Dipl.-Ing. Granit Luzhnica

Leveraging Optimisations on Spatial Acuity for Conveying Information through Wearable Vibrotactile Displays

DISSERTATION

for the attainment of the degree of

Doctor of Engineering Sciences (Dr. techn.)

submitted to

Graz University of Technology

Supervisor:
Univ.-Prof. Dipl.-Inf. Dr. Stefanie Lindstaedt

Advisor:
Dr.techn. MSc Eduardo Veas

Graz, May 20, 2019

EIDESSTATTLICHE ERKLÄRUNG

AFFIDAVIT

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Das in TUGRAZonline hochgeladene Textdokument ist mit der vorliegenden Masterarbeit/Diplomarbeit/Dissertation identisch.

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master's thesis/diploma thesis/doctoral dissertation.

Datum / Date

Unterschrift / Signature

Abstract

With the proliferation of wearable devices, the vibrotactile capabilities are accessible to a substantial number of end-users. Currently, the primary utilisation of vibrotactile feedback is to provide additional support to visual channel interaction. Nevertheless, haptic feedback is capable of transmitting rich information without the need to perceive it through auditory or visual channels. Although haptic reading using finger-tips has been used for blind individuals using Braille encoding or simply for information transmission using Morse Code, they both suffer from the practical aspects which makes them hard to adopt in wearable device technologies. In the meantime, several approaches have been proposed to encode information through spatial encoding which mainly reuses a vibrotactile Braille-like encoding in other body parts (e.g. hands, arms). However, spatial stimulation suffers from effects such as masking and haptic illusions which are a result of the low resolution of the skin. Other methods use temporal encoding or variations in amplitude/frequency or complex patterns to encode information which either require a long training period or are very limited in the number of distinct patterns users can distinguish which limits their applicability of them in real-world applications.

This thesis investigates the use of spatial encoding of information by leveraging the spatial acuity of the simulated locations. Moreover, informed by numerous studies on the limitations, perception properties and resolution of the skin, it takes the tasks of designing encoding techniques that avoid as much as possible the drawbacks of the low resolution of the skin or in some cases it leverages them for encoding information. It first proposes overlapping spatiotemporal patterns as an efficient way of encoding discrete abstract symbols and meanings (e.g. letters of the alphabet), which are optimised to deliver good identification accuracy and allow constructing very fast speed messages. The patterns leverage the sensitivity of locations in order to prioritise the activation of actuators which in turn increases the perception

and identification of such patterns. Furthermore, it investigates body locations for suitable for conveying information using wearable displays. Additionally, it investigates conveying of complex messages (e.g. words) by combining two or more simple abstract symbols in a series. It identifies problems with such a conveying of simple discrete and compounds messages, and it proposes methods to overcome them. Additionally, it proposes and evaluates interaction methods for navigating through complex messages. Besides encoding discrete symbols, this thesis investigates conveying continuous values (e.g. continuous numbers) using phantom sensation. It proposes sensitivity adjusted perpetual models which predict more accurately the perception of the user and thus result in a lower error of decoding the encoded value.

Acknowledgments

The work in this thesis would not have been possible without the support and collaboration with many brilliant people.

First of all, I would like to thank my advisor Dr Eduardo Veas, for his fruitful discussions throughout the entire process of this thesis. His guidance went beyond my thesis, into the overall research conducting process, which helped me become the researcher I am today. I appreciate his encouragement and freedom to follow my ideas to the full potential, while occasionally challenging my views in order to bring provide other perspectives. Above all, Eduardo is a great and caring person on a personal level who has been a great mentor along my PhD journey. I would like to also thank my supervisor, Prof. Dr. Stefanie Lindstaedt for the encouraging feedback and the great research opportunities at the know Center.

I also would like to thank Prof. Roderick Murray-Smith for having me at Glasgow University as a visiting researcher in the summer of 2016. The two months of work and collaboration with his team were very fruitful. Furthermore, it is during my time there that I started developing an interest in using optimisation techniques for human-computer interaction, which then later were applied to my research and proved to be very successful approaches.

Further, I would like to thank my friends for their continuous support. My best friend Andi Rexha for the professional and personal discussions, advice and encouragements throughout the thesis. Additionally, I would like to thank my friends Lirim Elshani and Milot Gashi for their great help on recruiting participants for my user studies.

