBody is not an exact measurement device. Due to its open design, it is especially vulnerable to artifacts, such as cable movements or electromagnetic noise. While the implemented filtering steps mitigate these effects, proper setup routines are still vital. Similarly, since EMBody does not provide adjustable gain settings, it represents a trade-off between being able to recognize small, fine-grained motions and extensive movements. The size of

²²Please consult the manual for details at <https://github.com/HCUM/embody/tree/master/manual>.

the electrodes particularly influences the resulting signal. Hence, users are encouraged to choose electrodes according to the desired usage scenario.

Second, our toolkit works best for isometric muscle activation, i.e. continuous muscle activity without visible movement [206]. Isotonic muscle contraction can be problematic, e.g., recognizing movements over a period of multiple seconds. Here, correlation based on previously calibrated templates might be more suitable (cf. Section 4.3). The modular structure of the accompanying software allows for adapting it for that purpose.

Third, for more complex movements, one might require more channels than EMBody can accommodate. Even using up to six channels requires extensive cable management. As EMBody was designed to allow for using all muscles, we do not include cable arrangements in the system. Custom cable solutions can help to alleviate this issue by combining multiple leads or even go one step further, such as PhysioSkin [217] and PolySense [122].

Finally, despite being a mobile prototype, EMBody still requires an active Wi-Fi connection for broadcasting. Subsequently, a mobile application on a smartphone providing an access point and data processing capabilities is needed for true mobility.

6.2.6 Target Audience

EMBody is tailored for a wider audience, supporting interaction designers, researchers and engineers. Consequently, EMBody offers different levels of depth and complexity in signal processing and classification. In the following, we outline two typical use case scenarios to showcase the diverse needs of EMBody's users. The first scenario describes how a VR interaction designer uses EMBody to realize dynamic interaction in a sword-fighting game. The latter scenario deals with a researcher collecting electromyograms throughout an experiment for prosthesis control. We showcase how both scenarios can be integrated in proficiency-aware systems.

Scenario 1: Using EMG Input in VR Prototypes

An interface designer wants to extend their VR application using the standard workflow (Figure 6.1). They want to sense how strongly the user is gripping the VR controller. The designer decides to include two different grip modes, normal and hard, in their sword fighting game. Thus, grip strength has a direct impact on the sword's momentum, influencing the player's ability to attack and parry. This scenario is an example for EMBody's potential user group composed of interaction designers and application developers. Their requirements include easy access to interpreted EMG data, while allowing for fast iterations among possible gestures. The underlying processing and interpretation of the EMG data is secondary and not of interest to this user group.

The designer first needs to find a suitable electrode location. A pragmatic approach to this question is to observe one's own muscle movements and place a pair of electrodes on the involved muscles. A tight grip mostly activates muscles in the forearm, hence the designer

decides to place electrodes on the underside of the respective forearm (Figure 6.2). Afterwards, they complete the setup by connecting the electrodes to the prototype as depicted in the manual²³. The designer specifies the two recording channels (one for each forearm) and checks the signal via the live view. They observe whether their movements trigger changes in the displayed signal as illustrated in Figure 6.3.

In his application, the designer is only interested in detecting a tight grip with either hand. Thus, they provide the labels: *LEFT* and *RIGHT* and create an empty calibration. EMBody automatically adds a *NULL_CLASS*, which represents any other motion. The designer starts the calibration process and provides a tight hand grip when prompted (Figure 6.4), making sure to relax and perform other relevant sword swinging motions in between.

EMBody reports an average accuracy of 95.6%. If need be, EMBody provides them with the means to reiterate the gesture calibration using a different electrode configuration or gesture labeling. Here, the designer is satisfied with the result and starts the live classification. EMBody now switches to the live view (Figure 6.5) and continuously processes the incoming EMG data, predicting if the user tightens his grip for either hand. The designer incorporates the live classification into their VR application by accessing the UDP network stream in Unity. They affix the prototype on the belt of the user and put the powerbank in their trouser pocket. As EMBody continues to relay live classifications of grip force, the designer can focus on tweaking parameters relevant for the game, e.g., how much stronger a strike should be when the sword is tightly gripped.

In a proficiency-aware scenario, the calibrated data could be used to detect a proper sword strike. EMBody offer numerous opportunities to do so: (1) by intentionally calibrating for good and bad strokes, allowing the toolkit to distinguish between them; (2) by using a classifier that outputs confidence in its prediction, only accepting high confidence strokes; (3) by providing an appropriate gold standard (the perfect stroke) for a correlation-based approach (cf. Section 4.3). While (1) requires no modification, (2) includes minor changes to `trainClassifierModel` and `makePrediction`. To realize (3), one must additionally modify calibration (`onRawCalibrationDataAvailable`).

Scenario 2: Designing Experiments for EMG-based Input

A researcher familiar with Electromyography recording wants to use EMBody to find a suitable set of classification features to recognize ten different hand movements when using up to six channels connected to the forearm. Contrary to the designer scenario, this user group includes signal processing experts and researchers as well as engineers and machine learning developers. They require close control over the data processing pipeline allowing them to customize vital steps if need be. Rapid prototyping of different designs is secondary. This user group focuses on signal accuracy and body physiology, requiring detailed views of the recorded data. This use case is strongly inspired by the conducted research in Chapter 4 and directly relates to EMG-based proficiency-aware systems.

²³ Available at <https://github.com/HCUM/embody/tree/master/manual>.

Since flexion and abduction of the wrist as well as controlling finger movements involved various muscles in the forearm, the researcher decides to place the electrodes in two rings around the forearm (one closer to the elbow, one near the wrist) to capture most of the involved muscles. This also helps them to generalize their approach more easily, as exact knowledge of the forearm anatomy is thus not required to place electrodes. Contrary to Scenario 1, the researcher makes use of the unipolar measurement technique²⁴ as shown in Figure 6.9. They use a custom hardware device to collect the data and send a UDP stream, adhering to EMBody’s protocol, to deliver the data.

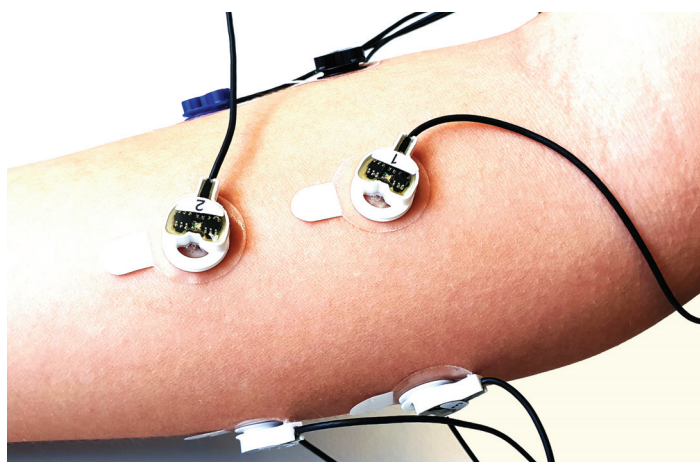


Figure 6.9: Unipolar measurement with six sensing electrodes (white, two on backside) and one reference electrode (blue). Ground electrode (black) serves for noise reduction only.

The researcher connects their UDP stream to EMBody and selects all recording channels. In the live view (Figure 6.3), they observe whether their movements trigger changes in the displayed signal and initially assess the viability of the current electrode configuration by comparing different channels and respective signals. The researcher notes possible improvements regarding the configuration but decides to do an initial recording. They define the ten gestures and start a complete calibration process. Using EMBody’s save functionality, the researcher exports a dataset with annotated ground truth from the calibration. They repeat the process with two different electrode configurations and fine tune their classification algorithm and extracted features in their own work environment. Here, EMBody provides this user group with easy access to annotated data streams for research purposes.

After establishing a sufficient model, the researcher incorporates their classification algorithm into EMBody by extending its *ClassificationManager*. Henceforth, EMBody will use the tailored algorithm to process incoming data, taking care of signal acquisition and routing. For the actual experiment, the researcher attaches the electrodes in the optimal configuration for every participant and executes the calibration procedure. The researcher connects to EMBody live classification via UDP and relays the current prediction to the prosthesis. They record accuracy metrics in a manual task and questionnaire responses for later analysis.

²⁴one GND/REF electrode each, several measurement electrodes.

6.2.7 Formative Evaluation

Organizing two workshops which featured rapid prototyping of EMG interfaces was a key element in designing and implementing EMBody. This way we assured that the final version of our system reflected the needs of the HCI community. In this section, we illustrate how we established requirements and challenges for EMBody during the workshops. We used formative evaluation to understand the qualities necessary for EMBody to enable exploring EMG-based systems.

Workshops: Initial Feedback and Refining Requirements

We organized two experimental workshops titled "Using Physiological Sensing for Embodied Interaction" for university students in HCI. Thirty-one and 36 students, including bachelor, master, and Ph.D. students participated in the workshops. Figure 6.10 shows one workshop location during the hands-on sessions. Participants learned about physiological sensing (mainly EMG) and created their own EMG-controlled devices in hands-on tutorials. The workshops provided an opportunity to verify whether an initial version of the toolkit offered easy entry for EMG-based interface design. It included a hardware prototype using the Bluetooth protocol to transmit data and a set of processing script to receive the data. After being presented with the system and its functionality, participants were instructed to first define their own ideas (e.g., an EMG-controlled musical instrument) and formulate a concept for their prototype. Over the course of the two-day workshops, participants successfully developed fully functional prototypes, which they evaluated in small user studies. Finally, each student team presented their work in front of the students and teachers.

The goal of the workshops was to evaluate the suitability of the toolkit for rapid prototyping and identify possible design flaws. Every participant group was able to successfully create a fully functional prototype system that employed EMG as an input modality. Examples²⁵ included:

- *SmartSpine*: helping the user to correctly lift heavy loads by placing electrodes on the legs and back.
- *Flappy Bird*: controlling the game Flappy Bird via flapping one's arms (electrodes on the arms).
- *Muscle PIN*: biometric authentication via muscle flex patterns (electrodes on the forearm).
- *Dance Avatar*: a puppet mimicking the user's every move; electrodes on arms and legs.
- *Lunar Lander*: a collaborative game where players control a lunar lander probe. Two players steer the probe via electrodes on the forearm.

²⁵ Selected examples are shown in EMBody's video at <https://github.com/HCIUM/embody>.



Figure 6.10: Hands-on session during one of the workshops.

- *Canoeing*: a four-player game where teams of two compete in a canoe competition. Each team needs to maintain a consistent paddling motion (electrodes on the forearm.)

The breath of ideas generated in the workshops as well as the fact that students at varying levels of HCI and technical competence were able to rapidly build functional systems show that the initial toolkit effectively supported rapid prototyping. Most importantly, the workshops enabled us to verify if the toolkit fulfilled the requirements and what parts of EMBody needed improvement. The prototypes were highly *mobile*—workshop participants built prototypes using different muscles and in different location. Workshop attendees were also able to effectively perform *data acquisition* as they all successfully connected muscle sensing to application input. All the members of the diverse audience in the workshops were actively involved in building the prototype, thus showing that the toolkit offered an *abstraction* level that was effectively used by the participants. Finally, participants with expert signal processing knowledge were able to add advanced computation to their prototypes by taking advantage of the toolkit’s *modularity*.

The workshops also enabled us to identify key areas for improvement for EMBody. First, some participants experienced issues with Bluetooth connectivity. To alleviate that issue, we redesigned the toolkit to rely solely on the WiFi connection. Second, we observed that the workshop participants spent a significant part of their prototyping time designing algorithms for recognizing movements. This was especially true for those who did not have extensive signal processing experience. Consequently, we decided that EMBody should include pre-defined gesture detection tools that could be customized by expert user. The updated

software package includes a default classifier that can be used with no knowledge of EMG gesture recognition as outlined in Section 6.2.4.

Sample Applications

The next step in our process was to develop systems that would enable experimental studies. To demonstrate EMBody's versatility and verify the correctness of the workflow, we built several interactive systems which used EMG input for different purposes.

Choosing the Right Input Control for EMG

To better understand a user's perception of their own muscle control, we endeavored to investigate how users perceive different input control mappings in a steering law experiment. The goal was to keep a moving ball as close as possible to a predefined trajectory (Figure 6.11). Two EMG channels were used, each controlling one horizontal direction, while the ball moved upwards on a screen. Electrodes were placed on the respective forearm. The study apparatus is depicted in Figure 6.11.

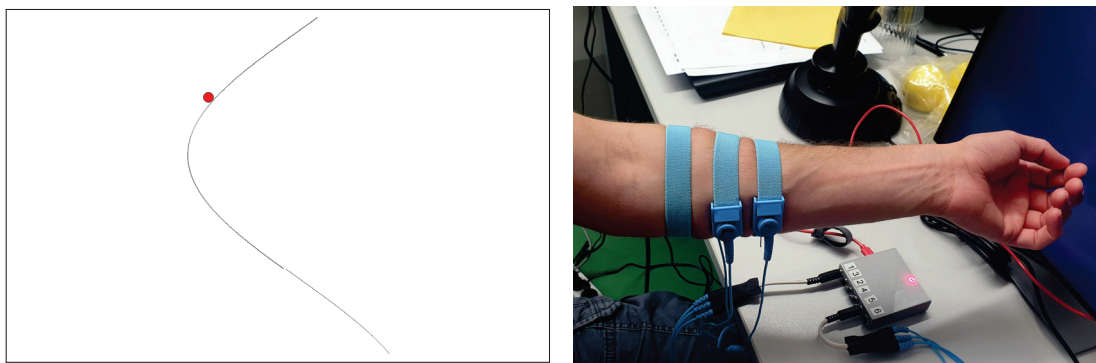


Figure 6.11: Study stimulus (left) and electrode placement (right) for EMG input controls.

Here, EMBody allowed us to test a series of electrode locations and suitable muscle groups as well as various input mapping functions in a rapid fashion for a final study. We conducted a within-subject experiment using three different modalities to control the ball: a joystick as baseline and two EMG-based controls (position and rate control). While position control directly changed the ball's position based on the recorded power of muscle activation (controlling its velocity), rate control influenced the acceleration in either direction. Different ball speeds were introduced as an additional independent variable.

We measured the average deviation from the given line as well as participants' responses to our questionnaire about ease of use, their perceived fatigue and their feeling of control²⁶.

Figure 6.12 illustrates the questionnaire responses w.r.t. the modality. While the Joystick clearly outperformed the EMG-based controls in terms of control and fatigue (lower

²⁶All on a visual-analog scale from 0 to 100.

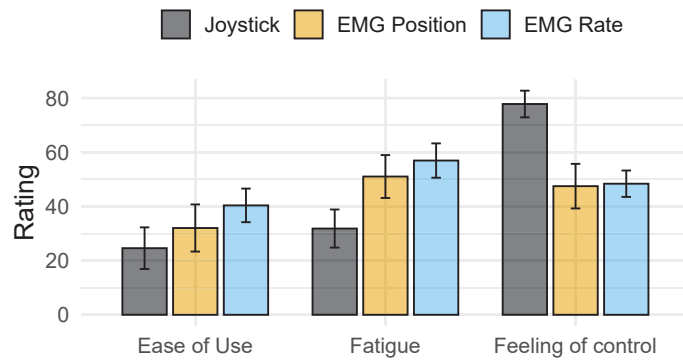


Figure 6.12: Averaged responses for our questionnaires (ease of use, fatigue, feeling of control) given modality. Visual-analog scale from 0 to 100.

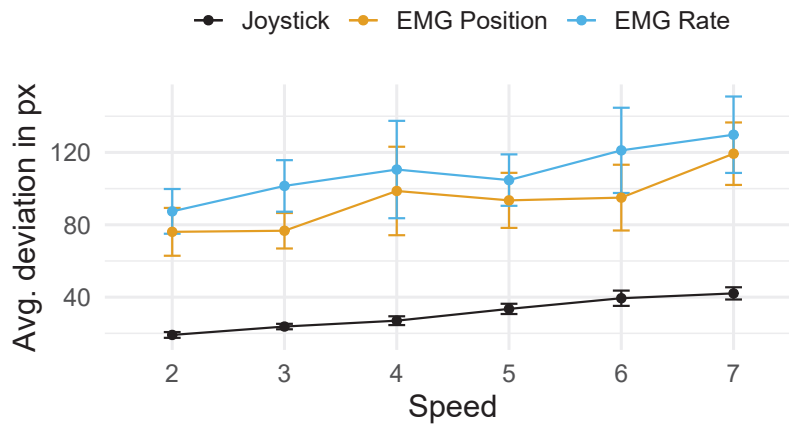


Figure 6.13: Average Deviation (L2 norm) in pixels given different speeds and modalities.

is better), EMG-based inputs were preferred in terms of ease of use. In a preliminary evaluation (one-way ANOVA and Tukey posthoc comparisons) of ten participants ($8m/2f$), we found that the feeling of control via the joystick was significantly higher ($F(2,27) = 7.7, p < 0.01$) than for both EMG-based controls. Additionally, rate control was significantly ($F(2,27) = 3.4, p < 0.05$) more fatiguing than the joystick.

The superior control of the joystick baseline can be seen in Figure 6.13 (significantly different to position and rate control). We observed a significant linear effect²⁷ of ball speed for all modalities. There was no interaction effect between speed and modality. Additionally, position control did not significantly outperform rate control for EMG-based input.

While EMG-based input was lacking in control for steering tasks, there was a tendency for improved ease of use. In line with this finding, we believe that EMG is better suited as a

²⁷Linear mixed model analysis. Fixed: modality, speed. Random: trialnumber, participant.

secondary input modality, e.g. for hands-free interaction. Here, EMBody supports designers in prototyping the right placement for electrodes while ensuring adequate control.

EMG as Secondary Modality for Manual Tasks

Similar to the previous work, we also investigated the suitability of EMG as a secondary modality for manual tasks. In tasks where both hands are occupied, interrupting the task to perform additional actions often induces a focus shift that potentially breaks the workflow. Here, we evaluated if EMG-based gestures can provide a seamless way of triggering additional actions, similar to our work with EMPiano (cf. Section 5.3), but now with a low-cost prototype. In a mimicked soldering task (see Figure 6.14), where users could additionally control the third hand, we compared EMG with direct manual interaction and a pedal-based system.

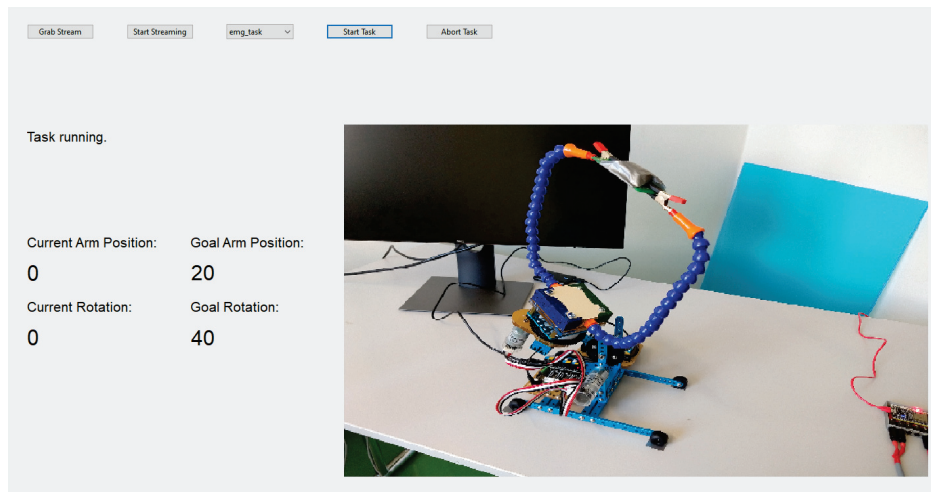


Figure 6.14: Task instructions as seen by participants. Final target position was given in rotation angles (left) and visualized as a picture (right). The platform was actuated through a robot for the EMG- and pedal-based condition.

We found that EMG allowed for a flexible way of defining gestures to move the third hand, while the pedal-based systems induced mapping errors of rotation directions for some users. In this work, EMBody allowed us to calibrate and classify the individual user gestures during the experiment, streaming predictions directly to our robot platform that actuated the third hand.

Taking a Look Inside

Besides explicit interaction, EMBody allows to explore EMG for implicit interaction. Here, we evaluated to what degree insights into one's own muscle activation can be beneficial in learning motor tasks. The full project description and implications of FitBack are detailed in Section 4.3 and Section 5.1, respectively. This section will focus on how EMBody aided in the execution of the study.

We used EMBody to quickly connect recognized muscle motions to alternative feedback modalities via the UDP stream. This allowed us to focus on possible designs for the feedback. The detection of correct exercise form was detected using an adapted version of EMBody’s machine learning algorithm. Here, EMBody allowed us to evaluate several feedback options for a specific task, while relying on the same mobile signal acquisition system. Changing feedback was easily possible and only depended on how we wanted to present the muscular activity.

6.2.8 Summative Evaluation

During the development process, we iteratively refined EMBody to address upcoming challenges. In line with Ledo et al. [179] who suggested multiple evaluation strategies and goals for HCI toolkits, we first evaluated an instance of the *usage* of EMBody in workshops, as described above. In order to establish the capabilities of the final version of the prototype, we further evaluated the toolkit through a series of expert interviews.

Participants

We recruited five HCI experts who participated in at least one of our workshops. All participants were male and aged $\bar{x} = 27.6y$ ($s = 3.6y$). No remuneration was provided for the interview. Table 6.1 details the profiles of the participants. We chose interviewees so that they would be member of the primary target audience of the toolkit—HCI researchers with varying level of technical knowledge and different research foci. The participants had varying levels of experience with prototyping. Apart from one, all interviewees were prototyping at least once a month involving microcontrollers, AR/VR applications and small electronics projects.

ID	Age	Gender	Profession	Areas of expertise	Experience	Frequency
P1	31	male	PhD student	UX design, SW dev.	7	Once a month
P2	29	male	Postdoc	HCI, sports	7	Twice a week
P3	30	male	PhD student	HCI, AR/VR	6	Once a month
P4	22	male	Student	HCI, participatory design	5	Once a week
P5	26	male	PhD student	Machine learning, NLP	3	Once a year

Table 6.1: Participant profiles in the interviews, including their areas of expertise as well as prototyping experience (7-item Likert scale) and frequency.

Interview script

At the start of the interview, we asked the participants about their experience in the workshops, specifically what challenges they faces while realizing their project. Afterwards,

participants watched a video of the final version of EMBody. We then inquired about the participants' initial perceptions of EMBody and its exploration-centered workflow. Next, we asked about the challenges and opportunities they saw in using the toolkit. Finally, we discussed possible applications of EMBody in the participants' research work.

Analysis

All five interviews were recorded (total duration 1 : 42h) and transcribed verbatim. We wanted to conduct a focused analysis of a moderate volume of qualitative data. Consequently, we used the pragmatic approach to thematic analysis [24]. We established an initial coding tree by open-coding a representative 20% of the material by two researchers and aligning the codes. The rest of the interviews were then split between the coders and analyzed by a single researcher. In a final session, we refined codes and identified recurring themes in the data.

Results

The final discussion resulted in the following high-level themes: GRADUAL LEVELS OF FIDELITY, TARGET AUDIENCES, TRANSPARENCY and CHALLENGES IN WORKING WITH EMG. We further detail the contents of each theme in the following.

Gradual Levels of Fidelity

EMBody aims to support a wide range of user groups, from novices to experts. In our interviews, participants appreciated the simplicity of starting to work with EMG-based interaction as the toolkit allowed them to focus on designing gesture and movements without the need to bother with signal interpretation:

I could simply just start with coding my interaction [...] as a designer don't have to deal with the signal and all. (P2)

This new approach would also enable a bit more non-technical people who actually don't care about a signal, just about the application to get into EMGs sensing. For example, people from non-computational fields that just want to try out something, I think, for them it's really more accessible then. (P5)

It also became evident that experts benefited from the modular data processing pipeline. Moreover, they would appreciate an in-depth classification report already in EMBody's base version:

I think it's important to be able to get the raw data, maybe someone wants to do some work on machine-learning algorithm on it and do something else with this data. I think it's important to have most or all data and have this abstraction. Different people can use it differently, depending on the flow. (P3)

I want it to be as accurate as possible, so I would want to have access to everything to be able to customize everything. (P4)

Target Audiences

Participants remarked on the suitability of EMBody for different audiences, such as people with no experience in computer science. The exploration-centered workflow allowed them to quickly grasp the idea of EMG-based interaction and supported an easy entry:

What would they have to do? Let's say the box itself, the board itself would be nicely presented in a nice cover. You just have to plug it in. Easy, that's something you do all the time. Connect it to wifi, so that probably you just immediately opens the wifi and you can connect it from the computer. That's something you always do, so that should be possible I guess. Then afterwards, installing the application. That's easy to do with the installer. You don't have to compile it. Then that would be obviously very feasible as well (P1)

However, participants noted that the final version of EMBody focused more on curious audience and researchers who wanted to conduct EMG experiments. Here, participants discussed the possibility of introducing different user modes within the application:

I think this is one of the disadvantages and one of the issues. It's not optimized for the public (P2)

What I can imagine is that you have these abstractions layers for expert users that really want to see the signal. They're really interested in how this classification actually works, and you have something, let's say, novice mode where you don't see that. (P2)

Transparency

Presenting users with the live signal and the respective gesture predictions allowed users to "get a feel of what the signals looks like" (P4). This greatly increased the transparency of the toolkit and its data processing pipeline:

When you're doing live prediction, it's nice to have the signal in front of you and the prediction just next to it, so you know if you've done something wrong. (P4)

I think you get a lot of transparency towards the user, whoever that may be. Could be a regular user, could be a researcher. You see how the signal's actually working, you see the signal, and so on (P2)

Furthermore, EMBody's workflow was immediately recognized, closely following related procedures for prototyping and experimenting with physiological signals:

I think we used the very same procedure. We first set up the prototype, so we connected electrodes, then we— well, first, the gesture we wanted to use were defined beforehand but we still had to calibrate. After the calibration, we trained the model and we started live classification to use a prototype. It's the same workflow. (P4)

Interviewees appreciated that all of EMBody's source code and documentation is open-source, ensuring transparency of the workflow and algorithms:

Yes. I think that's very convenient to have. Also, it's open-source. They can want to extend it I guess. (P1)

(Talking about Myo armband) It was not open source. This is maybe a key difference between the commercial product and your project. They did several gestures, but their accuracy was not very well. It was not really clear how they classified it and so the product failed and the company is no more (P2)

Challenges in Working with EMG

Participants further commented on the challenges they experienced with EMG as a modality. First, electrode location was critical in achieving consistent results and not always straightforward:

There it was like not really clear, "Where do I have to put these two electrodes and where do I have to put the ground electrodes? What does this actually mean? Does this has an influence where I put the ground electrode?" (P2)

Second, the lack of generalizability over multiple persons that is inherent in EMG was difficult to address when prototyping with multiple users:

Overall, what worked very well is that when you put it approximately at the same place, it worked very well again, but only if you put it on the same person. (P2)

Having access to six channels also meant dealing with a lot of cables for the electrodes. Interviewees remarked that this could place a heavy burden on users.

I don't know if you used all of these six channels, but it could be quite heavy on the participant. (P4)

6.2.9 Discussion

From an initial set of requirements, we further refined and addressed challenges for EMG-based prototyping and experimentation resulting in the final version of EMBody, contributing to **RQ4a**. Through various stages of evaluations, we confirmed that EMBody meets the requirements. However, our work also highlights ongoing challenging in EMG-based interaction.

Mobility and Data Acquisition

EMBody's goal is to provide a *mobile* platform, allowing for easy *data acquisition* of muscle activity via electromyograms. Throughout the workshops, we identified that the Bluetooth protocol used in an early version was unreliable and intractable, especially when working in groups and with multiple devices. Hence, the final version of EMBody relies on a WiFi connection using the UDP protocol. This constitutes a compromise between reliability, the necessary setup time and resources. While relying on WiFi for connectivity meant the necessity of increased power (Section 6.2.3), we found this to be negligible and could keep a small form factor. Our presented sample applications make use of this new prototype, showcasing that *mobile* scenarios, such as fitness exercises, are possible without obstructing the user.

An added benefit of switching to WiFi was the fact that we could realize a simple setup process via a captive portal, allowing users to easily configure the device during first use, without the need to flash the firmware of the microcontroller. Additionally, it simplified prototyping when working with multiple receiver applications. We confirmed the feasibility of this *data acquisition* setup and connectivity via WiFi in the presented sample applications.

Switching to a more powerful microcontroller also meant that we could increase the channel count to a maximum of six channels. We found that the original version using one EMG shield²⁸ per channel was very cumbersome to use when employing more than two channels. The final version of EMBody natively supports up to six channels for *data acquisition*, thus enabling simultaneous exploration of electrode configurations. The live view of EMBody's software application conveniently allows for visual debugging of these configurations as confirmed by our sample applications and interviews.

Abstraction Layers through Modular Structure

One major objective during EMBody's development process was to make it accessible to a wide range of user groups, including novices as well as experts in physiological sensing, but also tailoring to different professions, such as designers, developers and engineers. We realized this objective using a modular structure encapsulating EMBody's data-centric processing pipeline (Section 6.2.4). The software grants experts a high degree of control over how data is processed and interpreted. Likewise, novices are aided by the exploration-centered workflow (Section 6.2.2), guiding them throughout the creation process, while hiding technical complexities in EMBody's base version. Extensive documentation, including electrode setup and best practices, is provided²⁹. We first informally verified this procedure during the workshops by guiding the students, confirming its viability. We later successfully instantiated the workflow in our sample applications and confirmed in the interviews that it was comprehensible and easy to follow. Moreover, the workflow closely draws from standard workflows when working with physiological sensing. Hence, experts felt immediately at home and quickly identified extensions points suitable for customization.

²⁸<https://www.olimex.com/Products/Duino/Shields/SHEILD-EKG-EMG/open-source-hardware>

²⁹ Available at <https://github.com/HCUM/embody>.

Towards More Accessible EMG Input for HCI

Our toolkit effectively lowers entry barriers for researchers and designers to begin exploring EMG-based input. Our evaluation of EMBody highlighted several challenges for future EMG input systems to further support developing interactive systems.

EMG measurements require placing electrodes on muscles and connecting electrode measurements to measurement units. We observed that researchers using EMBody were eager to experiment with multiple muscles and, consequently, measurement channels. This offers the opportunity of sensing complex movements, effectively increasing the fidelity of the motions that an EMG system can detect. However, a high number of channels results in a high number of cables to be connected, which may be cumbersome. Thus, future EMG toolkits should include advanced cable management. Despite cable-based solutions posing certain problems, cables are still the technology of choice for HCI prototyping. While textile wearable electrodes are being researched, they are often muscle-specific and may not support diverse users, e.g. [229]. Wireless electrodes would require individual power sources which would increase their mass.

Another finding from the evaluation of our toolkit is the fact that EMG measurements are highly person-dependent. As a consequence, EMG systems require individual calibration and detection is based on values specific to the user. EMBody includes calibration routines, but the need for explicit calibration does increase the complexity of interacting with prototypes which use EMG input. We envision that future EMG tools for HCI researchers should explore if implicit calibration methods can be used. This could be achieved by integrating calibration in tasks. There is a need for developing methods similar to ad hoc calibration in eye tracking, e.g. [158].

Finally, developing and evaluating EMBody enabled us to observe how designing EMG input was part of an interaction design process. One of the overarching ideas behind our toolkit was enabling designers to focus on the nature of the interaction technique they were designing and emphasize the limitations of the sensing modality. While we did observe that EMBody eliminated initial barriers to using EMG, EMG input still produces additional constraints in the design process. Limiting input to the muscles monitored or the need to place electrodes are likely to have a significant impact on how a design team develops an interactive artefact which uses EMG. Future work should address this challenge and study how designers can consider EMG as a input modality and be implicitly aware of the EMG design space without investing time in extensive EMG prototyping.

6.2.10 Conclusion and Outlook

In this chapter, we introduced EMBody—a data-centric toolkit for rapid prototyping and experimentation with EMG. We provided details of the design of the toolkit and information on how to access the open-source resources needed to build it. We also illustrated the utility of EMBody by reporting on workshops with students, presenting two systems that make use of the toolkit and a final evaluation through expert interviews.

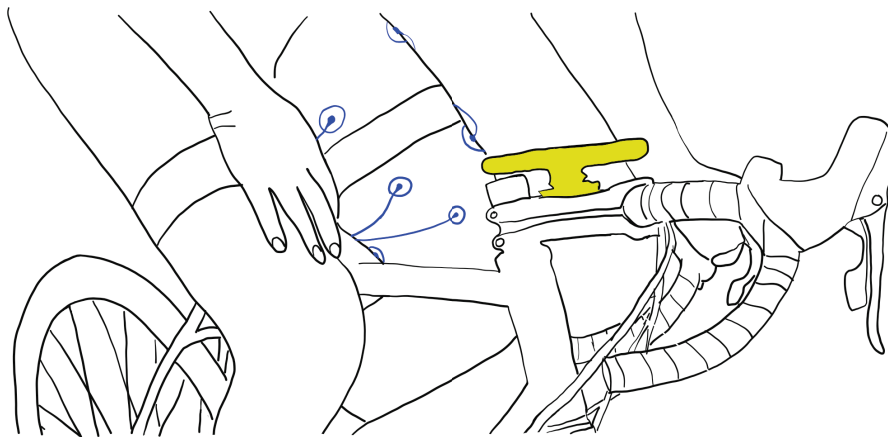


Figure 6.15: A speculative future usage scenario for EMBody. A cyclist monitors their quadriceps activity (using electrodes, marked in blue) during a bike ride on their smartphone (yellow). We envision that EMBody will foster experimentation with EMG.

For the domain of proficiency-aware systems, it was important for us to lower entry barriers for practitioners into EMG-based interactive systems. We believe this to be a vital step towards the applicability of EMG for future proficiency-aware systems.

We concluded that the EMBody toolkit can help practitioners focus on designing the interface and feedback, reducing the need to troubleshoot data acquisition and interpretation. Additionally, experts appreciated its modular structure and data processing pipeline, confirming that EMBody successfully tailored to the needs of a wider audience. This provides valuable insights with regard to its applicability in proficiency-aware systems

In contrast to off-the-shelf products, EMBody provides a full exploration-centered workflow from data acquisition through calibration and model training for live predictions. We envision that further iterations of our toolkit can be driven by the community, enabling access to electromyography for less technology-proficient practitioners, while allowing experts to benefit from improved algorithms. This would enable end-users to experiment with EMG for understanding their own bodies (see Figure 6.15 for an example scenario). In combination with recent advances in sensing technologies [217, 122], we hope that our work can help establish EMG as a key modality for future embodied interaction, including the domain of proficiency-aware systems.

6.3 Chapter Summary

Investigating how to facilitate proficiency awareness in interactive systems has been a two-stage process. While we have evaluated how to design for proficiency awareness in Chapter 5 (RQ3), this chapter has introduced our toolkit approach to implementing proficiency-

aware systems (**RQ4**). With EMBody, we highlighted how important ready access to physiological data is for developing and advancing future EMG-based systems.

However, this observation is not limited to EMG as a modality. We believe that toolkits and proper abstraction can have far-reaching implications for the acceptability of novel modalities. Especially those that are largely unknown to potential users. During the course of this thesis, we have realized that most participants did not know about EMG before. Yet, introducing them to the modality was straightforward and often led to a quick understanding of its capabilities. This also shows that users are the experts when it comes to understanding their own bodily signals.

We envision that further development on toolkits can help substantiate our concept of proficiency-aware systems. We have already shown that this vision is not limited to physiological measurements but is applicable for a variety of user data. For example, web-based toolkits could easily allow developers to create engaging websites that allow users to further their proficiency. In this regard, we feel that toolkits pave the way towards actual applications for the end user.

CONCLUSION AND FUTURE WORK

Chapter 7

Conclusion and Future Work

In this thesis, we introduced the concept of proficiency-aware systems to promote a class of systems that stimulate users and encourage personal growth. Proficiency-aware systems can be a means to create engaging experiences [250], where users can fully reflect on their proficiency and their use of the system, striving for excellence.

Proficiency-aware systems may specifically be designed and engineered towards supporting the user in understanding their own proficiency. This work argues that while the rise of *intelligence* in interactive systems supports task efficiency, it is rarely employed for the user's personal growth. Instead, reasons for adaptation towards the user are mostly performance-oriented, opening up the risk for *de-skilling*, where users increasingly rely on interactive systems and neglect their own proficiency progression. In this thesis, we highlighted that most interactive systems have the inherent potential to draw on the idea of proficiency awareness and encourage users in advancing their proficiency during the interaction with the system.

However, teaching users is challenging and no commonly understood and validated process explains how humans acquire new skills [277, 171], that could readily be applied to interactive systems. Often, learning is highly task-specific and governed by the individual strategies of the user. Especially in the early stages of a learning process, users tend to be overconfident in their proficiency [258], creating an illusion of knowledge. Consequently, it can be challenging to find the right balance between encouraging proficiency improvements or supporting the user through performance-oriented adaptation.

The findings of this thesis include, most importantly, the fact that when users are informed about their proficiency, they can reflect on their performance. Such a user-driven interpretation is a key element of proficiency-aware systems as users make necessary connections between their task approach and communicated data by themselves (cf. Chapter 2). Thus, we can potentially create systems that actively engage users in furthering their proficiency and scale well with a variety of tasks, as illustrated in this thesis. Often a basic representation is sufficient for users to connect this information to their task approach and reflect on it (cf. writing task in Section 5.2). Despite the challenging premise — creating systems that *teach* users — it is rather simple to design interactive systems with proficiency awareness in

mind. Providing the users with the means to reflect on their data is often already sufficient (cf. FitBack in Section 5.1).

In this regard, this thesis first investigated engineering challenges of proficiency-aware systems in Part I in five task domains (**RQ1**, **RQ2**). To do so, we employed two sensing modalities, gaze and electromyography (EMG), to recognize users' cognitive and physical abilities. This information was ultimately used as an estimation of their proficiency. In combination with our toolkit EMBody (**RQ4**), we contribute a set of **engineering guidelines** for proficiency-aware systems (Section 7.2) to aid further development. In Part II, we researched how this estimation of proficiency can be used to facilitate proficiency awareness in interactive systems (**RQ3**, **RQ4**). We used EMG again as a modality to provide users with insights into their bodily signals, but also showed design qualities of simple metrics for proficiency awareness in everyday computer work. We summarized these findings in a set of **design guidelines** for future proficiency-aware systems.

7.1 Designing Proficiency-Aware Systems

We first summarize our findings on designing proficiency awareness, which directly address **RQ3**. While most of these originate from Chapter 5, contributions in other chapters informed and refined this final set.

RQ3

What are design criteria for proficiency awareness in interactive systems?

Proficiency Feedback Makes Performances Tangible For The User

Providing proficiency feedback allows the users to reflect on their performance. It supports them in becoming aware of their own proficiency and connecting it to their current performance. Here, proficiency is potentially something the user can identify with, as it can be broken down into simpler concepts than actual task performance. Recall the editor example from Chapter 2. A possible task performance metric could be the number of copies sold. Yet, many factors influence this metric, such as the popularity of the magazine, current relevance of the topic, or even vacation cycles of potential buyers, apart from the actual quality of the article, which might be of interest to the editor. While it is hard to provide this particular metric, there are almost always good proxies for proficiency (cf. Section 5.2) that can support users in making their performance tangible. Thus, **finding suitable proficiency proxies is a key design task for proficiency-aware systems**.

At the same time, users need to spend resources on improving their proficiency. While intrusive feedback can be beneficial for task performance, it often inhibits proficiency improvement, as necessary resources are not available. Choosing the right feedback type and

granularity is a trade-off between primary task performance and potential proficiency benefits for the user. Thus, designing proficiency awareness includes considering how to increase productivity through proficiency feedback, and contributing ways to do so in a user-friendly and effective manner for the user to reflect on their proficiency. As outlined above, these proficiency metrics do not necessarily have to be performance-oriented, as long as the user knows how to connect them to the task. In other words, **instead of developing highly customized systems with delicate algorithms, the user is included in the process and provides the final interpretation of available data.**

Consequently, if any proxy for a user's proficiency can be acquired with ease, it should always be available as feedback. Users are unlikely to object to this (cf. Sections 4.2 and 5.2) and are most often able to make the connection to the current task, even if the signal is complex (cf. EMG). In this thesis, we have provided guidelines for several tasks on how, when, and how often to add feedback. We envision that systems can guide users to preferable task completion strategies optimized for their proficiency level.

It remains essential to deliver feedback related to the task domain, which can be exploited by the user. Carefully selecting the right proficiency feedback allows the user to concentrate on other aspects of the task and allows developers to easily integrate proficiency aspects, potentially generalizing them for multiple systems.

Consider Ethical Implications For The Individual

Our work also showed that even subtle differences in how computer-based assessments of proficiency scores are presented could significantly affect the user. The score might not always be used to enhance performance but could also judge particular individuals' abilities. Regardless of how it was derived, the proficiency score stays a form of computer judgment to be endured by the individual, especially since computer-based assessments historically have had negative connotations [72]. While this thesis highlights the benefits of proficiency awareness to render tasks easier to perform whilst improving user experience, we must emphasize that this can be abused to limit the user's autonomy. When machines are telling us what to do, designers must carefully outweigh potential adverse effects on the personal autonomy [32] of users. An aggressive system might intimidate users, compelling them to certain actions. A loss of autonomy can potentially lead to "ubiquitous human computing" [342], what Zittrain [342] dubbed "harvesters of human mindpower". Consequently, **users must always be given the opportunity to opt out of proficiency assessment, allowing them to make their own judgments about the morality of their work** [342]. Ethical pitfalls can be avoided if ethics-oriented design methods are used early in the design process, e.g., adversary design fictions [17], exposing the system's potential for malicious intent.