Special gratitude goes also to my participants who dedicated their time and effort for me to be able to collect data and prove my hypotheses. Without their dedication, my findings would not have been possible.

I want to thank my family for the unconditional love, care and support that they have given me throughout my entire life. My mother Sahadate who dedicated her life to her children. She taught me to work hard and aspire to be the best version of myself. My late father Mehmet, who was my early mentor in life. My sister, Dafina, who always cheered me up. My other sister, Donika, who was always caring and supportive. Beyond, that, she helped me a great deal with recruiting participants for the work conducted in this thesis. My brother Enxhell, who despite being much younger than me, always provided good advice for me, including for the work provided in this thesis.

Lastly and most importantly, I would like to thank my wife, Arbenore, who was a part of this journey with me. Arbenore was always there for me to provide incredible encouragement and support, even in my worst moments. Additionally, she played a significant active role in recruiting participants for my user studies. Thank you also to my son Jon, who continuously fills my life with joy and happiness.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Wearable and mobile devices are already a part of our everyday life. They provide assistance to daily activities and enrich them with additional information collected by the sensors within them. The primary feedback modalities of mobiles and wearables are visual and auditory. As such, they compete for the user's visual and auditory attention and distract the user from important tasks. Alternative display modalities, such as tactile displays, can reduce demands on the predominant visual display, but are largely under-utilised [Brewster and Brown, 2004]. With the broad spread of wearable devices, vibrotactile capabilities are accessible to a substantial number of end-users. Currently, the primary utilisation of vibrotactile feedback is to provide additional support to visual channel interaction. Nevertheless, haptic feedback is capable of transmitting rich information without the need to perceive it through auditory or visual channels. The information can be delivered in the form of **tactons** which are defined to be vibrotactile patterns representing abstract meanings [Brewster and Brown, 2004]. Such information could offer many benefits to provide or enhance multitasking, interactions on-the-go and sensory substitution.

Multitasking is common not only in work activities [González and Mark, 2004] but also in runtime daily activities such as eating [Hellmich, 2004]. Interaction with mobile devices requires visual attention and thus creates distractions which on-the-go might even be dangerous. Visual distractions from mobile devices while driving [Chen, 2009, News, 2017] or waking on the street [Christina London and Rascon, 2014] can lead to life-threatening situations for drivers, passengers and or pedestrians. Although several countries have already taken regulatory measures by restricting the use of mobile phone while driving, such regulations seem not to be sufficient as they are not always respected [Chen, 2009, News, 2017]. Providing methods of perceiving information through haptics, might help with multitasking and avoid visual distractions.

For individuals that lack perception channels such as visual or auditory, vision and audio could be captured using technological solutions such as a camera or microphone and then mapped to another sensory modality such as haptic. This method is commonly referred to as sensory substitution. Attempts to provide a sensory substitution for auditory impaired users date back to 1924 [Gault, 1924] where the entire speech signal of spoken words is transformed into vibrotactile stimuli. Since then, such techniques have been used to provide hearing aid solutions which supplement lip reading [Milnes et al., 1996, Yuan et al., 2005, Galvin et al., 1999, Ranjbar, 2008, Rönnberg et al., 1998, Weisenberger et al., 1991, Reed and Delhorne, 2003, Scott and Filippo, 1977, Galvin et al., 2001, Phillips et al., 1994] but they are insufficient to serve as hearing solutions [Novich and Eagleman, 2015]. Substitution of vision has been explored [White et al., 1970, Bliss et al., 1970] where images captured by a camera are imprinted in the body using vibrotactile stimuli. Nevertheless, due to the low resolution of the skin and perception limitations, such methods are limited to conveying very basic shapes (e.g. lines, circles, squares). While the sensory substitution using wearable devices would be of great potential, means of efficiently conveying information through haptics in a wearable form are still a challenge.

The primary focus of this work is to investigate methods for conveying information (e.g., a numerical value or a symbol) using wearable haptic displays, where information is presented to the user via touch, typically using vibrotactile motors (**vibromotors**). This differs a lot from stimulating touch effects that mimic virtual objects for enhancing user experiences. Conveying information can be used, for example, to perceive natural language messages (e.g. words) encoded in vibrotactile patterns [Geldard, 1957, Luzhnica et al., 2016b]. The proposed methods can be applied in a broad range of applications. For instance, users would be able to receive and understand the messages and notification from the mobile phone without even having to get it out of the pocket. Deaf users would be able to use speech to text (captured by a smartphone) and text to tactile to fully understand other persons talking to them. Individuals with hand amputee or prosthesis that lack the tactile and kinesthetic sensation on the hand would be able to receive the magnitude of applied pressure to an object while grasping for a closed loop interaction. Workers in factories could receive tactile instructions allowing them to focus visually and auditory on their work. Numerous other scenarios could benefit from wearable haptic displays, and most importantly many barriers of technology (wireless communication, batteries, integration to fabrics) for making such haptic displays fully wearable are already overcome.