Offer Multiple Feedback Options

It is a good idea to offer a selection of feedback options, varying in type and granularity. If possible, task-specific feedback, such as relating to muscle power in fitness tasks, is to be favored over showing the raw EMG signal. For fully observable task domains, i.e., where

the difficulty of the task can be governed by the system (cf. EMGuitar in Section 4.2), the system should **offer a choice of basic, more abstract feedback and detailed feedback to allow for a wider user audience**. Similarly, designers of tutoring systems can separate their feedback visualizations from the estimated proficiency as provided by the system.

Depending on who governs the proficiency-aware system, the frequency of feedback needs to be adjusted accordingly. Here, gamification elements have proven beneficial to remedy annoying feedback if the frequency cannot be chosen freely. However, elements should be aligned with the task at hand. Playful feedback for a serious task is likely to alienate users.

Design With Social Context In Mind

Social interactions dictate how we approach tasks and interactions with systems. On the one side, proficiency-aware systems can exploit these social interactions to create incentives for users to become better. On the other side, proficiency-aware systems might only unfold their full potential in the right social context. Especially when systems substitute coaches, a tandem approach combines the strengths of both coach and system. During the course of this work, we have presented various systems that benefit from such a social relationship, such as our musical tutoring systems EMGuitar and EMPiano, but also FitBack has shown strong potential, not only complementing training but increasing its effectiveness. Consequently, **it can be beneficial to consider the social context of a proficiency-aware system and its potential users**.

While social interactions can be constructive for the effectiveness of proficiency-aware systems, they also imply the necessity to design for multiple stakeholders of the system. Controlling the system's agency (i.e., the one who is in control of the adaptation) is straightforward if only one user is involved (cf. EMGuitar in Section 4.2), as the user chooses whether or not to give up control and trust in the system's adaptation. However, if multiple stakeholders are present, interests might not necessarily align.

In a coaching scenario (cf. FitBack in Section 5.1), the coach gains a tangible tool to highlight wrong exercise form while the student gains objective truth, increasing their trust in the coach's advice. Here, interests align, and a symbiotic relationship can be achieved. In our text production use case (cf. Section 5.2), this might not be the case. The chief editor and newspaper owner are potentially interested in the proficiency of their editors as well. Even the developer could make use of the information to improve the system. This creates an ethical dilemma in who should be given access to the user's proficiency. As the principal nature of a proficiency-aware system aims to improve the user's proficiency, this should take priority. To rule out malicious intents, design methods as outlined previously can be applied.

7.2 Engineering Proficiency-Aware Systems

To complement our guidelines on how to design proficiency awareness, we present a set of engineering guidelines as well. These guidelines directly address **RQ1**, **RQ2**, and **RQ4**. In

this thesis, we advanced the understanding of how to create interactive systems that make use of proficiency to adapt the interaction to and with the user (**RQ1, RQ2**). Our toolkit approach to implementing proficiency-aware systems highlights how complex modalities for proficiency estimation can be made available to a wider audience (**RQ4**).

RQ1, RQ2, RQ4

What are suitable metrics inferred from a user's gaze for proficiency estimation?

How accurately can we estimate proficiency from a user's muscle activity?

How can we support a wide audience in realizing proficiency-aware systems?

Gaze As Ubiquitous Predictor

During the research conducted in this thesis, we found that **gaze can be a powerful modality to estimate proficiency for proficiency-aware systems but is subject to task constraints**, such as the required granularity of proficiency levels and the complexity of the visual stimulus.

If complexity and demands for granularity are low, gaze can provide robust proficiency estimation across users, even for a short interaction duration. While these premises might seem impractical at first, there exists a variety of tasks that only require binary proficiency estimation often correlated with ad hoc interaction, such as reading a bus schedule in a foreign language or buying a public transport ticket in an unknown city. Most often, these are the scenarios where a proficiency-aware system can rarely provide extensive proficiency feedback but might highlight useful shortcuts to separating relevant from irrelevant information for the user.

If fine-grained detection is needed in a complex task, gaze metrics that generalize across users are usually not sufficient. Complex visual stimuli justify person-dependent algorithms to ensure accurate proficiency estimation. Added motoric aspects further impede accurate classification, often calling for prolonged gaze recordings. Auxiliary sources of context can support estimation in this scenario, such as combining achieved learning progress in a language tutoring system with the current gaze data to identify difficult words for the user.

Gaze can be ambiguous; a property that should always be respected when realizing gaze-based proficiency-aware systems. Personal strategies that are not necessarily related to proficiency can govern the user's viewing patterns. Nevertheless, domain experts could exploit recognized task-solving strategies to identify teaching methods.

Electromyography for Fine Motor Tasks

In contrast to gaze, **electromyography has high distinctiveness**, i.e., the recorded muscle activity correlates well with the executed movement. **Yet, it suffers from poor general-**

ization across users. Often EMG-based systems require prior calibration before usage and have extensive setup procedures.

Nevertheless, detecting fine-grained motor tasks is easily possible with high-grade hardware. Here, EMG offers high accuracy when it comes to estimating correct posture and movement forms. In this thesis, we have confirmed technical requirements, such as electrode count and placement, derived from related work for posture-based (cf. EMGuitar), as well as for motion-based (cf. FitBack, EMPiano) task recognition.

While low-cost devices for EMG are available, their signal quality only allows recognizing coarse motor tasks. However, the setup is simplified, and with EMBody we provide a toolkit that enables users to prototype and experiment with EMG. This is especially important to make EMG a viable candidate for proficiency estimation as the need for calibration prevents most ad hoc systems. Generating awareness through easy access for a wide audience is vital for future EMG research and possibly adopting implicit calibration mechanisms from other domains.

Start With Simple Metrics

In one particular use case (cf. Section 5.2), we have highlighted that it is not necessary to employ complex sensing modalities to communicate proficiency to the user. **Proficiency-aware systems are not limited to physiological computing.** Like context-aware systems, they are just as effective where computing systems assist us in performing everyday tasks.

Especially when there is no agreed-on concept for proficiency for the given task domain, simple metrics that contribute to a user's proficiency are effective and support users in understanding their proficiency. Often, systems already collect user data for adaptation. It is worthwhile to consider how this data can be used to communicate proficiency to the user, allowing an existing system to become proficiency-aware without the need for extensive modification.

In general, this approach holds true for sensing-based systems as well. Readily deployable modalities can already provide sufficient information for the system about the user's proficiency, such as a gaze-based system for reading proficiency instead of measuring the user's brain activity to estimate their understanding of a text. Similarly, a user's typing proficiency should preferably be analyzed first via linguistics and keystrokes before applying other measures. The same principle applies to calculating metrics for a given modality. This thesis has shown that **there is no need for complex algorithms that provide elaborate visualizations of collected sensing data.** Users are good at recognizing and interpreting their own data. Consequently, we suggest starting simple before introducing the limitations of complex metrics.

Toolkits For Ubiquitous Technologies Are Enablers

Being able to estimate a user's proficiency is the backbone of every proficiency-aware system. Yet, toolkits that support researchers in prototyping novel systems are only sparsely

available. Especially for complex modalities, we advocate the necessity of readily available toolkits that lower the entry barrier for less technically-versed practitioners. **We need to give researchers and designers the opportunity to experiment with novel and meaningful sensing modalities** to advance future proficiency-aware systems.

We believe that this holds true even for metrics derived from everyday computer work, such as linguistics or keystrokes. Having ready access to those metrics — ethical implications kept in mind — can provide users with insights unbeknownst to them. In other words, knowing that one’s average typing speed when writing research papers is higher in the evening hours can guide users in changing their work schedule accordingly. To create an interest for such insights among users and developers, we also need toolkits for software applications that draw attention to the potential of proficiency awareness in everyday computer work.

7.3 Limitations

The aforementioned guidelines for proficiency-aware systems and the majority of the work conducted throughout this thesis are based on empirical studies conducted in a constrained lab environment. Most notably, this allowed us to sustain a high internal validity of the conducted studies. This was essential to draw implications with high confidence for the individual use cases. Nevertheless, for a concept like proficiency, where external influences can have a high impact on the system’s perceived proficiency of the user, it remains to be investigated how these factors affect the user experience for proficiency-aware systems. An in-the-wild investigation could provide additional guidelines for production environments.

Especially mid-term factors might pose a challenge for proficiency-aware systems that rely on the user’s prolonged interaction. Subtle changes in measurements, e.g., due to fatigue (slower eye movements, muscle exhaustion), can throw off classification algorithms. If these factors are likely to occur during the interaction, developers should consider modelling them separately to allow robust proficiency estimation. Long-term proficiency changes are generally less of a problem for the system but more an issue of proper calibration towards the user.

Apart from external factors that dynamically influence a user’s proficiency, internal aspects such as user interest or cognitive load may hinder proper adaptation. This is most often the case for ambiguous modalities, as these are not only a proxy for proficiency but might indicate other cognitive processes as well. In particular, gaze-based systems are affected through the interference of user interest [147] or increased cognitive load [289]. Yet, cognitive load can be an excellent proxy for proficiency (cf. Section 3.2), as in, users that are less proficient need to spend more cognitive resources. Consequently, when relying on possibly ambiguous modalities, it is worthwhile to consider what is being measured.

Proficiency-aware systems share a close connection to intelligent tutoring systems, albeit on a theoretical level. While the latter actively employ learning strategies, proficiency-aware

systems simply provide the means for a user to advance their skills. This user-driven interpretation allows users to employ their own strategies and prevents the system from pursuing a potentially detrimental strategy. However, future systems might want to actively engage learning strategies if the task domains allow for it. Thus, it remains an open challenge to design proficiency-aware systems that actively teach users.

7.4 Future Work

In this thesis, we have presented design and engineering guidelines to create proficiency-aware systems, allowing the detection of proficiency as well as facilitating proficiency awareness for interactive systems. While the previous section has already teased open challenges and opportunities for future work, we highlight specific starting points and time frames in this section.

The recent rise of artificial intelligence through neural networks and deep learning [235] approaches has broad implications for the estimation of concepts such as proficiency from user collected data. Throughout this thesis, we have relied on robust and established machine learning algorithms to classify data. Here, deep learning approaches can boost detection, especially for proficiency-aware systems that are based on user-bound and prolonged interaction. Potential projects based on this work include detection of office workplace activities via EMG, e.g., through wearables or integrated into office appliances or gaze-contingent piano tutoring systems that automatically focus on difficult segments of a piece.

Proficiency awareness is not only limited to single computing systems. Especially with the proliferation of IoT devices and infrastructure, estimating proficiency can be decentralized and propagated to registered receivers. The idea of proficiency-aware environments sees the user in the focus, and their proficiency is published for other trusted devices that know how to make sense of it. Here, this thesis provides the starting point to identify technical and design challenges that need to be investigated and addressed in a larger research framework.

While the two previously introduced projects address mostly single-user scenarios of proficiency-aware systems, this next project addresses the research of proficiency-aware computing systems. As outlined previously, the ethical implications of such systems need to be carefully considered and evaluated, both for involved users and third parties. Moreover, a realization for everyday computing systems requires extensive research of suitable feedback visualizations and modalities, applied algorithms, and user interaction. A final outcome shows how to create proficiency-aware systems for everyday computer work that encourage users to advance their own proficiency. This research's main challenge lies in the fact that the system should be seamlessly integrated into already existing workflows. Consequently, this project is best suited as work in an interdisciplinary research cluster as it requires expertise in numerous domains, including psychology, human-computer interaction, visualization, software development, algorithms and possibly sociology.

7.5 Final Remarks

At the onset of this thesis, we argued that *de-skilling* might become a severe problem for interactive systems. Consequently, we presented the concept of proficiency-aware systems: systems that (1) support the user in accomplishing their task and (2) facilitate an understanding of their proficiency. The research in this thesis aims to prevent de-skilling from happening by providing users with engaging experiences that foster personal growth.

But can de-skilling also be a positive outcome for the user and, in particular, for associated systems? The assistance provided by interactive systems allows us to take a holistic approach to solve a given task. The availability of freed-up resources is a necessity to acquire novel skills that surpass what a proficiency-aware system might measure. Recalling the example from the introduction in this thesis: is it really so bad that the younger generation struggles in reading maps? Do we not gain more from technology that allows us to navigate without having to concentrate on paper maps? While it is a nice skill to have, interest will always drive people to acquire *old* skills.

The potential to acquire novel skills should never be inhibited by technology but rather encouraged and promoted. While it may seem that the concept of proficiency-aware systems contradicts this idea at first glance, skill discovery and acquisition of novel skills happen most often at expert levels. Features of tools might only become *available* once mastery is reached, such as exploiting a blade's profile for cutting; an implicit form of proficiency awareness. From this perspective, proficiency-aware systems are essential in supporting users to become experts, but also letting them get there themselves rather than helping them stay along a beaten track.

LISTINGS AND BIBLIOGRAPHY

List of Acronyms

ABD	average blink duration
AFD	average fixation duration
AI	artificial intelligence
ANOVA	analysis of variance
EDA	electrodermal activity
EEG	electroencephalography
EMG	electromyography
EOG	electrooculography
fNIRS	functional near-infrared spectroscopy
GSR	galvanic skin response
HCI	human-computer interaction
IMU	inertial measurement unit
LSL	lab streaming layer
NASA-TLX	NASA task load index
PAS	proficiency-aware system
RFR	refixation ratio
RMS	root mean square
RQ	research question
SVM	support vector machine
TBR	total blink ratio

LIST OF FIGURES

2.1	Depiction of a proficiency-aware system. Relevant skills are inferred and supplied to the system. Adaptation and estimation mutually influence each other until an equilibrium is reached. Additionally, a proficiency-aware system communicates proficiency back to the user.	16
3.1	Apparatus showing LCD monitor with attached eye-tracking device (left) and participant during the study (right).	34
3.2	Violin plots (width normalized per group) showing the distribution of the average fixation duration (AFD) grouped by proficiency according to CEFR [51]. Density plots depict the sample distribution over the proficiency groups.	38
3.3	Violin plots (width normalized per group) showing the distribution of the refixation ratio (RFR) grouped by proficiency according to CEFR [51]. Density plots depict the sample distribution over the proficiency groups.	38
3.4	Violin plots (width normalized per group) showing the distribution of the average blink duration (ABD) grouped by proficiency according to CEFR [51]. Density plots depict the sample distribution over the proficiency groups.	39
3.5	Violin plots (width normalized per group) showing the distribution of the total blink ratio (TBR) grouped by proficiency according to CEFR [51]. Density plots depict the sample distribution over the proficiency groups.	39
3.6	Top three levels of the built decision tree using J48 with standard parameters. ABD and AFD are given in milliseconds; TBR and RFR are given as ratios. Leaf nodes (rectangles) indicate predicted class (0 for NP; 1 for P) and show (correctly classified/misclassified) instances. Values are rounded for visual clarity.	42
3.7	Complete decision tree using J48 and heavy pruning. ABD and AFD are given in milliseconds; RFR is given as ratio. Leaf nodes (rectangles) indicate predicted class (0 for NP; 1 for P) and show (correctly classified/misclassified) instances. Values are rounded for visual clarity.	43
3.8	Accuracy, true and false positive rate of proficient (class P: A1 to C2) and non-proficient (class NP: X) class given different maximum observation times of recorded gaze data.	45
3.9	Accuracy, true and false positive rate of proficient (class P: A1 to C2) and non-proficient (class NP: X) class given different sampling rates of the underlying gaze dataset (logarithmic scale).	46
3.10	The heatmaps of two exemplary participants showcasing different eye movement patterns when solving the same programming task. The choice for a heatmap visualization and omission of AOIs boxes was done for visual clarity.	52
3.11	Study setup including keyboard, loudspeaker, and monitor with attached eye tracker.	55

3.12	Confusion matrices depicting the accuracy of the final prediction algorithm for difficult passages grouped by three sensitivity levels (from left to right: highest to lowest). There is a total of $N = 1230$ samples with 386 difficult parts.	56
4.1	Electrode configuration with two rings around the forearm each consisting of five electrodes.	67
4.2	Schematic depiction of our apparatus.	67
4.3	Depiction of used and discarded (red) signal parts. One box corresponds to one bar whereas each dotted line marks a new window.	68
4.4	Confusion matrices with respect to the classifier’s performance on the test-data. The total number of instances per class varies over the three participants because of the random starting position in the pattern and the train-test split.	71
4.5	Average F1 scores of the classifier given window size. All values are calculated using the RMS and RMS ratios feature groups.	72
4.6	Maximum average F1 score for different EMG channel configurations. Features include RMS and their ratios with a 250ms window size.	73
4.7	Study procedure for the evaluation of EMGuitar.	76
4.8	Boxplots illustrating the results of the first two questionnaires.	78
4.9	Possible feedback visualization for chord grips and chord changes. Colors range from green (flawless play) to red (faulty play).	78
4.10	Schematic depiction of a second-order hidden Markov model (HMM) for chord estimates. The most-likely sequence is calculated based on the estimated transition and observation probabilities.	79
4.11	FitBack: Hardware prototype to record EMG data.	84
4.12	FitBack visualizations showing the same EMG data. Abstract representations <i>Bars</i> (left) and <i>Circles</i> (middle), and the detailed representation <i>Lines</i> (right).	85
4.13	Electrode placement on biceps (left) and deltoid (middle) muscles. Reference electrodes can be seen close to the scapula. On the right side: participant during the study.	87
4.14	Confusion matrices (associated truth labels on the y axis) between coach—system (left), experts—coach (middle), experts—system (right). Note that percentages are based on within-class instances (per row), as class distribution is highly skewed.	87
4.15	Expert ratings given the shown visualization and sound feedback. Significant differences are marked with *.	88
5.1	Example EMG data for a correct (left) and incorrect (right) bicep curl. The incorrect example clearly shows excessive deltoid use (orange) compared to bicep use (blue). Time (in samples) on the x-axis; muscle activation (as RMS values) on the y-axis.	99

5.2	FitBack visualizations showing the same EMG data. Abstract representations <i>Bars</i> (left) and <i>Circles</i> (middle), and the detailed representation <i>Lines</i> (right).	100
5.3	Mean scores for Questions Q1-Q5; assessing ranking and perception of feedback (see Table 5.1).	102
5.4	Participants during the exploration phase performing different exercises. . .	107
5.5	Example view of our proficiency-aware web form, showing the progress bar and ranking condition.	118
5.6	Deviation of readability score over time. The red-shaded corridor shows standard deviation over the filtered data (red line). Condition <i>Revision</i> (middle) additionally shows timings for the revision prompt (black vertical lines; thick line being the mean). Note that all trials have been rescaled to the median answer time and start at $t = 100$ s.	123
5.7	Ratings for questions Q1a-Q5a (see Table 5.6). Significant differences are marked with *.	124
5.8	Gamification implementation of our proficiency-aware web form. The view on top shows the emoji condition. The bottom view shows the progress bar and social comparison statement. Both gamification views depict the continuous condition.	127
5.9	Final readability scores given both factors: <i>Feedback Type</i> and <i>Gamification Type</i> . Scores for <i>Feedback Type (Revision - Continuous)</i> are significantly different. The shaded green area marks the target zone for the readability score (60 to 70).	129
5.10	SIMS (left) and IMI (right) subscale ratings categorized by <i>Gamification Type</i> . Significant differences are marked with *.	130
5.11	Deviation of readability score over time. Red-shaded corridor shows standard deviation over the filtered data (red line). Conditions based on revision feedback (top row) additionally show timings for the revision prompt (black vertical lines; thick line being the mean). Note that all trials have been rescaled to the median answer time and start at $t = 100$ s.	131
5.12	Ratings for questions Q1b-Q7b (see Table 5.6). Q3b, Q5b, Q6b and Q7b are significant for <i>Feedback Type</i> . Q7b is additionally significant for <i>Gamification Type</i> ; marked with *.	132
5.13	Mean scores for Question A1-A4; assessing perceived accuracy and intuitiveness of both conditions (see Table 5.10). Error bars show standard error. Significant differences between conditions are marked with *.	140
6.1	Workflow when designing with EMBody.	151
6.2	EMBody works with a variety of electrodes. Here, they are affixed with straps to the forearm. First, the ground electrode is placed (left) on a location with little muscle fiber, e.g. close to the elbow. Afterwards, the sensing electrodes are placed on the muscle belly (right).	151

6.3	Checking the EMG signal in the live view using different views of the EMG data. Filtered signal on the left; generated RMS features on the right. See Section 6.2.4 for algorithm details.	152
6.4	EMBody calibrating for a gesture labeled <i>LEFT</i> . Corresponding gesture by the user (left arm) on the right.	153
6.5	Live view with live classification. Note that the classification is also streamed via UDP.	154
6.6	Diagram of the EMG system: The system senses muscular activity with electrodes. These signals are amplified and sent in UDP packets over WiFi. . . .	155
6.7	EMBody’s hardware prototype showing the six audio jacks on the front to connect electrodes and the microcontroller in the middle. The three amplifiers (each supporting two channels) are placed underneath the microcontroller. A lid (removed in this picture) is also provided.	156
6.8	Classification pipeline within <code>ClassificationManager</code> . Note the different flows for the calibration and live classification phases.	157
6.9	Unipolar measurement with six sensing electrodes (white, two on backside) and one reference electrode (blue). Ground electrode (black) serves for noise reduction only.	162
6.10	Hands-on session during one of the workshops.	164
6.11	Study stimulus (left) and electrode placement (right) for EMG input controls.	165
6.12	Averaged responses for our questionnaires (ease of use, fatigue, feeling of control) given modality. Visual-analog scale from 0 to 100.	166
6.13	Average Deviation (L2 norm) in pixels given different speeds and modalities.	166
6.14	Task instructions as seen by participants. Final target position was given in rotation angles (left) and visualized as a picture (right). The platform was actuated through a robot for the EMG- and pedal-based condition.	167
6.15	A speculative future usage scenario for EMBody. A cyclist monitors their quadriceps activity (using electrodes, marked in blue) during a bike ride on their smartphone (yellow). We envision that EMBody will foster experimentation with EMG.	174

LIST OF TABLES

1.1	Overview of the research questions that guided the work in this thesis.	7
3.1	Example questions used in our study in four different languages.	33
3.2	Languages used in this study and respective proficiencies present in our participant base.	34
3.3	Accuracies and F-Measures for selected classifiers.	41
3.4	Classifier statistics on J48 given proficient users (class P) and non-proficient users (class NP).	42
3.5	Classifier statistics on cost-sensitive J48 (standard parameters) given proficient users (class P) and non-proficient users (class NP).	43
3.6	Classifier statistics on J48 with different class distribution: highly-proficient (class HP: C1 and C2) and less-proficient (class LP: X, A1, A2, B1 and B2) users. Levels according to [51].	44
3.7	Example programs used in the study given the calculated difficulty.	51
3.8	Statistical results of paired t-tests and Wilcoxon signed-rank tests (bold: significant).	55
4.1	Weighted class-averaged F1 score for given feature groups (averaged over all participants).	70
4.2	Averaged processor time required to compute the respective feature groups for a five minute recording on a MacBookPro ⁵	71
4.3	Inter-person classifier performance using the RMS and RMS ratios feature groups. Training was done on the other two subjects' data.	72
4.4	Questionnaire (7-point Likert) for both conditions (automatic, manual) adapted from Yuksel et al. [338].	76
4.5	Questionnaire (5-point Likert and free text) exclusively for the automatic condition.	77
4.6	NASA-TLX, UMUX and flow experience score per condition.	89
5.1	Additional questions after each feedback type. Q1 from <i>extremely confusing</i> to <i>easy to understand</i> ; Q2-Q5 from <i>not at all</i> to <i>very much</i> ; all visual analog scale (0 to 100).	101
5.2	Information about the participants in our study. The sports professionals specialized in different sports and had different levels of experience.	107
5.3	Study designs and respective conditions.	117
5.4	Writing self-assessment questions (adapted from [46]). Five-item Likert scale from <i>incompetent</i> to <i>very competent</i>	121

5.5	Writing quality assessment by participants for their proposal text (adapted from [137]). Ten-item Likert scale from <i>very strongly disagree</i> to <i>very strongly agree</i>	121
5.6	Additional questions for conditions <i>Revision</i> and <i>Continuous</i> ; from <i>strongly disagree</i> to <i>strongly agree</i> ; all visual analog scale (0 to 100).	122
5.7	TCT, NASA-TLX, final FRES and its deviation from the target zone (60 to 70).	122
5.8	Additional questions for all conditions; from <i>strongly disagree</i> to <i>strongly agree</i> ; all visual analog scale (0 to 100).	128
5.9	Distribution of collected data entries over all conditions in Study II.	129
5.10	Additional questions for both conditions (<i>PW</i> , <i>EMG</i>) and both tasks (<i>Notesheet</i> , <i>Improvisation</i>); all visual analog scale (0 to 100).	139
6.1	Participant profiles in the interviews, including their areas of expertise as well as prototyping experience (7-item Likert scale) and frequency.	168

Bibliography

- [1] E. D. Adrian and D. W. Bronk. The discharge of impulses in motor nerve fibres. *J Physiol*, 67(2):i3–151, March 1929. ISSN 0022-3751. URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1402822/>.
- [2] Christine Ahlheim. Die gefährliche Mischung aus Halbwissen und Selbstüberschätzung. *Deutsche Apotheker Zeitung*, September 2020.
- [3] Seoyoung Ahn, Conor Kelton, Aruna Balasubramanian, and Greg Zelinsky. Towards Predicting Reading Comprehension From Gaze Behavior. In *ACM Symposium on Eye Tracking Research and Applications*, ETRA '20 Short Papers, pages 1–5, New York, NY, USA, June 2020. Association for Computing Machinery. ISBN 978-1-4503-7134-6. doi: 10.1145/3379156.3391335.
- [4] E. Aiello, D. H. Gates, B. L. Patrilli, K. D. Cairns, M. Meister, E. A. Clancy, and P. Bonato. Visual EMG Biofeedback to Improve Ankle Function in Hemiparetic Gait. *Conf Proc IEEE Eng Med Biol Soc*, 2005:7703–7706, 2005. ISSN 1557-170X. doi: 10.1109/IEMBS.2005.1616297.
- [5] O.S. Akinola, B.O. Akinkunmi, and T.S. Alo. A data mining model for predicting computer programming proficiency of computer science undergraduate students. *African Journal of Computing and ICT*, 2012.
- [6] Florian Alt, Alireza Sahami Shirazi, Albrecht Schmidt, and Julian Mennenöh. Increasing the User’s Attention on the Web: Using Implicit Interaction Based on Gaze Behavior to Tailor Content. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*, NordiCHI '12, pages 544–553, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1482-4. doi: 10.1145/2399016.2399099.
- [7] Christoph Amma, Thomas Krings, Jonas Böer, and Tanja Schultz. Advancing Muscle-Computer Interfaces with High-Density Electromyography. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 929–938, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3145-6. doi: 10.1145/2702123.2702501.

-
- [8] Center for Disease Control and Prevention. Early Release of Selected Estimates Based on Data From the 2018 National Health Interview Survey. May 2019. URL <https://www.cdc.gov/nchs/nhis/releases/released201905.htm>.
- [9] Fraser Anderson, Tovi Grossman, Justin Matejka, and George Fitzmaurice. YouMove: Enhancing movement training with an augmented reality mirror. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, pages 311–320, New York, NY, USA, October 2013. Association for Computing Machinery. ISBN 978-1-4503-2268-3. doi: 10.1145/2501988.2502045.
- [10] B. Arslan, A. Brouse, J. Castet, R. Lehembre, C. Simon, J. J. Filatriau, and Q. Noirhomme. A Real Time Music Synthesis Environment Driven with Biological Signals. In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, volume 2, pages II–II, May 2006. doi: 10.1109/ICASSP.2006.1660557.
- [11] Amid Ayobi, Paul Marshall, and Anna L Cox. Trackly : A customisable and pictorial self-tracking app to support agency in multiple sclerosis self-care. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1–15, 2020.
- [12] Nikola Banovic, Fanny Chevalier, Tovi Grossman, and George Fitzmaurice. Triggering triggers and burying barriers to customizing software. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 2717–2726, New York, NY, USA, May 2012. Association for Computing Machinery. ISBN 978-1-4503-1015-4. doi: 10.1145/2207676.2208666.
- [13] Gabriel Barata, Sandra Gama, Joaquim Jorge, and Daniel Gonçalves. Improving participation and learning with gamification. In *Proceedings of the First International Conference on Gameful Design, Research, and Applications*, Gamification '13, pages 10–17, New York, NY, USA, 2013. Association for Computing Machinery. ISBN 978-1-4503-2815-9. doi: 10.1145/2583008.2583010.
- [14] D. T. Barry, K. E. Gordon, and G. G. Hinton. Acoustic and surface EMG diagnosis of pediatric muscle disease. *Muscle Nerve*, 13(4):286–290, April 1990. ISSN 0148-639X. doi: 10.1002/mus.880130403.
- [15] Eric P.S. Baumer. Reflective informatics: Conceptual dimensions for designing technologies of reflection. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 585–594, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3145-6. doi: 10.1145/2702123.2702234.
- [16] Eric P.S. Baumer, Vera Khovanskaya, Mark Matthews, Lindsay Reynolds, Victoria Schwanda Sosik, and Geri Gay. Reviewing reflection: On the use of reflection in interactive system design. In *Proceedings of the 2014 Conference on Designing Interactive Systems*, DIS '14, pages 93–102, New York, NY, USA, June 2014. Association for Computing Machinery. ISBN 978-1-4503-2902-6. doi: 10.1145/2598510.2598598.

-
- [17] Eric P.S. Baumer, Timothy Berrill, Sarah C. Botwinick, Jonathan L. Gonzales, Kevin Ho, Allison Kundrik, Luke Kwon, Tim LaRowe, Chanh P. Nguyen, Fredy Ramirez, Peter Schaedler, William Ulrich, Amber Wallace, Yuchen Wan, and Benjamin Weinfeld. What Would *You* Do? Design Fiction and Ethics. In *Proceedings of the 2018 ACM Conference on Supporting Groupwork*, GROUP '18, pages 244–256, New York, NY, USA, January 2018. Association for Computing Machinery. ISBN 978-1-4503-5562-9. doi: 10.1145/3148330.3149405.
- [18] Djamila Romaiissa Beddiar, Brahim Nini, Mohammad Sabokrou, and Abdenour Hadid. Vision-based human activity recognition: A survey. *Multimedia Tools and Applications*, 79(41):30509–30555, November 2020. ISSN 1573-7721. doi: 10.1007/s11042-020-09004-3.
- [19] Roman Bednarik and Markku Tukiainen. An eye-tracking methodology for characterizing program comprehension processes. In *Proceedings of the 2006 Symposium on Eye Tracking Research & Applications*, pages 125–132. ACM, 2006.
- [20] Marco Belloli, Stefano Giappino, Fabio Robustelli, and Claudio Somaschini. Drafting Effect in Cycling: Investigation by Wind Tunnel Tests. *Procedia Engineering*, 147: 38–43, 2016. ISSN 18777058. doi: 10.1016/j.proeng.2016.06.186.
- [21] Steve Benford, Holger Schnädelbach, Boriana Koleva, Rob Anastasi, Chris Greenhalgh, Tom Rodden, Tom Rodden, Jonathan Green, Ahmed Ghali, Tony Pridmore, Bill Gaver, Andy Boucher, Brendan Walker, Sarah Pennington, Albrecht Schmidt, Hans Gellersen, Anthony Steed, and Anthony Steed. Expected, Sensed, and Desired: A Framework for Designing Sensing-based Interaction. *ACM Trans. Comput.-Hum. Interact.*, 12(1):3–30, March 2005. ISSN 1073-0516. doi: 10.1145/1057237.1057239.
- [22] David Benyon. Adaptive systems: A solution to usability problems. *User Model User-Adap Inter*, 3(1):65–87, March 1993. ISSN 1573-1391. doi: 10.1007/BF01099425.
- [23] Ralf Biedert, Georg Buscher, Sven Schwarz, Jörn Hees, and Andreas Dengel. Text 2.0. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10, pages 4003–4008, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-930-5. doi: 10.1145/1753846.1754093.
- [24] Ann Blandford, Dominic Furniss, and Stephann Makri. Qualitative HCI research: Going behind the scenes. *Synthesis lectures on human-centered informatics*, 9(1): 1–115, 2016.
- [25] Susanne Bødker. Applying activity theory to video analysis: How to make sense of video data in human-computer interaction. In *Context and Consciousness: Activity Theory and Human-Computer Interaction*, pages 147–174. Massachusetts Institute of Technology, USA, December 1995. ISBN 978-0-262-14058-4.

-
- [26] W. Boeckh-Behrens and Buskies W. *Fitness-Krafttraining*. Rowohlt-Taschenbuch-Verlag, 2000.
- [27] Anne-Claire Bourland, Peter Gorman, Jess McIntosh, and Asier Marzo. Project Telepathy: Targeted Verbal Communication Using 3D Beamforming Speakers and Facial Electromyography. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, pages 1508–1515, New York, NY, USA, 2017. ACM. ISBN 978-1-4503-4656-6. doi: 10.1145/3027063.3053129.
- [28] Robin Brewer, Meredith Ringel Morris, and Anne Marie Piper. "Why Would Anybody Do This?": Understanding older adults' motivations and challenges in crowd work. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 2246–2257, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 978-1-4503-3362-7. doi: 10.1145/2858036.2858198.
- [29] Joseph D. Bronzino. *The Biomedical Engineering Handbook, Third Edition*. Taylor & Francis, April 2006. ISBN 978-0-8493-2124-5.
- [30] Andreas Bulling, Daniel Roggen, and Gerhard Tröster. It's in Your Eyes: Towards Context-awareness and Mobile HCI Using Wearable EOG Goggles. In *Proceedings of the 10th International Conference on Ubiquitous Computing*, UbiComp '08, pages 84–93, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-136-1. doi: 10.1145/1409635.1409647.
- [31] Andreas Bulling, Jamie A. Ward, Hans Gellersen, and Gerhard Tröster. Eye Movement Analysis for Activity Recognition. In *Proceedings of the 11th International Conference on Ubiquitous Computing*, UbiComp '09, pages 41–50, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-431-7. doi: 10.1145/1620545.1620552.
- [32] Sarah Buss and Andrea Westlund. Personal Autonomy. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2018 edition, 2018. URL <https://plato.stanford.edu/archives/spr2018/entries/personal-autonomy/>.
- [33] Guy Thomas Buswell. *How People Look at Pictures*. Chicago: University of Chicago Press, 1935.
- [34] D. F. Butcher and W. A. Muth. Predicting performance in an introductory computer science course. *Communications of the ACM*, 1985.
- [35] Evan A Byrne and Raja Parasuraman. Psychophysiology and adaptive automation. *Biological Psychology*, 42(3):249–268, February 1996. ISSN 0301-0511. doi: 10.1016/0301-0511(95)05161-9.
- [36] J. Carroll and M. Rosson. Paradox of the active user. In *Interfacing Thought: Cognitive Aspects of Human-Computer Interaction*. The MIT Press, 1987.

- [37] Diogo V. Carvalho, Eduardo M. Pereira, and Jaime S. Cardoso. Machine learning interpretability: A survey on methods and metrics. *Electronics*, 8(8), 2019. ISSN 2079-9292. doi: 10.3390/electronics8080832.
- [38] Jared E. Cechanowicz, Carl Gutwin, Scott Bateman, Regan Mandryk, and Ian Stavness. Improving player balancing in racing games. In *Proceedings of the First ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '14, pages 47–56, New York, NY, USA, October 2014. Association for Computing Machinery. ISBN 978-1-4503-3014-5. doi: 10.1145/2658537.2658701.
- [39] Center for Self-Determination Theory. Intrinsic Motivation Inventory. URL <http://selfdeterminationtheory.org/intrinsic-motivation-inventory/>.
- [40] Anargyros Chatzitofis, Nicholas Vretos, Dimitrios Zarpalas, and Petros Daras. Three-dimensional monitoring of weightlifting for computer assisted training. In *Proceedings of the Virtual Reality International Conference: Laval Virtual, VRIC '13*, pages 1–7, New York, NY, USA, March 2013. Association for Computing Machinery. ISBN 978-1-4503-1875-4. doi: 10.1145/2466816.2466820.
- [41] Shiwei Cheng, Zhiqiang Sun, Lingyun Sun, Kirsten Yee, and Anind K. Dey. Gaze-Based Annotations for Reading Comprehension. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 1569–1572, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3145-6. doi: 10.1145/2702123.2702271.
- [42] Tim Chester. 'Sedated by software': No one knows how to read maps anymore, experts say. *Mashable*, May 2015. URL <https://mashable.com/2015/05/01/map-reading-knowledge/?europa=true>.
- [43] Pei-Yu (Peggy) Chi and Yang Li. Weave: Scripting cross-device wearable interaction. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 3923–3932, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 978-1-4503-3145-6. doi: 10.1145/2702123.2702451.
- [44] Christopher Gilbert and Donald Moss. Biofeedback and Biological Monitoring. In *Handbook of Mind-Body Medicine for Primary Care*. Thousand Oaks, CA: SAGE Publications, Inc, 2003. URL <http://dx.doi.org/10.4135/9781452232607.n8>.
- [45] Andy Cockburn, Carl Gutwin, Joey Scarr, and Sylvain Malacria. Supporting Novice to Expert Transitions in User Interfaces. *ACM Comput. Surv.*, 47(2):31:1–31:36, November 2014. ISSN 0360-0300. doi: 10.1145/2659796.
- [46] Richard T. Cole, Larry A. Hembroff, and Andrew D. Corner. National Assessment of the Perceived Writing Skills of Entry-Level PR Practitioners. *Journalism & Mass Communication Educator*, 64(1):9–26, March 2009. ISSN 1077-6958, 2161-4326. doi: 10.1177/107769580906400102.

-
- [47] Sunny Consolvo, David W. McDonald, and James A. Landay. Theory-driven design strategies for technologies that support behavior change in everyday life. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 405–414, New York, NY, USA, April 2009. Association for Computing Machinery. ISBN 978-1-60558-246-7. doi: 10.1145/1518701.1518766.
- [48] Amy Cook, Jessica Hammer, Salma Elsayed-Ali, and Steven Dow. How Guiding Questions Facilitate Feedback Exchange in Project-Based Learning. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 1–12, New York, NY, USA, May 2019. Association for Computing Machinery. ISBN 978-1-4503-5970-2. doi: 10.1145/3290605.3300368.
- [49] Enrico Costanza, Samuel A. Inverso, and Rebecca Allen. Toward Subtle Intimate Interfaces for Mobile Devices Using an EMG Controller. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, pages 481–489, New York, NY, USA, 2005. ACM. ISBN 978-1-58113-998-3. doi: 10.1145/1054972.1055039.
- [50] Enrico Costanza, Samuel A. Inverso, Rebecca Allen, and Pattie Maes. Intimate Interfaces in Action: Assessing the Usability and Subtlety of Emg-based Motionless Gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 819–828, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-593-9. doi: 10.1145/1240624.1240747.
- [51] Council of Europe. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Applied Linguistics Non Series. Cambridge University Press, 2001. ISBN 978-0-521-00531-9.
- [52] CD Cozean, William S Pease, and SL Hubbell. Biofeedback and functional electric stimulation in stroke rehabilitation. *Archives of physical medicine and rehabilitation*, 69(6):401–405, 1988.
- [53] Ronald V. Croce. The effects of EMG biofeedback on strength acquisition. *Biofeedback and Self-Regulation*, 11(4):299–310, December 1986. ISSN 1573-3270. doi: 10.1007/BF01000166.
- [54] Palle Dahlstedt, Patrik Karlsson, Katarina Widell, and Tony Blomdahl. YouHero Making an Expressive Concert Instrument from the GuitarHero Controller. In *NIME*, pages 403–406, 2014.
- [55] Claudia Daudén Roquet and Corina Sas. Body matters: Exploration of the human body as a resource for the design of technologies for meditation. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, pages 533–546. Association for Computing Machinery, New York, NY, USA, 2020. ISBN 978-1-4503-6974-9. URL <https://doi.org/10.1145/3357236.3395499>.

-
- [56] F. Delavier. *Strength Training Anatomy*. Anatomy Series. Human Kinetics, 2010. ISBN 978-0-7360-9226-5.
- [57] John N Demos. *Getting Started with Neurofeedback*. WW Norton & Company, 2005.
- [58] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. From game design elements to gamefulness: Defining "Gamification". In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, MindTrek '11, pages 9–15, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0816-8. doi: 10.1145/2181037.2181040.
- [59] Guy Deutscher. *The Unfolding of Language: An Evolutionary Tour of Mankind's Greatest Invention*. Henry Holt and Company, 2006. ISBN 978-1-4668-3783-6.
- [60] Anind K. Dey. *Providing Architectural Support for Building Context-Aware Applications*. PhD thesis, Georgia Institute of Technology, 2000.
- [61] Anind K Dey. Understanding and Using Context. *Personal and ubiquitous computing*, page 4, 2001.
- [62] Felix Dietz, Matthias Hoppe, Jakob Karolus, Pawel W. Wozniak, Albrecht Schmidt, and Tonja Machulla. VRsneaky: Stepping into an audible virtual world with gait-aware auditory feedback. In Regina Bernhaupt, Florian 'Floyd' Mueller, David Verweij, Josh Andres, Joanna McGrenere, Andy Cockburn, Ignacio Avellino, Alix Goguey, Pernille Bjøn, Shengdong Zhao, Briane Paul Samson, and Rafal Kocielnik, editors, *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, CHI 2020, Honolulu, HI, USA, April 25-30, 2020*, pages 1–4. ACM, 2020. doi: 10.1145/3334480.3383168.
- [63] Sidney D'Mello, Andrew Olney, Claire Williams, and Patrick Hays. Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of Human-Computer Studies*, 70(5):377–398, May 2012. ISSN 1071-5819. doi: 10.1016/j.ijhcs.2012.01.004.
- [64] Marco Donnarumma, Baptiste Caramiaux, and Atau Tanaka. Muscular Interactions Combining EMG and MMG sensing for musical practice. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Seoul, 2013. KAIST. URL <http://research.gold.ac.uk/10634/>.
- [65] Paul Dourish. Embodied interaction: Exploring the foundations of a new approach to HCI. 1999.
- [66] Steven Dow, Anand Kulkarni, Scott Klemmer, and Björn Hartmann. Shepherding the crowd yields better work. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work, CSCW '12*, pages 1013–1022, New York, NY, USA, February 2012. Association for Computing Machinery. ISBN 978-1-4503-1086-4. doi: 10.1145/2145204.2145355.