This research spans the creation of wearable haptic prototypes, stimulation methods, information encoding, user training and methods for interacting with such wearable devices. One critical aspect of a haptic display is the encoding of information as it should be optimised to be perceivable, understandable and to have a high throughput. That is why this thesis proposes overlapping spatiotemporal patterns (OST) which are still short but are designed to provide a better discriminability compared to the spatial patterns. A major challenge, when encoding a vocabulary of symbols in a small number of haptic actuators, is maintaining high throughput and accuracy. For that, this work proposes methods which optimise the encoding of information which maximises the perception and the comprehension of information. Training time is also important as it can take several hours to train users to recognise different encoded messages [Geldard, 1957, Luzhnica et al., 2016b]. There are also scenarios where understanding quantitative values approximately is sufficient (e.g. lap progress in a video game, the strength of the grasp). For such scenarios, this work aims at constructing intuitive methods which require conveying information with minimum or no training at all. The process of conveying information through wearable devices is passive and the user has no control of information flow. Thus, it might be beneficial to equip users with interactions and empower them for controlling the flow, which is explored in this work as well.

Overall, this work addresses the following research questions:

(RQ1) Constructing patterns optimised for throughput and perception: Do the overlapping spatiotemporal patterns result in better identification accuracy than the baseline spatial patterns on the hands and forearms?

(RQ2) SkinReading - conveying natural messages: Are overlapping spatiotemporal patterns suitable for vibrotactile skin reading on the hands and forearms? More specifically, what performance on the recognition of letters and words can participants achieve with few hours of training? **(RQ3) Conveying inaccuracy tolerant quantitative values**: For scenarios where high precision is not required, can we encode continuous values using a discrete number of actuators using phantom sensation? More precisely, how well (with what accuracy) can values be decoded by users and does sensitive adjustment increase such encoding/decoding accuracy? **(RQ4) Interactions for skin reading**: What interactions are necessary for skin reading? What are the preferred modalities for such interactions?

1.1 Addressing Research Questions

This section describes how I addressed the research question of the thesis. Figure 1.1 presents the relations of the research questions and illustrates in which cases the output of one research question informs the others.

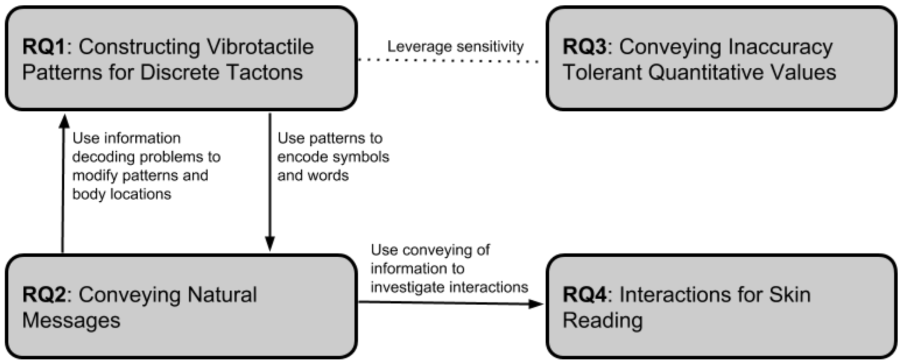

Figure 1.1: An overview of the research questions of this thesis and their flow of information. The patterns and body locations proposed in RQ1 are used to encode symbols and then words for investigating RQ2. The results of RQ2 are used to reiterate RQ1 in order to find other suitable body positions for OST patterns. The findings of RQ2 (the need for repetition) first motivate the RQ4 and then the methods of conveying information in RQ2 are used as a basis for investigating RQ4. RQ3 is detached from the rest of research questions as it still deals with conveying values but not discrete ones as in the case of RQ1 and RQ2. However, both RQ1 and RQ3 share the same vision of leveraging the sensitivity of location (spatial acuity) to improve the performance in perception.