-
- [67] Heiko Drewes and Albrecht Schmidt. Interacting with the Computer Using Gaze Gestures. In *Proceedings of the 11th IFIP TC 13 International Conference on Human-Computer Interaction - Volume Part II*, INTERACT'07, pages 475–488, Berlin, Heidelberg, 2007. Springer-Verlag. ISBN 978-3-540-74799-4. doi: http://dx.doi.org/10.1007/978-3-540-74800-7_43.
- [68] Andrew T. Duchowski. *Eye Tracking Methodology - Theory and Practice, Third Edition*. Springer, 2017. ISBN 978-3-319-57881-1. doi: 10.1007/978-3-319-57883-5.
- [69] Andrew T. Duchowski. Gaze-based interaction: A 30 year retrospective. *Computers & Graphics*, 73:59–69, June 2018. ISSN 0097-8493. doi: 10.1016/j.cag.2018.04.002.
- [70] Andrew T. Duchowski, Nathan Cournia, and Hunter Murphy. Gaze-contingent displays: A review. *Cyberpsychol Behav*, 7(6):621–634, December 2004. ISSN 1094-9313. doi: 10.1089/cpb.2004.7.621.
- [71] Don Samitha Elvitigala, Denys J.C. Matthies, L  ic David, Chamod Weerasinghe, and Suranga Nanayakkara. GymSoles: Improving Squats and Dead-Lifts by Visualizing the User's Center of Pressure. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 1–12, New York, NY, USA, May 2019. Association for Computing Machinery. ISBN 978-1-4503-5970-2. doi: 10.1145/3290605.3300404.
- [72] Harold P. Erdman and Sharon W. Foster. Ethical Issues in the Use of Computer-Based Assessment. *Computers in Human Services*, 3(1-2):71–87, August 1988. ISSN 0740-445X. doi: 10.1300/J407v03n01_06.
- [73] Eurostat. How much do Europeans exercise? 2017. URL <https://ec.europa.eu/eurostat/web/products-eurostat-news/-/DDN-20170302-1>.
- [74] Kate C. Ewing, Stephen H. Fairclough, and Kiel Gilleade. Evaluation of an Adaptive Game that Uses EEG Measures Validated during the Design Process as Inputs to a Biocybernetic Loop. *Front Hum Neurosci*, 10, May 2016. ISSN 1662-5161. doi: 10.3389/fnhum.2016.00223.
- [75] Stephen H. Fairclough. Fundamentals of physiological computing. *Interact Comput*, 21(1-2):133–145, January 2009. ISSN 0953-5438. doi: 10.1016/j.intcom.2008.10.011.
- [76] Chloe Fan, Jodi Forlizzi, and Anind K. Dey. A spark of activity: Exploring informative art as visualization for physical activity. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, UbiComp '12, pages 81–84, New York, NY, USA, September 2012. Association for Computing Machinery. ISBN 978-1-4503-1224-0. doi: 10.1145/2370216.2370229.

- [77] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. LIBLINEAR - a library for large linear classification. 2008. URL <http://www.csie.ntu.edu.tw/~cjlin/liblinear/>.
- [78] Janet Feigenspan, Christian Kästner, Jörg Liebig, Sven Apel, and Stefan Hanenberg. Measuring programming experience. In *Program Comprehension (ICPC), 2012 IEEE 20th International Conference On*, pages 73–82. IEEE, 2012.
- [79] Matt Feusner and Brian Lukoff. Testing for Statistically Significant Differences Between Groups of Scan Patterns. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications, ETRA '08*, pages 43–46, New York, NY, USA, 2008. ACM. ISBN 978-1-59593-982-1. doi: 10.1145/1344471.1344481.
- [80] Andy Field and Graham Hole. *How to Design and Report Experiments*. Sage Publications, London, 2003.
- [81] Leah Findlater and Joanna McGrenere. A comparison of static, adaptive, and adaptable menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '04*, pages 89–96, New York, NY, USA, April 2004. Association for Computing Machinery. ISBN 978-1-58113-702-6. doi: 10.1145/985692.985704.
- [82] Leah Findlater and Joanna McGrenere. Beyond performance: Feature awareness in personalized interfaces. *International Journal of Human-Computer Studies*, 68(3): 121–137, March 2010. ISSN 1071-5819. doi: 10.1016/j.ijhcs.2009.10.002.
- [83] Kraig Finstad. The usability metric for user experience. *Interacting with Computers*, 22(5):323–327, May 2010. ISSN 0953-5438. doi: 10.1016/j.intcom.2010.04.004.
- [84] Gerhard Fischer. User Modeling in Human–Computer Interaction. *User Modeling and User-Adapted Interaction*, 11(1):65–86, March 2001. ISSN 1573-1391. doi: 10.1023/A:1011145532042.
- [85] Rudolph Flesch. A new readability yardstick. *Journal of Applied Psychology*, 32(3): 221–233, 1948. ISSN 1939-1854, 0021-9010. doi: 10.1037/h0057532.
- [86] Tom Foulsham, Joey T. Cheng, Jessica L. Tracy, Joseph Henrich, and Alan Kingstone. Gaze allocation in a dynamic situation: Effects of social status and speaking. *Cognition*, 117(3):319–331, December 2010. ISSN 1873-7838. doi: 10.1016/j.cognition.2010.09.003.
- [87] Gordon Fraser. Gamification of software testing. In *2017 IEEE/ACM 12th International Workshop on Automation of Software Testing (AST)*, pages 2–7. IEEE, 2017. doi: 10.1109/AST.2017.20.
- [88] Krzysztof Z. Gajos, Katherine Everitt, Desney S. Tan, Mary Czerwinski, and Daniel S. Weld. Predictability and accuracy in adaptive user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '08*,

-
- pages 1271–1274, New York, NY, USA, April 2008. Association for Computing Machinery. ISBN 978-1-60558-011-1. doi: 10.1145/1357054.1357252.
- [89] Tong Gao, Jessica R. Hullman, Eytan Adar, Brent Hecht, and Nicholas Diakopoulos. NewsViews: An automated pipeline for creating custom geovisualizations for news. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 3005–3014, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 978-1-4503-2473-1. doi: 10.1145/2556288.2557228.
- [90] Nadia Garcia-Hernandez, Karen Garza-Martinez, and Vicente Parra-Vega. Electromyography Biofeedback Exergames to Enhance Grip Strength and Motivation. *Games Health J*, 7(1):75–82, February 2018. ISSN 2161-7856. doi: 10.1089/g4h.2017.0054.
- [91] Stavros Garzonis, Simon Jones, Tim Jay, and Eamonn O'Neill. Auditory icon and earcon mobile service notifications: Intuitiveness, learnability, memorability and preference. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 1513–1522, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 978-1-60558-246-7. doi: 10.1145/1518701.1518932.
- [92] L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, and L. Kagal. Explaining explanations: An overview of interpretability of machine learning. In *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, pages 80–89, 2018. doi: 10.1109/DSAA.2018.00018.
- [93] Christiane Glatz, Stas S. Krupenia, Heinrich H. Bülthoff, and Lewis L. Chuang. Use the right sound for the right job: Verbal commands and auditory icons for a task-management system favor different information processes in the brain. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13. Association for Computing Machinery, New York, NY, USA, 2018. ISBN 978-1-4503-5620-6. URL <https://doi.org/10.1145/3173574.3174046>.
- [94] Joseph H Goldberg and Xerxes P Kotval. Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics*, 24(6):631–645, October 1999. ISSN 0169-8141. doi: 10.1016/S0169-8141(98)00068-7.
- [95] Saul Greenberg and Chester Fitchett. Phidgets: Easy development of physical interfaces through physical widgets. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology*, UIST '01, pages 209–218, Orlando, Florida, November 2001. Association for Computing Machinery. ISBN 978-1-58113-438-4. doi: 10.1145/502348.502388.
- [96] Saul Greenberg and Ian H. Witten. Adaptive personalized interfaces—A question of viability. *Behaviour & Information Technology*, 4(1):31–45, 1985. doi: 10.1080/01449298508901785.

-
- [97] Miriam Greis, Jakob Karolus, Hendrik Schuff, Paweł W. Woźniak, and Niels Henze. Detecting uncertain input using physiological sensing and behavioral measurements. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia, MUM '17*, pages 299–304, Stuttgart, Germany, November 2017. Association for Computing Machinery. ISBN 978-1-4503-5378-6. doi: 10.1145/3152832.3152859.
- [98] Tobias Grosse-Puppendahl, Yannick Berghoefer, Andreas Braun, Raphael Wimmer, and Arjan Kuijper. OpenCapSense: A rapid prototyping toolkit for pervasive interaction using capacitive sensing. In *2013 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 152–159, March 2013. doi: 10.1109/PerCom.2013.6526726.
- [99] Frédéric Guay, Robert J. Vallerand, and Céline Blanchard. On the Assessment of Situational Intrinsic and Extrinsic Motivation: The Situational Motivation Scale (SIMS). *Motivation and Emotion*, 24(3):175–213, September 2000. ISSN 1573-6644. doi: 10.1023/A:1005614228250.
- [100] Brian Guenter, Mark Finch, Steven Drucker, Desney Tan, and John Snyder. Foveated 3D Graphics. *ACM Trans. Graph.*, 31(6):164:1–164:10, November 2012. ISSN 0730-0301. doi: 10.1145/2366145.2366183.
- [101] Guy Brook-Hart and Simon Haines. *Complete Advanced*. Cambridge University Press, second edition edition, 2014.
- [102] Juan Haladjian. The Wearables Development Toolkit: An Integrated Development Environment for Activity Recognition Applications. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 3(4):134:1–134:26, December 2019. doi: 10.1145/3369813.
- [103] Juan Haladjian, Katharina Bredies, and Bernd Brügge. Interactex: An integrated development environment for smart textiles. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers, ISWC '16*, pages 8–15, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 978-1-4503-4460-9. doi: 10.1145/2971763.2971776.
- [104] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The WEKA data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18, November 2009. ISSN 1931-0145. doi: 10.1145/1656274.1656278.
- [105] Perttu Hämäläinen. Interactive video mirrors for sports training. In *Proceedings of the Third Nordic Conference on Human-Computer Interaction*, pages 199–202. Association for Computing Machinery, New York, NY, USA, October 2004. ISBN 978-1-58113-857-3. URL <https://doi.org/10.1145/1028014.1028044>.

-
- [106] Juho Hamari and Jonna Koivisto. Measuring flow in gamification: Dispositional flow scale-2. *Computers in Human Behavior*, 40:133–143, 2014. doi: 10.1016/j.chb.2014.07.048.
- [107] Juho Hamari, Jonna Koivisto, Harri Sarsa, et al. Does gamification Work?-A literature review of empirical studies on gamification. In *HICSS*, volume 14, pages 3025–3034, 2014.
- [108] Nils Y. Hammerla and Thomas Plötz. Let’s (Not) Stick Together: Pairwise Similarity Biases Cross-validation in Activity Recognition. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp ’15*, pages 1041–1051, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3574-4. doi: 10.1145/2750858.2807551.
- [109] Faizan Haque, Mathieu Nancel, and Daniel Vogel. Myopoint: Pointing and Clicking Using Forearm Mounted Electromyography and Inertial Motion Sensors. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI ’15*, pages 3653–3656, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3145-6. doi: 10.1145/2702123.2702133.
- [110] Sandra G. Hart. Nasa-task load index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9):904–908, 2006. doi: 10.1177/154193120605000909.
- [111] Sandra G. Hart and Lowell E. Staveland. Development of NASA-TLX (task load index): Results of empirical and theoretical research. In Peter A. Hancock and Najmedin Meshkati, editors, *Human Mental Workload*, volume 52 of *Advances in Psychology*, pages 139–183. North-Holland, 1988. doi: 10.1016/S0166-4115(08)62386-9.
- [112] Björn Hartmann, Leith Abdulla, Manas Mittal, and Scott R. Klemmer. Authoring sensor-based interactions by demonstration with direct manipulation and pattern recognition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’07*, pages 145–154, San Jose, California, USA, April 2007. Association for Computing Machinery. ISBN 978-1-59593-593-9. doi: 10.1145/1240624.1240646.
- [113] G. Hefftner and G. G. Jaros. The electromyogram (EMG) as a control signal for functional neuromuscular stimulation. II. Practical demonstration of the EMG signature discrimination system. *IEEE Transactions on Biomedical Engineering*, 35(4): 238–242, April 1988. ISSN 0018-9294. doi: 10.1109/10.1371.
- [114] G. Hefftner, W. Zucchini, and G. G. Jaros. The electromyogram (EMG) as a control signal for functional neuromuscular stimulation. I. Autoregressive modeling as a means of EMG signature discrimination. *IEEE Transactions on Biomedical Engineering*, 35(4):230–237, April 1988. ISSN 0018-9294. doi: 10.1109/10.1370.

-
- [115] Herbert A. Simon. Rational choice and the structure of the environment. *Psychological Review*, (63(2)), 1956. doi: 10.1037/h0042769.
- [116] Leonard A. Hermens and Jeffrey C. Shlimmer. A Machine-Learning Apprentice for the Completion of Repetitive Forms. *IEEE Expert: Intelligent Systems and Their Applications*, 9(1):28–33, February 1994. ISSN 0885-9000. doi: 10.1109/64.295135.
- [117] A. Hernandez Arieta, R. Katoh, H. Yokoi, and Y. Wenwei. Development of a Multi-DOF Electromyography Prosthetic System Using the Adaptive Joint Mechanism, 2006.
- [118] Lotta Hirvenkari, Johanna Ruusuvoori, Veli-Matti Saarinen, Maari Kivioja, Anssi Peräkylä, and Riitta Hari. Influence of Turn-Taking in a Two-Person Conversation on the Gaze of a Viewer. *PLoS One*, 8(8), August 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0071569.
- [119] Lars Erik Holmquist. Intelligence on tap: Artificial intelligence as a new design material. *Interactions*, 24(4):28–33, June 2017. ISSN 1072-5520. doi: 10.1145/3085571.
- [120] Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halzszka Jarodzka, and Joost van de Weijer. *Eye Tracking: A Comprehensive Guide to Methods and Measures*. OUP Oxford, September 2011. ISBN 978-0-19-162542-8.
- [121] A. Holtermann, P. J. Mork, L. L. Andersen, H. B. Olsen, and K. Sjøgaard. The use of EMG biofeedback for learning of selective activation of intra-muscular parts within the serratus anterior muscle: A novel approach for rehabilitation of scapular muscle imbalance. *J Electromyogr Kinesiol*, 20(2):359–365, April 2010. ISSN 1873-5711. doi: 10.1016/j.jelekin.2009.02.009.
- [122] Cedric Honnet, Hannah Perner-Wilson, Marc Teyssier, Bruno Fruchard, Jürgen Steimle, Ana C. Baptista, and Paul Strohmeier. PolySense: Augmenting Textiles with Electrical Functionality using In-Situ Polymerization. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pages 1–13, Honolulu, HI, USA, April 2020. Association for Computing Machinery. ISBN 978-1-4503-6708-0. doi: 10.1145/3313831.3376841.
- [123] Matthias Hoppe, Jakob Karolus, Felix Dietz, Pawel W. Wozniak, Albrecht Schmidt, and Tonja-Katrin Machulla. VRsneaky: Increasing presence in VR through gait-aware auditory feedback. In Stephen A. Brewster, Geraldine Fitzpatrick, Anna L. Cox, and Vassilis Kostakos, editors, *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI 2019, Glasgow, Scotland, UK, May 04-09, 2019*, page 546. ACM, 2019. doi: 10.1145/3290605.3300776.
- [124] Steven Houben and Nicolai Marquardt. WatchConnect: A toolkit for prototyping smartwatch-centric cross-device applications. In *Proceedings of the 33rd Annual*

-
- ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 1247–1256, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 978-1-4503-3145-6. doi: 10.1145/2702123.2702215.
- [125] Steven Houben, Connie Golsteijn, Sarah Gallacher, Rose Johnson, Saskia Bakker, Nicolai Marquardt, Licia Capra, and Yvonne Rogers. Physikit: Data engagement through physical ambient visualizations in the home. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 1608–1619, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 978-1-4503-3362-7. doi: 10.1145/2858036.2858059.
- [126] Dandan Huang, Melanie Tory, Bon Adriel Aseniero, Lyn Bartram, Scott Bateman, Sheelagh Carpendale, Anthony Tang, and Robert Woodbury. Personal Visualization and Personal Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics*, 21(3):420–433, March 2015. ISSN 2160-9306. doi: 10.1109/TVCG.2014.2359887.
- [127] Donny Huang, Xiaoyi Zhang, T. Scott Saponas, James Fogarty, and Shyamnath Gollakota. Leveraging Dual-Observable Input for Fine-Grained Thumb Interaction Using Forearm EMG. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, UIST '15, pages 523–528, New York, NY, USA, November 2015. Association for Computing Machinery. ISBN 978-1-4503-3779-3. doi: 10.1145/2807442.2807506.
- [128] Scott E. Hudson and Jennifer Mankoff. Rapid construction of functioning physical interfaces from cardboard, thumbtacks, tin foil and masking tape. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology*, UIST '06, pages 289–298, Montreux, Switzerland, October 2006. Association for Computing Machinery. ISBN 978-1-59593-313-3. doi: 10.1145/1166253.1166299.
- [129] Alice Hughes. Millions of millennials have never read a paper map, study claims. *Independent UK*, May 2019.
- [130] Wonil Hwang and Gavriel Salvendy. Number of people required for usability evaluation: The 10 ± 2 rule. *Communications of the ACM*, 53(5):130, May 2010. ISSN 00010782. doi: 10.1145/1735223.1735255.
- [131] Aulikki Hyrskykari, Päivi Majaranta, Antti Aaltonen, and Kari-Jouko Rähkä. Design Issues of iDICT: A Gaze-assisted Translation Aid. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, ETRA '00, pages 9–14, New York, NY, USA, 2000. ACM. ISBN 978-1-58113-280-9. doi: 10.1145/355017.355019.
- [132] Aulikki Hyrskykari, Päivi Majaranta, and Kari-Jouko Rähkä. Proactive response to eye movements. In *INTERACT*, volume 3, pages 129–136. IOS press Amsterdam, 2003.

- [133] Shamsi T. Iqbal and Eric Horvitz. Notifications and awareness: A field study of alert usage and preferences. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work, CSCW '10*, pages 27–30, New York, NY, USA, February 2010. Association for Computing Machinery. ISBN 978-1-60558-795-0. doi: 10.1145/1718918.1718926.
- [134] Robert J.K. Jacob and Keith S. Karn. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, 2003.
- [135] Lawrence R. James, Robert G. Demaree, and Gerrit Wolf. Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, 69(1):85–98, 1984. ISSN 1939-1854(Electronic),0021-9010(Print). doi: 10.1037/0021-9010.69.1.85.
- [136] Halszka Jarodzka, Kenneth Holmqvist, and Marcus Nyström. A Vector-based, Multi-dimensional Scanpath Similarity Measure. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications, ETRA '10*, pages 211–218, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-994-7. doi: 10.1145/1743666.1743718.
- [137] Adam C. Johnson, Joshua Wilson, and Rod D. Roscoe. College student perceptions of writing errors, text quality, and author characteristics. *Assessing Writing*, 34:72–87, October 2017. ISSN 10752935. doi: 10.1016/j.asw.2017.10.002.
- [138] David Johnson, Isabelle Dufour, Daniela Damian, and George Tzanetakis. Detecting Pianist Hand Posture Mistakes for Virtual Piano Tutoring. In *Proceedings of the International Computer Music Conference*, page 167, 2016.
- [139] Rose M.G. Johnson, Janet van der Linden, and Yvonne Rogers. MusicJacket: The Efficacy of Real-time Vibrotactile Feedback for Learning to Play the Violin. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems, CHI EA '10*, pages 3475–3480, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-930-5. doi: 10.1145/1753846.1754004.
- [140] Peter Ju, Leslie Pack Kaelbling, and Yoram Singer. State-based Classification of Finger Gestures from Electromyographic Signals. In *In ICML*, pages 439–446, 2000.
- [141] Marcel A. Just and Patricia A. Carpenter. A theory of reading: From eye fixations to comprehension. *Psychological review*, 87(4):329, 1980. URL <http://psycnet.apa.org/journals/rev/87/4/329/>.
- [142] Martin Kaltenbrunner and Ross Bencina. ReactIVision: A computer-vision framework for table-based tangible interaction. In *Proceedings of the 1st International Conference on Tangible and Embedded Interaction, TEI '07*, pages 69–74, New York, NY, USA, 2007. Association for Computing Machinery. ISBN 978-1-59593-619-6. doi: 10.1145/1226969.1226983.