In summary, ten user studies were conducted to investigates the proposed research questions and some extensive details related to them. For each of the two first research questions, four user studies were conducted, until the overlapping spatiotemporal patterns were fully investigated (including the effects of sensitivity prioritisation), the wearable vibrotactile display layouts and the encoding were optimised and resulted in a very high comprehension accuracy from participants. Figure 1.2 sketches the relationship between research questions and user studies in chronological order. In addition, it provides information on how user studies inform each other. Moreover, as illustrated in Figure 1.2, the first two research questions follow an iterative process, where the User Study 1 (in Section 3.1) conducted to investigate the first research question informs the User Study 5 (in Section 4.1) conducted for the second research question. However, due to issues found in the User Study 5, the process is iterated, and more studies were conducted to reiterate layouts and also patterns investigated in the first research question which then are used to inform other studies for the second research question.

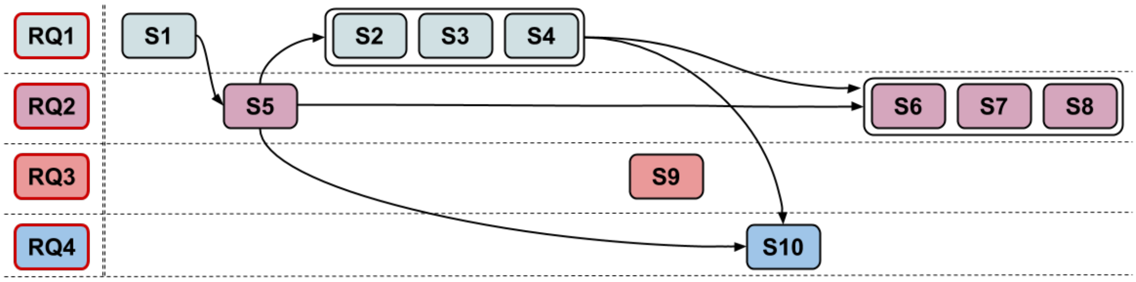

Figure 1.2: An overview of the studies, their flow of information and their contribution to particular research questions. The x-axis represents the chronological order, which could be thought as the time of execution but it is not linear. The arrows illustrate how some user studies build on top of the evidence and knowledge produced by other user studies. Note that the main interaction is between research questions 1 and 2 as I go back and forth to iterate the proposed solutions until the results are improved at a satisfactory level.

1.1.1 RQ1: Constructing vibrotactile Patterns Optimised for Throughput and Perception

Do the overlapping spatiotemporal patterns result in better identification accuracy than the baseline spatial patterns on the hands and forearms?

This research question takes the challenge of finding vibrotactile patterns which are optimised to be short so that they can allow high throughput when combined in complex messages and at the same time are highly distinguishable so that they can be identified correctly. This problem is addressed in Chapter 3 where the prioritised overlapping spatiotemporal (OST) patterns are proposed as a suitable alternative on maximising throughput and accuracy. Four user studies investigate the accuracy of such vibrotactile patterns compared to spatial patterns and also investigate suitable body positions for such vibrotactile patterns.

The first user study (in Section 3.1) proposes and evaluates overlapping spatiotemporal (OST) in comparison with spatial patterns. It also proposes and investigates three wearable vibrotactile layouts for perceiving such patterns. The results of this study favoured two wearable layouts as being suitable for the perception of such patterns. It also revealed that OST patterns are perceived more accurately than spatial patterns. A second user study (in Section 3.2) proposes and evaluates intensity varying spatial patterns, aiming at achieving the perception performance of OST patterns and at the same time reducing the duration of stimulation. The study revealed that varying the intensity of the individual locus of spatial patterns does not contribute to an increase in perception.

Additionally, a third user study (in Section 3.3) builds on the hypothesis that the correct perception of such vibrotactile patterns can be boosted by prioritising the activation of vibromotors based on spatial acuity (sensitivity of the location). Results show that indeed prioritising the activation of vibromotors based on the sensitivity of locus increases the identification accuracy significantly. The fourth user study dealing with RQ1 (in Section 3.3), extends one of the hand based layout proposed in the user study 1 (in Section 3.1) by adding more actuators and then investigates the perception of OST patterns in the proposed new layout. The end result of this user study is a layout with eight suitable locations of actuators which are suitable for perceiving OST patterns accurately.