-
- [143] Jakob Karolus and Albrecht Schmidt. Proficiency-Aware Systems: Adapting to the User's Skills and Expertise. In *Proceedings of the 7th ACM International Symposium on Pervasive Displays*, PerDis '18, pages 1–2, Munich, Germany, June 2018. Association for Computing Machinery. ISBN 978-1-4503-5765-4. doi: 10.1145/3205873.3210708.
- [144] Jakob Karolus and Paweł W. Woźniak. Proficiency-aware systems: Designing for user reflection in context-aware systems. *it - Information Technology*, 63(3):167–175, 2021. doi: doi:10.1515/itit-2020-0039.
- [145] Jakob Karolus, Paweł W. Wozniak, and Lewis L. Chuang. Towards using gaze properties to detect language proficiency. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction, Gothenburg, Sweden, October 23 - 27, 2016*, page 118. ACM, 2016. doi: 10.1145/2971485.2996753.
- [146] Jakob Karolus, Paweł W. Wozniak, Lewis L. Chuang, and Albrecht Schmidt. Robust Gaze Features for Enabling Language Proficiency Awareness. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 2998–3010, Denver, Colorado, USA, May 2017. Association for Computing Machinery. ISBN 978-1-4503-4655-9. doi: 10.1145/3025453.3025601.
- [147] Jakob Karolus, Patrick Dabbert, and Paweł W. Wozniak. I Know What You Want: Using Gaze Metrics to Predict Personal Interest. In *The 31st Annual ACM Symposium on User Interface Software and Technology Adjunct Proceedings*, UIST '18 Adjunct, pages 105–107, Berlin, Germany, October 2018. Association for Computing Machinery. ISBN 978-1-4503-5949-8. doi: 10.1145/3266037.3266116.
- [148] Jakob Karolus, Hendrik Schuff, Thomas Kosch, Paweł W. Wozniak, and Albrecht Schmidt. EMGuitar: Assisting Guitar Playing with Electromyography. In *Proceedings of the 2018 Designing Interactive Systems Conference*, DIS '18, pages 651–655, Hong Kong, China, June 2018. Association for Computing Machinery. ISBN 978-1-4503-5198-0. doi: 10.1145/3196709.3196803.
- [149] Jakob Karolus, Annika Kilian, Thomas Kosch, Albrecht Schmidt, and Paweł W. Wozniak. Hit the Thumb Jack! Using Electromyography to Augment the Piano Keyboard. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, DIS '20, pages 429–440, New York, NY, USA, July 2020. Association for Computing Machinery. ISBN 978-1-4503-6974-9. doi: 10.1145/3357236.3395500.
- [150] Jakob Karolus, Felix Bachmann, Thomas Kosch, Albrecht Schmidt, and Paweł W. Woźniak. Facilitating bodily insights using electromyography-based biofeedback during physical activity. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction*, MobileHCI '21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 978-1-4503-8328-8. doi: 10.1145/3447526.3472027.

- [151] Jakob Karolus, Francisco Kiss, Caroline Eckerth, Nicolas Viot, Felix Bachmann, Albrecht Schmidt, and Paweł W. Woźniak. EMBody: A data-centric toolkit for EMG-Based interface prototyping and experimentation. *Proc. ACM Hum.-Comput. Interac.*, 5(EICS):195:1–195:29, 2021. doi: 10.1145/3457142.
- [152] Paweł Kasprowski and Katarzyna Harezlak. Implicit calibration using predicted gaze targets. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*, ETRA '16, pages 245–248, New York, NY, USA, March 2016. Association for Computing Machinery. ISBN 978-1-4503-4125-7. doi: 10.1145/2857491.2857511.
- [153] Geoff Kaufman, Mary Flanagan, and Sukdith Punjasthitkul. Investigating the impact of 'emphasis frames' and social loafing on player motivation and performance in a crowdsourcing game. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 4122–4128, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 978-1-4503-3362-7. doi: 10.1145/2858036.2858588.
- [154] Harmanpreet Kaur, Alex C. Williams, Anne Loomis Thompson, Walter S. Lasecki, Shamsi T. Iqbal, and Jaime Teevan. Creating Better Action Plans for Writing Tasks via Vocabulary-Based Planning. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW):86:1–86:22, November 2018. doi: 10.1145/3274355.
- [155] Alexander Kent. A Profession Less Ordinary? Reflections on the Life, Death and Resurrection of Cartography. *The bulletin of the Society of University Cartographers. Society of University Cartographers*, 48:7–16, September 2014.
- [156] Frederic Kerber, Michael Puhl, and Antonio Krüger. User-independent real-time hand gesture recognition based on surface electromyography. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '17, pages 1–7, New York, NY, USA, September 2017. Association for Computing Machinery. ISBN 978-1-4503-5075-4. doi: 10.1145/3098279.3098553.
- [157] Dagmar Kern, Paul Marshall, and Albrecht Schmidt. Gazemarks: Gaze-based Visual Placeholders to Ease Attention Switching. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 2093–2102, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-929-9. doi: 10.1145/1753326.1753646.
- [158] Mohamed Khamis, Ozan Saltuk, Alina Hang, Katharina Stolz, Andreas Bulling, and Florian Alt. TextPursuits: Using Text for Pursuits-based Interaction and Calibration on Public Displays. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '16, pages 274–285, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-4461-6. doi: 10.1145/2971648.2971679.

-
- [159] Annika Kilian, Jakob Karolus, Thomas Kosch, Albrecht Schmidt, and Paweł W. Woźniak. EMPiano: Electromyographic pitch control on the piano keyboard. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '21 Extended Abstracts)*, May 8–13, 2021, Yokohama, Japan, page 4. ACM, 2021. doi: 10.1145/3411763.3451556.
- [160] Francisco Kiss, Paweł W. Woźniak, Felix Scheerer, Julia Dominiak, Andrzej Romanowski, and Albrecht Schmidt. Clairbuoyance: Improving Directional Perception for Swimmers. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 1–12, New York, NY, USA, May 2019. Association for Computing Machinery. ISBN 978-1-4503-5970-2. doi: 10.1145/3290605.3300467.
- [161] G. Klein, D. D. Woods, J. M. Bradshaw, R. R. Hoffman, and P. J. Feltovich. Ten challenges for making automation a "team player" in joint human-agent activity. *IEEE Intelligent Systems*, 19(6):91–95, November 2004. ISSN 1941-1294. doi: 10.1109/MIS.2004.74.
- [162] Sebastian Kleinschmager and Stefan Hanenberg. How to rate programming skills in programming experiments?: A preliminary, exploratory, study based on university marks, pretests, and self-estimation. In *Proceedings of the 3rd ACM SIGPLAN Workshop on Evaluation and Usability of Programming Languages and Tools*, pages 15–24. ACM, 2011.
- [163] R. Benjamin Knapp and Hugh S. Lusted. A Bioelectric Controller for Computer Music Applications. *Computer Music Journal*, 14(1):42–47, 1990. ISSN 0148-9267. doi: 10.2307/3680115.
- [164] Kristina Knaving, Paweł W Woźniak, Jasmin Niess, Romina Poguntke, Morten Fjeld, and Staffan Björk. Understanding grassroots sports gamification in the wild. In *Proceedings of the 10th Nordic Conference on Human-Computer Interaction*, pages 102–113. ACM, 2018. doi: 10.1145/3240167.3240220.
- [165] Thomas Kosch, Mariam Hassib, and Albrecht Schmidt. The brain matters: A 3D real-time visualization to examine brain source activation leveraging neurofeedback. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, pages 1570–1576, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-4082-3. doi: 10.1145/2851581.2892484.
- [166] Thomas Kosch, Romina Kettner, Markus Funk, and Albrecht Schmidt. Comparing tactile, auditory, and visual assembly error-feedback for workers with cognitive impairments. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '16, pages 53–60, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 978-1-4503-4124-0. doi: 10.1145/2982142.2982157.

- [167] Thomas Kosch, Markus Funk, Albrecht Schmidt, and Lewis L. Chuang. Identifying cognitive assistance with mobile electroencephalography: A case study with in-situ projections for manual assembly. *Proc. ACM Hum. Comput. Interact.*, 2:11:1–11:20, 2018. doi: 10.1145/3229093.
- [168] Thomas Kosch, Jakob Karolus, Havy Ha, and Albrecht Schmidt. Your skin resists: Exploring electrodermal activity as workload indicator during manual assembly. In José Ignacio Panach, Jean Vanderdonckt, and Oscar Pastor, editors, *Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems, EICS 2019, Valencia, Spain, June 18-21, 2019*, pages 8:1–8:5. ACM, 2019. doi: 10.1145/3319499.3328230.
- [169] Heli Koskimäki, Pekka Siirtola, and Juha Röning. MyoGym: Introducing an open gym data set for activity recognition collected using myo armband. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers, UbiComp '17*, pages 537–546, New York, NY, USA, September 2017. Association for Computing Machinery. ISBN 978-1-4503-5190-4. doi: 10.1145/3123024.3124400.
- [170] Christian Kothe, David Medine, Chadwick Boulay, Matthew Grivich, and Tristan Stenner. Lab streaming layer. URL <https://github.com/sccn/labstreaminglayer>.
- [171] John W. Krakauer and Reza Shadmehr. Consolidation of motor memory. *Trends Neurosci.*, 29(1):58–64, January 2006. ISSN 0166-2236. doi: 10.1016/j.tins.2005.10.003.
- [172] Justin Kruger and David Dunning. Unskilled and Unaware of It: How Difficulties in Recognizing One’s Own Incompetence Lead to Inflated Self-Assessments. *Journal of Personality and Social Psychology*, 77:1121–1134, 1999.
- [173] Kai Kunze, Katsutoshi Masai, Masahiko Inami, Ömer Sacakli, Marcus Liwicki, Andreas Dengel, Shoya Ishimaru, and Koichi Kise. Quantifying Reading Habits: Counting How Many Words You Read. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp '15*, pages 87–96, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3574-4. doi: 10.1145/2750858.2804278.
- [174] David Kurlander and Steven Feiner. A history-based macro by example system. In *Proceedings of the 5th Annual ACM Symposium on User Interface Software and Technology, UIST '92*, pages 99–106, New York, NY, USA, December 1992. Association for Computing Machinery. ISBN 978-0-89791-549-6. doi: 10.1145/142621.142633.
- [175] Ralph La Forge. Aligning mind and body: Exploring the disciplines of mindful exercise. *ACSM’s Health & Fitness Journal*, 9(5):7–14, 2005.

-
- [176] David M. Lane, H. Albert Napier, S. Camille Peres, and Aniko Sandor. Hidden Costs of Graphical User Interfaces: Failure to Make the Transition from Menus and Icon Toolbars to Keyboard Shortcuts. *International Journal of Human-Computer Interaction*, 18(2):133–144, May 2005. ISSN 1044-7318, 1532-7590. doi: 10.1207/s15327590ijhc1802_1.
- [177] Jonathan Lazar, Jinjuan Feng, and Harry Hochheiser. *Research Methods in Human-Computer Interaction, 2nd Edition*. Morgan Kaufmann, 2017. ISBN 978-0-12-805390-4. URL <http://www.sciencedirect.com/science/book/9780128053904>.
- [178] David Ledo, Fraser Anderson, Ryan Schmidt, Lora Oehlberg, Saul Greenberg, and Tovi Grossman. Pineal: Bringing passive objects to life with embedded mobile devices. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 2583–2593, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 978-1-4503-4655-9. doi: 10.1145/3025453.3025652.
- [179] David Ledo, Steven Houben, Jo Vermeulen, Nicolai Marquardt, Lora Oehlberg, and Saul Greenberg. Evaluation Strategies for HCI Toolkit Research. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 1–17, Montreal QC, Canada, April 2018. Association for Computing Machinery. ISBN 978-1-4503-5620-6. doi: 10.1145/3173574.3173610.
- [180] Johnny C. Lee, Daniel Avrahami, Scott E. Hudson, Jodi Forlizzi, Paul H. Dietz, and Darren Leigh. The calder toolkit: Wired and wireless components for rapidly prototyping interactive devices. In *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, DIS '04, pages 167–175, Cambridge, MA, USA, August 2004. Association for Computing Machinery. ISBN 978-1-58113-787-3. doi: 10.1145/1013115.1013139.
- [181] B. F. LeVeau and C. Rogers. Selective training of the vastus medialis muscle using EMG biofeedback. *Phys Ther*, 60(11):1410–1415, November 1980. ISSN 0031-9023. doi: 10.1093/ptj/60.11.1410.
- [182] Ian Li, Anind K. Dey, and Jodi Forlizzi. Understanding my data, myself: Supporting self-reflection with ubicomp technologies. In *Proceedings of the 13th International Conference on Ubiquitous Computing*, UbiComp '11, pages 405–414, New York, NY, USA, September 2011. Association for Computing Machinery. ISBN 978-1-4503-0630-0. doi: 10.1145/2030112.2030166.
- [183] Calvin Liang, Jakob Karolus, Thomas Kosch, and Albrecht Schmidt. On the Suitability of Real-Time Assessment of Programming Proficiency using Gaze Properties. In *Proceedings of the 7th ACM International Symposium on Pervasive Displays*, PerDis '18, pages 1–2, Munich, Germany, June 2018. Association for Computing Machinery. ISBN 978-1-4503-5765-4. doi: 10.1145/3205873.3210702.

- [184] Rhema Linder, Shamsi T. Iqbal, and Jaime Teevan. Outsider Perspectives: Crowd-Based Feedback for Writing. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, pages 1–6, New York, NY, USA, April 2018. Association for Computing Machinery. ISBN 978-1-4503-5621-3. doi: 10.1145/3170427.3188602.
- [185] Markus Löchtefeld, Sven Gehring, Ralf Jung, and Antonio Krüger. guitarAR: Supporting Guitar Learning Through Mobile Projection. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, pages 1447–1452, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0268-5. doi: 10.1145/1979742.1979789.
- [186] C. J. De Luca. Physiology and Mathematics of Myoelectric Signals. *IEEE Transactions on Biomedical Engineering*, BME-26(6):313–325, June 1979. ISSN 0018-9294. doi: 10.1109/TBME.1979.326534.
- [187] Wendy E. Mackay. Triggers and barriers to customizing software. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '91, pages 153–160, New York, NY, USA, March 1991. Association for Computing Machinery. ISBN 978-0-89791-383-6. doi: 10.1145/108844.108867.
- [188] Jaime Madell and Sylvie Héébert. Eye movements and music reading: Where do we look next? *Music Perception*, 26(2):157–170, December 2008. ISSN 0730-7829. doi: 10.1525/mp.2008.26.2.157.
- [189] Päivi Majaranta and Kari-Jouko Räihä. Twenty Years of Eye Typing: Systems and Design Issues. In *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications*, ETRA '02, pages 15–22, New York, NY, USA, 2002. ACM. ISBN 978-1-58113-467-4. doi: 10.1145/507072.507076.
- [190] Sylvain Malacria, Joey Scarr, Andy Cockburn, Carl Gutwin, and Tovi Grossman. Skilometers: Reflective widgets that motivate and help users to improve performance. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, pages 321–330, St. Andrews, Scotland, United Kingdom, October 2013. Association for Computing Machinery. ISBN 978-1-4503-2268-3. doi: 10.1145/2501988.2501996.
- [191] Regan L. Mandryk, M. Stella Atkins, and Kori M. Inkpen. A Continuous and Objective Evaluation of Emotional Experience with Interactive Play Environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 1027–1036, New York, NY, USA, 2006. ACM. ISBN 978-1-59593-372-0. doi: 10.1145/1124772.1124926.
- [192] Marcel Manshanden. Piano books divided into 15 levels - PianoBookGuide. January 2021. URL <https://www.pianobookguide.com/piano-book-level-search>.

-
- [193] Sampada Marathe and S. Shyam Sundar. What drives customization? control or identity? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 781–790, New York, NY, USA, May 2011. Association for Computing Machinery. ISBN 978-1-4503-0228-9. doi: 10.1145/1978942.1979056.
- [194] Javier Marco, Eva Cerezo, and Sandra Baldassarri. ToyVision: A toolkit for prototyping tabletop tangible games. In *Proceedings of the 4th ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, EICS '12, pages 71–80, New York, NY, USA, 2012. Association for Computing Machinery. ISBN 978-1-4503-1168-7. doi: 10.1145/2305484.2305498.
- [195] Nicolai Marquardt and Saul Greenberg. Distributed physical interfaces with shared phidgets. In *Proceedings of the 1st International Conference on Tangible and Embedded Interaction*, TEI '07, pages 13–20, Baton Rouge, Louisiana, February 2007. Association for Computing Machinery. ISBN 978-1-59593-619-6. doi: 10.1145/1226969.1226973.
- [196] Nicolai Marquardt, Robert Diaz-Marino, Sebastian Boring, and Saul Greenberg. The proximity toolkit: Prototyping proxemic interactions in ubiquitous computing ecologies. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, UIST '11, pages 315–326, Santa Barbara, California, USA, October 2011. Association for Computing Machinery. ISBN 978-1-4503-0716-1. doi: 10.1145/2047196.2047238.
- [197] Linda Marschner, Sebastian Pannasch, Johannes Schulz, and Sven-Thomas Graupner. Social communication with virtual agents: The effects of body and gaze direction on attention and emotional responding in human observers. *International Journal of Psychophysiology*, 97(2):85–92, August 2015. ISSN 0167-8760. doi: 10.1016/j.ijpsycho.2015.05.007.
- [198] Rachel Martin, Ivanka Prichard, Amanda D Hutchinson, and Carlene Wilson. The role of body awareness and mindfulness in the relationship between exercise and eating behavior. *Journal of Sport and Exercise Psychology*, 35(6):655–660, 2013.
- [199] Pascual Martínez-Gómez and Akiko Aizawa. Recognition of Understanding Level and Language Skill Using Measurements of Reading Behavior. In *Proceedings of the 19th International Conference on Intelligent User Interfaces*, IUI '14, pages 95–104, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2184-6. doi: 10.1145/2557500.2557546.
- [200] Justin Matejka, Wei Li, Tovi Grossman, and George Fitzmaurice. Community-Commands: Command recommendations for software applications. In *Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology*, UIST '09, pages 193–202, New York, NY, USA, October 2009. Association for Computing Machinery. ISBN 978-1-60558-745-5. doi: 10.1145/1622176.1622214.

- [201] Michael Mauderer, David R. Flatla, and Miguel A. Nacenta. Gaze-Contingent Manipulation of Color Perception. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 5191–5202, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-3362-7. doi: 10.1145/2858036.2858320.
- [202] Joanna McGrenere, Ronald M. Baecker, and Kellogg S. Booth. An evaluation of a multiple interface design solution for bloated software. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '02, pages 164–170, New York, NY, USA, April 2002. Association for Computing Machinery. ISBN 978-1-58113-453-7. doi: 10.1145/503376.503406.
- [203] Jess McIntosh, Charlie McNeill, Mike Fraser, Frederic Kerber, Markus Löchtefeld, and Antonio Krüger. EMPress: Practical Hand Gesture Classification with Wrist-Mounted EMG and Pressure Sensing. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 2332–2342, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-3362-7. doi: 10.1145/2858036.2858093.
- [204] Andrew P. McPherson, Adrian Gierakowski, and Adam M. Stark. The Space Between the Notes: Adding Expressive Pitch Control to the Piano Keyboard. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 2195–2204, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1899-0. doi: 10.1145/2470654.2481302.
- [205] Timothy Horrell MEng, Ayman El-Baz PhD, Joshua Baruth MS, Allan Tasman MD, Guela Sokhadze, Christopher Stewart MD, and Estate Sokhadze PhD. Neurofeedback effects on evoked and induced EEG gamma band reactivity to drug-related cues in cocaine addiction. *Journal of Neurotherapy*, 14(3):195–216, 2010. doi: 10.1080/10874208.2010.501498.
- [206] Roberto Merletti and Dario Farina. *Surface Electromyography: Physiology, Engineering and Applications*. IEEE Press Series on Biomedical Engineering. John Wiley & Sons, March 2016. ISBN 978-1-119-08290-3.
- [207] Roberto Merletti and Philip A. Parker. *Electromyography: Physiology, Engineering, and Non-Invasive Applications*. IEEE Press Series on Biomedical Engineering. John Wiley & Sons, July 2004. ISBN 978-0-471-67580-8.
- [208] Csikszentmihalyi Mihaly. *Flow: The psychology of optimal performance*. 1990.
- [209] James W. Moore. What Is the Sense of Agency and Why Does it Matter? *Front Psychol*, 7, August 2016. ISSN 1664-1078. doi: 10.3389/fpsyg.2016.01272.
- [210] Norman C. Moore. A review of EEG biofeedback treatment of anxiety disorders. *Clinical Electroencephalography*, 31(1):1–6, 2000. doi: 10.1177/155005940003100105.

-
- [211] Jörg Müller, Florian Alt, Daniel Michelis, and Albrecht Schmidt. Requirements and Design Space for Interactive Public Displays. In *Proceedings of the 18th ACM International Conference on Multimedia*, MM '10, pages 1285–1294, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-933-6. doi: 10.1145/1873951.1874203.
- [212] Ganesh R. Naik, Dinesh Kant Kumar, Vijay Pal Singh, and Marimuthu Palaniswami. Hand Gestures for HCI Using ICA of EMG. In *Proceedings of the HCSNet Workshop on Use of Vision in Human-Computer Interaction - Volume 56*, VisHCI '06, pages 67–72, Darlinghurst, Australia, Australia, 2006. Australian Computer Society, Inc. URL <http://dl.acm.org/citation.cfm?id=1273385.1273397>.
- [213] Jeanne Nakamura and Mihaly Csikszentmihalyi. Flow theory and research. *The Oxford Handbook of Positive Psychology*, pages 195–206, 2009. doi: 10.1093/oxfordhb/9780195187243.013.0018.
- [214] Michael Nebeling, Theano Mintsy, Maria Husmann, and Moira Norrie. Interactive development of cross-device user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 2793–2802, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 978-1-4503-2473-1. doi: 10.1145/2556288.2556980.
- [215] Thi Thao Duyen T. Nguyen, Thomas Garnarcz, Felicia Ng, Laura A. Dabbish, and Steven P. Dow. Fruitful Feedback: Positive Affective Language and Source Anonymity Improve Critique Reception and Work Outcomes. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, pages 1024–1034, New York, NY, USA, February 2017. Association for Computing Machinery. ISBN 978-1-4503-4335-0. doi: 10.1145/2998181.2998319.
- [216] Scott Nicholson. A recipe for meaningful gamification. In *Gamification in Education and Business*, pages 1–20. Springer, 2015. doi: 10.1007/978-3-319-10208-5_1.
- [217] Aditya Shekhar Nittala, Arshad Khan, Klaus Kruttwig, Tobias Kraus, and Jürgen Steimle. PhysioSkin: Rapid Fabrication of Skin-Conformal Physiological Interfaces. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pages 1–10, New York, NY, USA, April 2020. Association for Computing Machinery. ISBN 978-1-4503-6708-0. doi: 10.1145/3313831.3376366.
- [218] Eslam Nofal, Georgia Panagiotidou, Rabee M. Reffat, Hendrik Hameeuw, Vanessa Boschloos, and Andrew Vande Moere. Situated tangible gamification of heritage for supporting collaborative learning of young museum visitors. *J. Comput. Cult. Herit.*, 13(1), February 2020. ISSN 1556-4673. doi: 10.1145/3350427.
- [219] Unaizah Obaidallah, Mohammed Al Haek, and Peter C.-H. Cheng. A Survey on the Usage of Eye-Tracking in Computer Programming. *ACM Comput. Surv.*, 51(1):5:1–5:58, January 2018. ISSN 0360-0300. doi: 10.1145/3145904.

- [220] Unaizah Obaidellah, Tanja Blascheck, Drew T. Guarnera, and Jonathan Maletic. A Fine-grained Assessment on Novice Programmers' Gaze Patterns on Pseudocode Problems. In *ACM Symposium on Eye Tracking Research and Applications, ETRA '20 Short Papers*, pages 1–5, New York, NY, USA, June 2020. Association for Computing Machinery. ISBN 978-1-4503-7134-6. doi: 10.1145/3379156.3391982.
- [221] Ayano Okoso, Takumi Toyama, Kai Kunze, Joachim Folz, Marcus Liwicki, and Koichi Kise. Towards Extraction of Subjective Reading Incomprehension: Analysis of Eye Gaze Features. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA '15*, pages 1325–1330, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3146-3. doi: 10.1145/2702613.2732896.
- [222] Florin Oprescu, Christian Jones, and Mary Katsikitis. I PLAY AT WORK - ten principles for transforming work processes through gamification. *Frontiers in psychology*, 5:14, 2014. doi: 10.3389/fpsyg.2014.00014.
- [223] Rita Orji, Gustavo F Tondello, and Lennart E Nacke. Personalizing persuasive strategies in gameful systems to gamification user types. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2018. doi: 10.1145/3173574.3174009.
- [224] Pascal Oser, Sebastian S. Feger, Pawel W. Wozniak, Jakob Karolus, Dayana Spagnuolo, Akash Gupta, Stefan Lüders, Albrecht Schmidt, and Frank Kargl. SAFER: Development and evaluation of an IoT device risk assessment framework in a multinational organization. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 4(3): 114:1–114:22, 2020. doi: 10.1145/3414173.
- [225] Oxford English Dictionary. Proficiency, March 2021. URL <https://www.oed.com>.
- [226] Se-yeon Park and Won-gyu Yoo. Effect of EMG-based feedback on posture correction during computer operation. *Journal of Occupational Health*, 54(4):271–277, 2012. doi: 10.1539/joh.12-0052-OA.
- [227] Richard Parncutt and Gary McPherson. *The Science & Psychology of Music Performance: Creative Strategies for Teaching and Learning*. Oxford University Press, Oxford and New York, 2002. ISBN 0-19-513810-4.
- [228] Prajwal Paudyal, Ayan Banerjee, and Sandeep K.S. Gupta. SCEPTRE: A Pervasive, Non-Invasive, and Programmable Gesture Recognition Technology. In *Proceedings of the 21st International Conference on Intelligent User Interfaces, IUI '16*, pages 282–293, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-4137-0. doi: 10.1145/2856767.2856794.
- [229] G. M. Paul, F. Cao, R. Torah, K. Yang, S. Beeby, and J. Tudor. A smart textile based facial EMG and EOG computer interface. *IEEE Sensors Journal*, 14(2):393–400, 2014.

-
- [230] D. Peleg, E. Braiman, E. Yom-Tov, and G. F. Inbar. Classification of finger activation for use in a robotic prosthesis arm. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 10(4):290–293, December 2002. ISSN 1534-4320. doi: 10.1109/TNSRE.2002.806831.
- [231] Erik Peper, Richard Harvey, and Naoki Takebayashi. Biofeedback an evidence based approach in clinical practice. *Japanese Journal of Biofeedback Research*, 36(1):3–10, 2009. doi: 10.20595/jjbf.36.1_3.
- [232] Monica Perusquía-Hernández, Saho Ayabe-Kanamura, Kenji Suzuki, and Shiro Kumanano. The Invisible Potential of Facial Electromyography: A Comparison of EMG and Computer Vision when Distinguishing Posed from Spontaneous Smiles. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 1–9, New York, NY, USA, May 2019. Association for Computing Machinery. ISBN 978-1-4503-5970-2. doi: 10.1145/3290605.3300379.
- [233] Max Pfeiffer, Tim Duentel, and Michael Rohs. Let your body move: A prototyping toolkit for wearable force feedback with electrical muscle stimulation. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '16, pages 418–427, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 978-1-4503-4408-1. doi: 10.1145/2935334.2935348.
- [234] Bastian Pfleging, Drea K. Fekety, Albrecht Schmidt, and Andrew L. Kun. A model relating pupil diameter to mental workload and lighting conditions. In Jofish Kaye, Allison Druin, Cliff Lampe, Dan Morris, and Juan Pablo Hourcade, editors, *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, San Jose, CA, USA, May 7-12, 2016, pages 5776–5788. ACM, 2016. doi: 10.1145/2858036.2858117.
- [235] Angkoon Phinyomark and Erik Scheme. EMG Pattern Recognition in the Era of Big Data and Deep Learning. *BDCC*, 2(3):21, August 2018. ISSN 2504-2289. doi: 10.3390/bdcc2030021.
- [236] Angkoon Phinyomark, Pornchai Phukpattaranont, and Chusak Limsakul. Feature reduction and selection for EMG signal classification. *Expert Systems with Applications*, 39(8):7420–7431, June 2012. ISSN 09574174. doi: 10.1016/j.eswa.2012.01.102.
- [237] Project Petrucci LLC. Piano pieces by level – IMSLP. January 2021. URL <https://imslp.org/wiki/Special:DiffPage/DiffMain/1>.
- [238] Uwe Proske and Simon C. Gandevia. The Proprioceptive Senses: Their Roles in Signaling Body Shape, Body Position and Movement, and Muscle Force. *Physiological Reviews*, 92(4):1651–1697, October 2012. ISSN 0031-9333. doi: 10.1152/physrev.00048.2011.

- [239] Marjaana Puurtinen. Eye on music reading: A methodological review of studies from 1994 to 2017. *Journal of Eye Movement Research*, 11(2):1–16, 2018.
- [240] Ross Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo, CA, 1993.
- [241] Pernilla Qvarfordt and Shumin Zhai. Conversing with the User Based on Eye-gaze Patterns. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, pages 221–230, New York, NY, USA, 2005. ACM. ISBN 978-1-58113-998-3. doi: 10.1145/1054972.1055004.
- [242] M.B.I. Raez, M.S. Hussain, and F. Mohd-Yasin. Techniques of EMG signal analysis: Detection, processing, classification and applications. *Biol Proced Online*, 8:11–35, March 2006. ISSN 1480-9222. doi: 10.1251/bpo115.
- [243] Raf Ramakers, Kashyap Todi, and Kris Luyten. PaperPulse: An integrated approach for embedding electronics in paper designs. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 2457–2466, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 978-1-4503-3145-6. doi: 10.1145/2702123.2702487.
- [244] Raf Ramakers, Fraser Anderson, Tovi Grossman, and George Fitzmaurice. RetroFab: A design tool for retrofitting physical interfaces using actuators, sensors and 3D printing. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 409–419, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 978-1-4503-3362-7. doi: 10.1145/2858036.2858485.
- [245] Keith Rayner. Eye movements in reading and information processing: 20 years of research. *Psychol Bull*, 124(3):372–422, November 1998. ISSN 0033-2909.
- [246] Keith Rayner and Alexander Pollatsek. Eye movements, the eye-hand span, and the perceptual span during sight-reading of music. *Current Directions in Psychological Science*, 6(2):49–53, 1997. doi: 10.1111/1467-8721.ep11512647.
- [247] Keith Rayner, Timothy J. Slattery, and Nathalie N. Bélanger. Eye movements, the perceptual span, and reading speed. *Psychonomic bulletin & review*, 17(6):834–839, 2010. doi: <http://dx.doi.org/10.3758/PBR.17.6.834>.
- [248] Erik D. Reichle, Alexander Pollatsek, Donald L. Fisher, and Keith Rayner. Toward a model of eye movement control in reading. *Psychological Review*, 105(1):125–157, 1998. ISSN 0033-295X. doi: 10.1037/0033-295X.105.1.125.
- [249] David Robinson and Victoria Bellotti. A preliminary taxonomy of gamification elements for varying anticipated commitment. In *Proc. ACM CHI 2013 Workshop on Designing Gamification: Creating Gameful and Playful Experiences*, 2013.

-
- [250] Yvonne Rogers. Moving on from Weiser's Vision of Calm Computing: Engaging UbiComp Experiences. In Paul Dourish and Adrian Friday, editors, *UbiComp 2006: Ubiquitous Computing*, Lecture Notes in Computer Science, pages 404–421, Berlin, Heidelberg, 2006. Springer. ISBN 978-3-540-39635-2. doi: 10.1007/11853565_24.
- [251] Richard M Ryan and Edward L Deci. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist*, 55(1):68, 2000.
- [252] Richard M Ryan and Heather Patrick. Self-determination theory and physical. *Hellenic journal of psychology*, 6:107–124, 2009.
- [253] Herman Saksono and Andrea G. Parker. Reflective informatics through family storytelling: Self-discovering physical activity predictors. *Conference on Human Factors in Computing Systems - Proceedings*, 2017-May:5232–5244, 2017. doi: 10.1145/3025453.3025651.
- [254] Alan W. Salmoni, Richard A. Schmidt, and Charles B. Walter. Knowledge of results and motor learning: A review and critical reappraisal. *Psychological Bulletin*, 95(3):355–386, 1984. ISSN 1939-1455(Electronic),0033-2909(Print). doi: 10.1037/0033-2909.95.3.355.
- [255] Dario D. Salvucci and John R. Anderson. Intelligent Gaze-added Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '00, pages 273–280, New York, NY, USA, 2000. ACM. ISBN 978-1-58113-216-8. doi: 10.1145/332040.332444.
- [256] Dario D. Salvucci and Joseph H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, pages 71–78. ACM, 2000. doi: <http://dx.doi.org/10.1145/355017.355028>.
- [257] Charles Lima Sanches, Koichi Kise, and Olivier Augereau. Japanese reading objective understanding estimation by eye gaze analysis. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, UbiComp '17, pages 121–124, New York, NY, USA, September 2017. Association for Computing Machinery. ISBN 978-1-4503-5190-4. doi: 10.1145/3123024.3123092.
- [258] Carmen Sanchez and David Dunning. Overconfidence among beginners: Is a little learning a dangerous thing? *Journal of Personality and Social Psychology*, 114(1):10–28, 2018. ISSN 1939-1315(Electronic),0022-3514(Print). doi: 10.1037/pspa0000102.
- [259] Anthony Santella and Doug DeCarlo. Robust Clustering of Eye Movement Recordings for Quantification of Visual Interest. In *Proceedings of the 2004 Symposium*

- on Eye Tracking Research & Applications*, ETRA '04, pages 27–34, New York, NY, USA, 2004. ACM. ISBN 978-1-58113-825-2. doi: 10.1145/968363.968368.
- [260] Anthony Santella, Maneesh Agrawala, Doug DeCarlo, David Salesin, and Michael Cohen. Gaze-based Interaction for Semi-Automatic Photo Cropping. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 771–780, New York, NY, USA, 2006. ACM. ISBN 978-1-59593-372-0. doi: 10.1145/1124772.1124886.
- [261] T Scott Saponas, Desney S. Tan, Dan Morris, and Ravin Balakrishnan. Demonstrating the Feasibility of Using Forearm Electromyography for Muscle-computer Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 515–524, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-011-1. doi: 10.1145/1357054.1357138.
- [262] T. Scott Saponas, Desney S. Tan, Dan Morris, Ravin Balakrishnan, Jim Turner, and James A. Landay. Enabling Always-available Input with Muscle-computer Interfaces. In *Proceedings of the 22Nd Annual ACM Symposium on User Interface Software and Technology*, UIST '09, pages 167–176, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-745-5. doi: 10.1145/1622176.1622208.
- [263] T. Scott Saponas, Desney S. Tan, Dan Morris, Jim Turner, and James A. Landay. Making muscle-computer interfaces more practical. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 851–854, New York, NY, USA, April 2010. Association for Computing Machinery. ISBN 978-1-60558-929-9. doi: 10.1145/1753326.1753451.
- [264] S. Sarcar, J. P. P. Jokinen, A. Oulasvirta, Z. Wang, C. Silpasuwanchai, and X. Ren. Ability-Based Optimization of Touchscreen Interactions. *IEEE Pervasive Computing*, 17(1):15–26, January 2018. ISSN 1558-2590. doi: 10.1109/MPRV.2018.011591058.
- [265] Sayan Sarcar, Jussi Jokinen, Antti Oulasvirta, Chaklam Silpasuwanchai, Zhenxin Wang, and Xiangshi Ren. Towards Ability-Based Optimization for Aging Users. In *Proceedings of the International Symposium on Interactive Technology and Ageing Populations*, ITAP '16, pages 77–86, New York, NY, USA, October 2016. Association for Computing Machinery. ISBN 978-1-4503-4746-4. doi: 10.1145/2996267.2996275.
- [266] Valkyrie Savage, Xiaohan Zhang, and Björn Hartmann. Midas: Fabricating custom capacitive touch sensors to prototype interactive objects. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, UIST '12, pages 579–588, Cambridge, Massachusetts, USA, October 2012. Association for Computing Machinery. ISBN 978-1-4503-1580-7. doi: 10.1145/2380116.2380189.
- [267] Valkyrie Savage, Colin Chang, and Björn Hartmann. Sauron: Embedded single-camera sensing of printed physical user interfaces. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, pages

-
- 447–456, St. Andrews, Scotland, United Kingdom, October 2013. Association for Computing Machinery. ISBN 978-1-4503-2268-3. doi: 10.1145/2501988.2501992.
- [268] Valkyrie Savage, Sean Follmer, Jingyi Li, and Björn Hartmann. Makers’ marks: Physical markup for designing and fabricating functional objects. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology, UIST ’15*, pages 103–108, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 978-1-4503-3779-3. doi: 10.1145/2807442.2807508.
- [269] Abraham. Savitzky and M. J. E. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8):1627–1639, 1964. doi: 10.1021/ac60214a047.
- [270] Joey Scarr, Andy Cockburn, Carl Gutwin, and Philip Quinn. Dips and ceilings: Understanding and supporting transitions to expertise in user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’11*, pages 2741–2750, New York, NY, USA, May 2011. Association for Computing Machinery. ISBN 978-1-4503-0228-9. doi: 10.1145/1978942.1979348.
- [271] M. H. Schieber. Muscular production of individuated finger movements: The roles of extrinsic finger muscles. *J. Neurosci.*, 15(1):284–297, January 1995. ISSN 0270-6474, 1529-2401. URL <http://www.jneurosci.org/content/15/1/284>.
- [272] B. Schilit, N. Adams, and R. Want. Context-Aware Computing Applications. In *1994 First Workshop on Mobile Computing Systems and Applications*, pages 85–90, December 1994. doi: 10.1109/WMCSA.1994.16.
- [273] Albrecht Schmidt. *Ubiquitous Computing - Computing in Context*. PhD thesis, The University of Lancaster, UK, 2003.
- [274] Albrecht Schmidt. Biosignals in human-computer interaction. *interactions*, 23(1): 76–79, December 2016. ISSN 1072-5520. doi: 10.1145/2851072.
- [275] Greg W. Schmidt and Gary L. Stein. Sport Commitment: A Model Integrating Enjoyment, Dropout, and Burnout. *Journal of Sport and Exercise Psychology*, 13(3): 254–265, September 1991. ISSN 1543-2904, 0895-2779. doi: 10.1123/jsep.13.3.254.
- [276] Martin Schmitz, Martin Stitz, Florian Müller, Markus Funk, and Max Mühlhäuser. ./Trilaterate: A fabrication pipeline to design and 3D print hover-, touch-, and force-sensitive objects. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI ’19*, pages 1–13, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 978-1-4503-5970-2. doi: 10.1145/3290605.3300684.
- [277] Wolfgang Schnotz and Christian Kürschner. A reconsideration of cognitive load theory. *Educational psychology review*, 19(4):469–508, 2007. doi: 10.1007/s10648-007-9053-4.

- [278] D.A. Schön. *The Reflective Practitioner: How Professionals Think in Action*. 1983.
- [279] Dominik Schuldhuis. *Human Activity Recognition in Daily Life and Sports Using Inertial Sensors*. PhD thesis, University of Erlangen-Nuremberg, Germany, 2019. URL <http://d-nb.info/1201240514>.
- [280] Seth J. Schwartz and Alan S. Waterman. Changing interests: A longitudinal study of intrinsic motivation for personally salient activities. *Journal of Research in Personality*, 40(6):1119–1136, 2006. ISSN 0092-6566. doi: 10.1016/j.jrp.2005.12.003.
- [281] William C. Scott, David Kaiser, Siegfried Othmer, and Stephen I. Sideroff. Effects of an EEG biofeedback protocol on a mixed substance abusing population. *The American Journal of Drug and Alcohol Abuse*, 31(3):455–469, 2005. doi: 10.1081/ADA-200056807.
- [282] Katie Seaborn and Deborah I Fels. Gamification in theory and action: A survey. *International Journal of human-computer studies*, 74:14–31, 2015.
- [283] Sara C. Sereno and Keith Rayner. Measuring word recognition in reading: Eye movements and event-related potentials. *Trends in cognitive sciences*, 7(11):489–493, 2003. doi: <http://dx.doi.org/10.1016/j.tics.2003.09.010>.
- [284] Teddy Seyed, Alaa Azazi, Edwin Chan, Yuxi Wang, and Frank Maurer. SoD-Toolkit: A toolkit for interactively prototyping and developing multi-sensor, multi-device environments. In *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces, ITS '15*, pages 171–180, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 978-1-4503-3899-8. doi: 10.1145/2817721.2817750.
- [285] Kennon M Sheldon, Andrew J Elliot, Youngmee Kim, and Tim Kasser. What is satisfying about satisfying events? Testing 10 candidate psychological needs. *Journal of personality and social psychology*, 80(2):325, 2001.
- [286] Roberta Shroyer. Actual readers versus implied readers: Role conflicts in office 97. *Technical Communication*, 47(2):238–240, 2000. ISSN 00493155, 1938369X. URL <http://www.jstor.org/stable/43748856>.
- [287] John L. Sibert, Mehmet Gokturk, and Robert A. Lavine. The Reading Assistant: Eye Gaze Triggered Auditory Prompting for Reading Remediation. In *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology, UIST '00*, pages 101–107, New York, NY, USA, 2000. ACM. ISBN 978-1-58113-212-0. doi: 10.1145/354401.354418.
- [288] Linda E. Sibert and Robert J. K. Jacob. Evaluation of Eye Gaze Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '00*, pages 281–288, New York, NY, USA, 2000. ACM. ISBN 978-1-58113-216-8. doi: 10.1145/332040.332445.