1.1.2 RQ2: SkinReading - Conveying Natural Messages

Are overlapping spatiotemporal patterns suitable for vibrotactile skin reading on the hands and forearms? More specifically, what performance on the recognition of letters and words can participants achieve with few hours of training?

The second research question investigates the feasibility of using such OST vibrotactile pattern to first encode discrete tactons representing the letters of English Alphabet and then investigate whether such symbols can be combined to form more complex messages such as words and phrases. This thesis first proposes an encoding of the English Alphabet to OST vibrotactile patterns. In four user studies presented in Chapter 4, users are trained to recognise the vibrotactile alphabet. Then, the studies investigate how well users can recognise symbols encoded by OST patterns, how well users can read words when such symbols are combined into words. Moreover, they identify problems related to the encoding of the symbols and construction of patterns and reiterate the entire process in order to maximise the recognition accuracy.

The first user study (in Section 4.1) proposes an encoding for all the letters of English Alphabet which maps every letter to an OST pattern. The encoding uses the frequency of the letters in the English language to construct an efficient encoding scheme. In addition, the study proposes a training program which is used to teach participants the proposed encoding. The study evaluates the performance of participants on recognition of letters and words after training using wearable vibrotactile displays on the hands and forearms. Its results show that participants are able to comprehend the information with a relatively high accuracy but also reveals that there are potential improvements related to layout and encoding.

Thus, the second user study (in Section 4.2) proposes a two-step optimisation process which optimises the layout and encoding. Furthermore, it evaluates the impact of such optimisation, showing drastic improvements in the comprehension of letters and words. Moreover, it investigates the decay of encoding knowledge over time as well as the transferability of encoding knowledge on the untrained body location.

When considering skin reading for real-world applications, several practical aspects should be taken into account as are crucial in defining its scope of applicability. For the skin reading to be useful in multitasking scenarios, it should be feasible to use it in parallel with other user activities. Therefore, the third user study (in Section 4.3) investigates whether participants can comprehend vibrotactile encoded symbols in the background while performing other tasks. Another vital factor to consider is the training procedure. Training is typically time-consuming and requires active participation by users which might pose challenges on adoption of such a technology. Thus the fourth user study (in Section 4.4) investigates the use of passive haptic learning for teaching users the vibrotactile skin reading. Such a method allows users to be trained while enjoying other activities (e.g. playing video games).

1.1.3 RQ3: Conveying Inaccuracy Tolerant Quantitative Values through Wearable Vibrotactile Displays

For scenarios where high precision is not required, can we encode continuous values using a discrete number of actuators using phantom sensation? More precisely, how well (with what accuracy) can users decode such encoded values and does sensitivity adjustment increase such encoding/decoding accuracy?

The third research question moves away from the discrete symbols and focuses on use cases where continuous numerical values need to be presented to the user. In discrete tactons mistaking two neighbouring values (e.g. A for B) is considered to be a high error. On the contrary, when dealing with numerical values, the magnitude of the error is very important. For instance, mistaking the value of 67% for 65% or 70% might not be a severe problem, but mistaking it for 15%, might be. Thus this research question targets use cases which are tolerant to some dose of inaccuracy.

The primary goal would be to develop a method to convey such continuous numerical values inspired by the existing visual presentation of such values (e.g. bar chart, progress bar) which would require no training on learning how to decode such values. Chapter 5 presents a user study which investigates how well the phantom sensation with the existing perceptual models that describe it, could be used to encode continuous numerical values. This chapter also proposes and evaluates sensitivity adjusted perceptual models that can be used to estimate the perception of the user better and thus increase the accuracy when decoding the value. Additionally, it proposes a data-driven approach to estimate the parameters (sensitivities of the locations) needed for sensitivity adjusted models. The results show that not only the phantom sensation could be used for encoding continuous numerical values, but also that the proposed sensitivity adjusted models can significantly increase the accuracy of comprehension.

1.1.4 RQ4: Interactions for Skin Reading

What interactions are necessary for skin reading? What is the preferred modality for such interactions?