-
- [289] Greg J. Siegle, Naho Ichikawa, and Stuart Steinhauer. Blink before and after you think: Blinks occur prior to and following cognitive load indexed by pupillary responses. *Psychophysiology*, 45(5):679–687, September 2008. ISSN 0048-5772. doi: 10.1111/j.1469-8986.2008.00681.x.
- [290] C. Estelle Smith, Eduardo Nevarez, and Haiyi Zhu. Disseminating research news in HCI: Perceived hazards, how-to’s, and opportunities for innovation. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI ’20, pages 1–13, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 978-1-4503-6708-0. doi: 10.1145/3313831.3376744.
- [291] Tony Souter. *Eye Movement, Memory and Tempo in the Sight Reading of Keyboard Music*. PhD thesis, University of Sydney / University of Sydney / Sydney Conservatorium of Music, January 2001. URL <http://hdl.handle.net/2123/8674>.
- [292] Constantine Stephanidis, Gavriel Salvendy, Margherita Antona, Jessie Y. C. Chen, Jianming Dong, Vincent G. Duffy, Xiaowen Fang, Cali Fidopiastis, Gino Fragomeni, Limin Paul Fu, Yinni Guo, Don Harris, Andri Ioannou, Kyeong-ah (Kate) Jeong, Shin’ichi Konomi, Heidi Krömker, Masaaki Kurosu, James R. Lewis, Aaron Marcus, Gabriele Meiselwitz, Abbas Moallem, Hirohiko Mori, Fiona Fui-Hoon Nah, Stavroula Ntoa, Pei-Luen Patrick Rau, Dylan Schmorrow, Keng Siau, Norbert Streitz, Wentao Wang, Sakae Yamamoto, Panayiotis Zaphiris, and Jia Zhou. Seven HCI Grand Challenges. *International Journal of Human–Computer Interaction*, 35(14): 1229–1269, August 2019. ISSN 1044-7318. doi: 10.1080/10447318.2019.1619259.
- [293] William Steptoe, Robin Wolff, Alessio Murgia, Estefania Guimaraes, John Rae, Paul Sharkey, David Roberts, and Anthony Steed. Eye-tracking for Avatar Eye-gaze and Interactional Analysis in Immersive Collaborative Virtual Environments. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, CSCW ’08, pages 197–200, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-007-4. doi: 10.1145/1460563.1460593.
- [294] Evan Strasnick, Sean Follmer, and Maneesh Agrawala. Pinpoint: A PCB debugging pipeline using interruptible routing and instrumentation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI ’19, pages 1–11, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 978-1-4503-5970-2. doi: 10.1145/3290605.3300278.
- [295] Paul Strohmeier, Narjes Pourjafarian, Marion Koelle, Cedric Honnet, Bruno Fruchard, and Jürgen Steimle. Sketching on-body interactions using piezo-resistive kinesiology tape. In *Proceedings of the Augmented Humans International Conference*, AHs ’20, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 978-1-4503-7603-7. doi: 10.1145/3384657.3384774.
- [296] Wolfgang Stuerzlinger, Olivier Chapuis, Dusty Phillips, and Nicolas Roussel. User interface facades: Towards fully adaptable user interfaces. In *Proceedings of the*

- 19th Annual ACM Symposium on User Interface Software and Technology*, UIST '06, pages 309–318, New York, NY, USA, October 2006. Association for Computing Machinery. ISBN 978-1-59593-313-3. doi: 10.1145/1166253.1166301.
- [297] C-H. Su and C-H. Cheng. A mobile gamification learning system for improving the learning motivation and achievements. *Journal of Computer Assisted Learning*, 31(3):268–286, 2015. doi: 10.1111/jcal.12088.
- [298] Leena Subramanian, John V. Hindle, Stephen Johnston, Mark V. Roberts, Masud Husain, Rainer Goebel, and David Linden. Real-time functional magnetic resonance imaging neurofeedback for treatment of parkinson’s disease. *Journal of Neuroscience*, 31(45):16309–16317, 2011. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.3498-11.2011.
- [299] Mathias Sundholm, Jingyuan Cheng, Bo Zhou, Akash Sethi, and Paul Lukowicz. Smart-mat: Recognizing and counting gym exercises with low-cost resistive pressure sensing matrix. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '14, pages 373–382, New York, NY, USA, September 2014. Association for Computing Machinery. ISBN 978-1-4503-2968-2. doi: 10.1145/2632048.2636088.
- [300] Dag Svanæs. Interaction design for and with the lived body: Some implications of merleau-ponty’s phenomenology. *ACM Trans. Comput.-Hum. Interact.*, 20(1), April 2013. ISSN 1073-0516. doi: 10.1145/2442106.2442114.
- [301] John Sweller. Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, 12(2):257–285, April 1988. ISSN 1551-6709. doi: 10.1207/s15516709cog1202_4.
- [302] John Sweller, Jeroen J. G. van Merriënboer, and Fred G. W. C. Paas. Cognitive Architecture and Instructional Design. *Educational Psychology Review*, 10(3):251–296, September 1998. ISSN 1573-336X. doi: 10.1023/A:1022193728205.
- [303] Atau Tanaka and R. Benjamin Knapp. 2002: Multimodal Interaction in Music Using the Electromyogram and Relative Position Sensing. In *A NIME Reader*, Current Research in Systematic Musicology, pages 45–58. Springer, Cham, 2017. ISBN 978-3-319-47213-3 978-3-319-47214-0. doi: 10.1007/978-3-319-47214-0_4.
- [304] Marc Teyssier, Gilles Bailly, Catherine Pelachaud, Eric Lecolinet, Andrew Conn, and Anne Roudaut. Skin-On Interfaces: A Bio-Driven Approach for Artificial Skin Design to Cover Interactive Devices. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, UIST '19, pages 307–322, New Orleans, LA, USA, October 2019. Association for Computing Machinery. ISBN 978-1-4503-6816-2. doi: 10.1145/3332165.3347943.
- [305] The 2020 Florida Statutes. Readable language in insurance policies (Title XXXVII, Chapter 627.4145). 2020.

-
- [306] Marian Theiss, Philipp M. Scholl, and Kristof Van Laerhoven. Predicting grasps with a wearable inertial and EMG sensing unit for low-power detection of in-hand objects. In *Proceedings of the 7th Augmented Human International Conference 2016, AH '16*, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 978-1-4503-3680-2. doi: 10.1145/2875194.2875207.
- [307] Laurentiu Toader, Nicolai B. K. Jensen, and Michael B. Holte. The Influence of Biofeedback on Exercise Correctness and Muscle Activity. In Anthony L. Brooks, Eva Brooks, and Nikolas Vidakis, editors, *Interactivity, Game Creation, Design, Learning, and Innovation*, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, pages 139–150, Cham, 2018. Springer International Publishing. ISBN 978-3-319-76908-0. doi: 10.1007/978-3-319-76908-0_14.
- [308] Gustavo F. Tondello, Rina R. Wehbe, Lisa Diamond, Marc Busch, Andrzej Marczewski, and Lennart E. Nacke. The gamification user types hexad scale. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play, CHI PLAY '16*, pages 229–243, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 978-1-4503-4456-2. doi: 10.1145/2967934.2968082.
- [309] Gustavo F. Tondello, Alberto Mora, and Lennart E. Nacke. Elements of gameful design emerging from user preferences. *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*, pages 129–142, 2017. doi: 10.1145/3116595.3116627.
- [310] Takumi Toyama, Daniel Sonntag, Andreas Dengel, Takahiro Matsuda, Masakazu Iwamura, and Koichi Kise. A Mixed Reality Head-mounted Text Translation System Using Eye Gaze Input. In *Proceedings of the 19th International Conference on Intelligent User Interfaces, IUI '14*, pages 329–334, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2184-6. doi: 10.1145/2557500.2557528.
- [311] L. J. Trejo, K. R. Wheeler, C. C. Jorgensen, R. Rosipal, S. T. Clanton, B. Matthews, A. D. Hibbs, R. Matthews, and M. Krupka. Multimodal neuroelectric interface development. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(2):199–203, June 2003. ISSN 1534-4320. doi: 10.1109/TNSRE.2003.814426.
- [312] Laia Turmo Vidal, Elena Márquez Segura, and Annika Waern. Movement Correction in Instructed Fitness Training: Design Recommendations and Opportunities. In *Proceedings of the 2018 Designing Interactive Systems Conference, DIS '18*, pages 1041–1054, New York, NY, USA, June 2018. Association for Computing Machinery. ISBN 978-1-4503-5198-0. doi: 10.1145/3196709.3196789.
- [313] Laia Turmo Vidal, Elena Márquez Segura, Christopher Boyer, and Annika Waern. Enlightened Yoga: Designing an Augmented Class with Wearable Lights to Support

- Instruction. In *Proceedings of the 2019 on Designing Interactive Systems Conference*, DIS '19, pages 1017–1031, New York, NY, USA, June 2019. Association for Computing Machinery. ISBN 978-1-4503-5850-7. doi: 10.1145/3322276.3322338.
- [314] Laia Turmo Vidal, Hui Zhu, and Abraham Riego-Delgado. BodyLights: Open-Ended Augmented Feedback to Support Training Towards a Correct Exercise Execution. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pages 1–14, New York, NY, USA, April 2020. Association for Computing Machinery. ISBN 978-1-4503-6708-0. doi: 10.1145/3313831.3376268.
- [315] April Tyack and Elisa D Mekler. Self-determination theory in HCI games research: Current uses and open questions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020. ISBN 978-1-4503-6708-0. doi: 10.1145/3313831.3376723.
- [316] George Tzanetakis. Natural Human-Computer Interaction with Musical Instruments. In *Digital Tools for Computer Music Production and Distribution*, pages 116–136. IGI Global, 2016.
- [317] Eduardo Velloso, Andreas Bulling, Hans Gellersen, Wallace Ugulino, and Hugo Fuks. Qualitative activity recognition of weight lifting exercises. In *Proceedings of the 4th Augmented Human International Conference*, AH '13, pages 116–123, New York, NY, USA, March 2013. Association for Computing Machinery. ISBN 978-1-4503-1904-1. doi: 10.1145/2459236.2459256.
- [318] Roel Vertegaal, Robert Slagter, Gerrit van der Veer, and Anton Nijholt. Eye Gaze Patterns in Conversations: There is More to Conversational Agents Than Meets the Eyes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '01, pages 301–308, New York, NY, USA, 2001. ACM. ISBN 978-1-58113-327-1. doi: 10.1145/365024.365119.
- [319] Nicolas Villar, James Scott, Steve Hodges, Kerry Hammil, and Colin Miller. .NET gadgeteer: A platform for custom devices. In *Proceedings of the 10th International Conference on Pervasive Computing*, Pervasive'12, pages 216–233, Newcastle, UK, June 2012. Springer-Verlag. ISBN 978-3-642-31204-5. doi: 10.1007/978-3-642-31205-2_14.
- [320] Scott R. Vrana. The psychophysiology of disgust: Differentiating negative emotional contexts with facial EMG. *Psychophysiology*, 30(3):279–286, May 1993. ISSN 1469-8986. doi: 10.1111/j.1469-8986.1993.tb03354.x.
- [321] Tina Caroline Walber, Ansgar Scherp, and Steffen Staab. Smart Photo Selection: Interpret Gaze As Personal Interest. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 2065–2074, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2473-1. doi: 10.1145/2556288.2557025.

-
- [322] Chiuan Wang, Hsuan-Ming Yeh, Bryan Wang, Te-Yen Wu, Hsin-Ruey Tsai, Rong-Hao Liang, Yi-Ping Hung, and Mike Y. Chen. CircuitStack: Supporting rapid prototyping and evolution of electronic circuits. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16, pages 687–695, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 978-1-4503-4189-9. doi: 10.1145/2984511.2984527.
- [323] Homer E. Weaver. Studies of ocular behavior in music reading. *Psychological Monographs*, 55(1):i–50, 1943. ISSN 0096-9753. doi: 10.1037/h0093537.
- [324] Mark Weiser. The computer for the 21st century. *Scientific American*, 265(3):66–75, January 1991.
- [325] Mark Weiser and John Seely Brown. *The Coming Age of Calm Technology*. 1996.
- [326] Michael Wessely, Theophanis Tsandilas, and Wendy E. Mackay. Shape-Aware Material: Interactive Fabrication with ShapeMe. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, UIST '18, pages 127–139, Berlin, Germany, October 2018. Association for Computing Machinery. ISBN 978-1-4503-5948-1. doi: 10.1145/3242587.3242619.
- [327] Jacob O. Wobbrock. Situationally aware mobile devices for overcoming situational impairments. In *Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, pages 1–18, Valencia Spain, June 2019. ACM. ISBN 978-1-4503-6745-5. doi: 10.1145/3319499.3330292.
- [328] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 143–146, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0228-9. doi: 10.1145/1978942.1978963.
- [329] Katrin Wolf and Jonas Willaredt. PickRing: Seamless interaction through pick-up detection. In *Proceedings of the 6th Augmented Human International Conference*, AH '15, pages 13–20, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 978-1-4503-3349-8. doi: 10.1145/2735711.2735792.
- [330] Klaus Wolters. *Handbuch Der Klavierliteratur: Klaviermusik Zu Zwei Händen*. Atlantis-Musikbuch. Atlantis-Musikbuch-Verl., Zürich, 5., vollst. rev. Aufl. edition, 2001. ISBN 3-254-00248-2.
- [331] Pawel W. Wozniak, Jakob Karolus, Florian Lang, Caroline Eckerth, Johannes Schöning, Yvonne Rogers, and Jasmin Niess. Creepy technology: What is it and how do you measure it? In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, Yokohama, Japan, May 08-13, 2021, page 13. ACM, 2021. doi: 10.1145/3411764.3445299.

- [332] Chi-Jui Wu, Steven Houben, and Nicolai Marquardt. EagleSense: Tracking people and devices in interactive spaces using real-time top-view depth-sensing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 3929–3942, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 978-1-4503-4655-9. doi: 10.1145/3025453.3025562.
- [333] Jishuo Yang and Daniel Wigdor. Panelrama: Enabling easy specification of cross-device web applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 2783–2792, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 978-1-4503-2473-1. doi: 10.1145/2556288.2557199.
- [334] Alfred Lukyanovich Yarbus. *Eye Movements and Vision*. New York: Plenum Press, 1967.
- [335] Kazuyo Yoshimura, Koichi Kise, and Kai Kunze. The eye as the window of the language ability: Estimation of English skills by analyzing eye movement while reading documents. In *Document Analysis and Recognition (ICDAR), 2015 13th International Conference On*, pages 251–255. IEEE, 2015. doi: <http://dx.doi.org/10.1109/ICDAR.2015.7333762>.
- [336] Bin Yu, Mathias Funk, Jun Hu, and Loe Feijs. StressTree: A metaphorical visualization for biofeedback-assisted stress management. In *Proceedings of the 2017 Conference on Designing Interactive Systems*, DIS '17, pages 333–337, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 978-1-4503-4922-2. doi: 10.1145/3064663.3064729.
- [337] Alvin Yuan, Kurt Luther, Markus Krause, Sophie Isabel Vennix, Steven P Dow, and Bjorn Hartmann. Almost an Expert: The Effects of Rubrics and Expertise on Perceived Value of Crowdsourced Design Critiques. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, pages 1005–1017, New York, NY, USA, February 2016. Association for Computing Machinery. ISBN 978-1-4503-3592-8. doi: 10.1145/2818048.2819953.
- [338] Beste F. Yuksel, Kurt B. Oleson, Lane Harrison, Evan M. Peck, Daniel Afergan, Remco Chang, and Robert JK Jacob. Learn Piano with BACH: An Adaptive Learning Interface That Adjusts Task Difficulty Based on Brain State. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 5372–5384, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-3362-7. doi: 10.1145/2858036.2858388.
- [339] Shumin Zhai, Carlos Morimoto, and Steven Ihde. Manual and Gaze Input Cascaded (MAGIC) Pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, pages 246–253, New York, NY, USA, 1999. ACM. ISBN 978-0-201-48559-2. doi: 10.1145/302979.303053.

-
- [340] X. Zhai, B. Jelfs, R. H. M. Chan, and C. Tin. Short latency hand movement classification based on surface EMG spectrogram with PCA. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 327–330, August 2016. doi: 10.1109/EMBC.2016.7590706.
- [341] Qiao Zhang, Shyamnath Gollakota, Ben Taskar, and Raj P.N. Rao. Non-intrusive Tongue Machine Interface. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems, CHI '14*, pages 2555–2558, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2473-1. doi: 10.1145/2556288.2556981.
- [342] Jonathan Zittrain. Ubiquitous human computing. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 366(1881):3813–3821, October 2008. ISSN 1364-503X, 1471-2962. doi: 10.1098/rsta.2008.0116.

APPENDIX

Contributing Publications

A substantial part of the submitted thesis is based on peer-reviewed publications in international venues. For all publications, I was the lead researcher of the projects and, for all but two, I am the first author of these. These exceptions were due to regulations of an internship program and an associated demo publication. All publications have been created in close collaborations with my co-authors but were primarily written by myself. In the following, I provide a detailed overview of the respective publications (published and in submission) and manuscripts, the co-authors, and my contribution to each of those.

For the following publications, I drafted the study design, implemented the prototypes, and conducted the evaluation (if applicable). I wrote most parts of the manuscript and led its preparation.

- J. Karolus, P. W. Woźniak, L. L. Chuang, and A. Schmidt. Robust gaze features for enabling language proficiency awareness. 2017. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, 2998–3010.
- J. Karolus, and P. W. Woźniak. Proficiency-Aware Systems: Designing for User Reflection in Context-Aware Systems. 2021. *IT - Information Technology*, 63(3), 167-175.
- J. Karolus and A. Schmidt. Proficiency-aware systems: Adapting to the user’s skills and expertise. 2018. In Proceedings of the 7th ACM International Symposium on Pervasive Displays (PerDis ’18). ACM, New York, NY, USA, Article 33, 2 pages.

For the following publication, I drafted the study design, implemented the prototypes, and conducted the evaluation. Minor parts of this paper are also based on student works. I wrote most parts of the manuscript and led its preparation.

- J. Karolus, F. Kiss, C. Eckerth, N. Viot, F. Bachmann, A. Schmidt, and P. W. Woźniak. EMBody: A data-centric toolkit for EMG-Based interface prototyping and experimentation. 2021. In Proc. ACM Hum.-Comput. Interac. 5 (EICS) Article 195, 29 pages.

The following publications are based on the supervision of student works. I co-crafted the study design, supported the implementation of prototypes and the study’s evaluation. Further evaluation for the publication was conducted by myself. I wrote most parts of the manuscript and led its preparation.

-
- J. Karolus, A. Kilian, T. Kosch, A. Schmidt, and P. W. Woźniak. Hit the thumb jack! Using electromyography to augment the piano keyboard. 2020. In Proceedings of the 2020 ACM Designing Interactive Systems Conference (DIS '20). ACM, New York, NY, USA, 420-440.
 - Associated demo publication: A. Kilian, J. Karolus, T. Kosch, A. Schmidt and P. W. Woźniak. EMPiano: Electromyographic Pitch Control on the Piano Keyboard. 2021. In Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI EA '21). ACM, New York, NY, USA, Article 196, 1–4.
 - J. Karolus, H. Schuff, T. Kosch, P. W. Woźniak, and A. Schmidt. EMGuitar: Assisting guitar playing with electromyography. 2018. In Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18). ACM, New York, NY, USA, 651-655.
 - C. Liang, J. Karolus, T. Kosch, and A. Schmidt. On the suitability of real-time assessment of programming proficiency using gaze properties. 2018. In Proceedings of the 7th ACM International Symposium on Pervasive Displays (PerDis '18). ACM, New York, NY, USA, Article 31, 2 pages.
 - J. Karolus, F. Bachmann, T. Kosch, A. Schmidt and P. W. Woźniak. 2021. Facilitating Bodily Insights Using Electromyography-Based Biofeedback during Physical Activity. In Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction (MobileHCI '21). ACM, New York, NY, USA, Article 14, 1–15.
 - J. Karolus, J. Sylupp, A. Schmidt, and P. W. Woźniak. 2021. EyePiano: Leveraging Gaze for Data-Driven Piano Tutoring. Submitted for review.

For the following publication, I co-drafted the study design, co-implemented the prototypes, and conducted the evaluation. I wrote most parts of the manuscript and led its preparation.

- J. Karolus, S. S. Feger, A. Schmidt and P. W. Woźniak. 2021. Designing Proficiency Awareness: Leveraging Reflective Feedback in Text Production. Submitted for review.

Eidesstattliche Versicherung

(Siehe Promotionsordnung vom 12.07.11, § 8, Abs. 2 Pkt. 5)

Hiermit erkläre ich an Eidesstatt, dass die Dissertation von mir selbstständig und ohne unerlaubte Beihilfe angefertigt wurde.

München, den 29. April 2021

Jakob Karolus