When perceiving information there is a need for repeating or navigating them. Users might need to repeat certain words from time to time as a result of attention breaks or simply due to misperception. In both visual and Braille reading such interactions occur very often even if readers are not aware of it as it occurs unconsciously and naturally. Readers jump backwards to revisit already visited letters and words [Larson, 2004, Rayner, 1998, Rayner et al., 2001, Rayner et al., 2010]. This phenomenon is known as back regression, and skilled readers make regressions back in $10 - 15\%$ of the reading time [Rayner, 1998, Rayner et al., 2001, Rayner et al., 2010]. Such regression is common practice also in Braille reading [Millar, 2003, Hughes et al., 2011]. In both cases, no technological solution is needed to provide these interactions as it comes naturally. In visual reading, this is achieved easily by shifting the fixation point visually. Similarly, in Braille, this can be easily achieved by moving the finger backwards. However, in vibrotactile skin reading, the user does not control what information is currently being conveyed, and thus interactions need to be provided to enable such control.

A user study presented in Chapter 6 investigates what interactions are necessary and what is the preferred modality of such interactions. The study reveals that users would prefer hand-based gesture interaction for such interactions. Then it maps the skin reading interactions to a set of meaningful hand gestures. Finally, it investigates what kind of sensors would be required to recognise gestures corresponding to interactions of skin reading.

1.2 Scientific Contributions

Overall this thesis provides guidelines for conveying information through wearable vibrotactile displays. This research spans the creation of wearable haptic prototypes, stimulation methods, information encoding, user training and methods for interacting with such wearable devices. The outcomes of this research are relevant to several communities including human-computer interaction, wearable computing, and psychophysics. A detailed list of contributions is given in the following:

(C1) Vibrotactile Patterns. Chapter 3 proposes overlapping spatiotemporal patterns (OST) which are shorter than sequential temporal patterns and can be identified more accurately than spatial patterns. Moreover, this thesis proposes to prioritise the activation of vibromotors based on spatial acuity to maximise the perception and identification accuracy.

(C2) Wearable Vibrotactile Display Design. Throughout this thesis, four layouts on the back of the hands and forearms are designed and evaluated for skin reading with the final one (in Section 4.2) being optimised for OST patterns which is very suitable for skin reading. Additionally, four other layouts on the forearm and upper arm are proposed to encode continuous numerical values using phantom sensation (in Section 5.3). Such layouts are able to encode continuous values of directional or circular nature.

(C3) Methods for Conveying and Encoding Letters of English Alphabet and Textual Information. Initially, a frequency based encoding of the English alphabet is proposed. Later another optimised encoding is proposed which delivers outstanding accuracy in terms of being comprehended by users. In addition, this thesis proposes methods for combining letters into words to form textual information.

(C4) Methods for Optimising the Encoding for a Given Language. Section 4.1.6 proposes a method and an algorithm to optimise the encoding of a language based on the bigram frequency of the given language. An evaluation reveals that the proposed optimised encoding delivers outstanding accuracy in terms of being comprehended by users.

(C5) Training Methods for Learning Skin Reading. Chapter 4 investigates different methods of training. It first (in Section 4.1) proposes an active training which uses visual, auditory and vibrotactile cues in order to train participants to associate the vibrotactile patterns with the encoded information. Then (in Section 4.4) it also investigates using passive haptic learning as a training method for skin reading. Section 4.1 also proposes a training program for teaching participants the skin reading.

(C6) Background Perception of Vibrotactile Symbols. Section 4.3 demonstrates that the vibrotactile messages (encoded by the proposed methods) can be comprehended in background while performing other primary tasks. Moreover, their performance is not affected by the absence of presence a primary task and vice versa. The performance of the primary task is not affected by the presence or absence of stimuli of background vibrotactile encoded symbols.

(C7) Knowledge Transferability of the Encoding in Untrained Body Parts. Section 4.2 investigates and evidences that the knowledge of encoding acquired during the training can be transferred to an untrained body part without any training. In other words, participants can be trained on one hand, and the use the wearable vibrotactile device on the other hand and still be able to recognise the encoded letters without any additional training.

(C8) **Methods for Conveying Continuous Numerical Values**. Chapter 5 proposes a method of using phantom sensation to encode a continuous numerical value in a chain of vibrotactile motors.

(C9) **Extending Perceptual Models for Phantom Sensation**. Chapter 5 extends the existing state of the art perceptual models by proposing sensitivity adjusted perceptual models which are better at estimating the perceived stimuli when applying the phantom effect. In turn, this increases the accuracy of comprehension for the encoded value using phantom sensation.

(C10) **Interaction Techniques for Skin Reading and Gesture Recognition**. Chapter 6 proposes to incorporate interactions during the process of skin reading. In addition to proposing and evaluating an interaction model, it proposes a mapping of such interaction to intuitive gestures which then can be recognised using wearable sensors.

(C11) **Wearable Sensory Substitution using Mobile Devices and Skin Reading**. This thesis proposes to use mobile phones to provide a solution for sensory substitution in combination with wearable vibrotactile devices. The solution uses mobile devices to capture the environment which then is processed, and its content is recognised using machine learning models. The recognise content is represented in textual form and then transmitted to the user through a wearable vibrotactile display. Besides the conceptual work, such concepts are also implemented in the form of mobile applications.

(C12) **Hand Gesture Recognition System**. In order to be able to interact with the skin reading using gestures, a gesture recognition system is constructed. A brief description of the methodology is given in Section 6.3.

The aforementioned contributions have been published in eight scientific peer-reviewed papers and one peer-reviewed poster described in the following:

(P1) **Luzhnica, G.**, Veas, E., and Pammer, V. (2016b). Skin Reading: Encoding Text in a 6-channel Haptic Display. *In Proceedings of the 2016 ACM International Symposium on Wearable Computers, ISWC '16*, pages 148- 155, New York, NY, USA. ACM.

This paper lays the foundation of the vibrotactile skin reading. It first introduces the overlapping spatiotemporal patterns (OST) and compares them against the sequentially spatial patterns in three wearable layouts. Moreover, it performs the first test of skin reading with four participants and shows that skin reading is feasible using OST patterns. The work presented in this paper includes the user study 1 described in the Section 3.1 and it serves as a template for the user study 5 described in Section 4.1.

(P2) **Luzhnica, G.** and Veas, E. (2017). Vibrotactile Patterns using Sensitivity Prioritisation. *In Proceedings of the 2017 ACM International Symposium on Wearable Computers, ISWC '17*, pages 148-155, New York, NY, USA. ACM.

This paper investigates more extensively the vibrotactile patterns for skin reading. It investigates the effect of sensitivity prioritisation in the perception of OST patterns. Additionally, it extends the hand based layout of vibrotactile wearable display already proposed in **P1**. The work reported in this paper includes user studies 2 (Section 3.2), 3 (Section 3.3) and 4 (Section 3.4).

(P3) **Luzhnica, G.** and Veas, E. (2019). Optimising the Encoding for Vibrotactile Skin Reading. *In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* ACM.

This paper builds upon **P1** by borrowing the skin reading training program and layouts proposed in **P1**. It first evaluates two wearable layouts with a frequency based encoding as proposed in **P1** with the goal of identifying systematic errors on recognition of letters and words which are related to wearable layout and encoding. Upon identification of such issues, it proposes a two-step optimisation process which optimises the layout and encoding. A second user study shows that such optimisation is very beneficial as drastically improves the ability of participants to recognise letters and words. The content of this paper is covered in Sections 4.1 and 4.2.

(P4) **Luzhnica, G**. and Veas, E. (2019b). Background Perception and Comprehension of Symbols Conveyed through Vibrotactile Wearable Displays. *In 24rd International Conference on Intelligent User Interfaces, IUI '19.* ACM.

This paper investigates how background perception of vibrotactile encoded messages affects other primary tasks. A user study shows that pre-trained participants were able to perform dual tasks where the primary task was a visual search task, and the secondary task was a comprehension of the vibrotactile encoded message. The study showed that participants were able to perform both of them with high accuracy and none of them affected each other. The content of this paper is described in the Section 4.3 and it illustrates the potential of skin reading in multitasking and multimodal scenarios.

(P5) **Luzhnica, G**., Veas, E., and Seim, C. (2018). *Passive Haptic Learning for Vibrotactile Skin Reading. In Proceedings of the 2018 ACM International Symposium on Wearable Computers, ISWC '18.* ACM.

This paper explores the possibility of using passive haptic learning (PHL) to train the encoding of skin reading while users perform other activities (play a game in this case). The primary goal is to make the training more engaging and less tiresome. A study shows that participants were able to learn several letters using PHL training, although not to the expected extent. The content of this paper is described in the Section 4.4.

(P6) **Luzhnica, G**., Stein, S., Veas, E., Pammer, V., Williamson, J., and Smith, R. M. (2017). Personalising Vibrotactile Displays through Perceptual Sensitivity Adjustment. *In Proceedings of the 2017 ACM International Symposium on Wearable Computers, ISWC '17*, pages 66-73, New York, NY, USA. ACM.

This paper proposes an encoding approach of the continuous numerical values inspired by visual progress bars. Such an encoding does not require training to memorise the meaning of encoding. However, it can only be used in situations where a degree of imprecision in the decoding of information is tolerable. The encoding of information combines a chain of vibromotors and the phantom effect to provide a sensation across the entire wearable vibrotactile display. Furthermore, this paper also extends state of the art perceptual models of phantom sensation by leveraging the spatial sensitivity. Such models result in a better decoding accuracy according to a study presented in Chapter 5.

(P7) **Luzhnica, G**. and Veas, E. (2018a). Investigating Interactions for Text Recognition using a Vibrotactile Wearable Display. *In 23rd International Conference on Intelligent User Interfaces, IUI '18*, pages 453-465. ACM.

This paper proposes to extend the interaction of skin reading by providing users means of interacting with the wearable vibrotactile displays. It proposes and evaluates an interaction concept that enables users to control the flow and navigate the presented text. Moreover, it evaluates preferred modality of such interaction. A study shows that users prefer gesture interaction as means of interacting with the device. Thus, the paper also maps the interaction to gestures and explores motions sensors and machine learning to provide a gesture recognition system to recognise such interactions. Its content is covered in Chapter 6.

(P8) **Luzhnica, G**. and Veas, E. (2018b). **(POSTER)** Skin Reading Meets Speech Recognition and Object Recognition for Sensory Substitution. *In Proceedings of the 2018 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2018 ACM International Symposium on Wearable Computers, UbiComp '18*, New York, NY, USA. ACM.

This work presents two concepts of mobile applications that target users with visual and hearing impairment using sensory substitution. Such concepts are also implemented (iOS and Android) and use object and speech recognition to recognise the environment surrounding the user and then provide such information to the user in the form of skin reading. The main goal of this work is to illustrate the potential of skin reading, and its content is described in Section 4.5.

(P9) **Luzhnica, G**., Simon, J., Lex, E., and Pammer, V. (2016a). A sliding window approach to natural hand gesture recognition using a custom data glove. *In 2016 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 81-90. IEEE.

This paper uses hand worn sensors to provide a gesture recognition system. This work does not directly contribute to the topics covered in this thesis. However, its gesture recognition system is borrowed by **P7** to enable users to control the information flow in the skin reading. Thus, due to the indirect contribution, the work of this paper is considered to be out of scope and not covered in this thesis. However, the content of this paper is provided in the Appendix (see Chapter A).

1.3 Collaborative Statement

The work described in this thesis and the aforementioned published paper has been achieved through the collaboration with other brilliant researchers. Their collaborations are summarised in the following:

Eduardo Veas: from Know Center and Graz University of Technology, provided advice, guidance and constructive feedback throughout all stages the work of this thesis. He also closely collaborated in the designing of the studies 1, 5 and 8 (Sections 3.1, 4.1 and 4.4). Additionally, with his expertise in statistical analysis, he guided and collaboratively contributed in analysing the results of the user studies 1 and 5 (Sections 3.1 and 4.1). He also collaborated in the writing of papers **P1**-**P8** and provided constructive feedback on different stages of the work presented in the **P9**.

Viktoria Pammer: from Know Center and Graz University of Technology, contributed with valuable feedback and discussions as well as participated in the writing of the papers **P1**, **P6** and **P9** which are part of Sections 3.1, 4.1 and Chapter 5.

Sebastian Stein, John Williamson and Roderick Murray Smith: from Georgia Tech University contributed with valuable feedback and discussions as well as in the writing of the paper **P6** which is now a part of Chapter 5.

Caitlyn Seim: from Georgia Institute of Technology with her expertise on passive haptic learning closely collaborated in the design of the user study 8 (provided in Section 4.4) which is published in the paper **P5** as well as the writing of the paper (**P5**).

Christopher Öjeling: built an Arduino based hardware prototype connected to nine vibromotors that could be controlled programmatically. A subset of such vibromotors and the device in all of the user studies described in this thesis.

1.4 Structure

The rest of this thesis is first followed by Chapter 2 which provides a comprehensive review of the background and related work. This is followed by Chapter 3 which aims at constructing patterns that are optimal for perception and throughput. This section presents four user studies which propose investigate overlapping spatiotemporal vibrotactile patterns and their perception details to collectively contribute to the Research Question 1.

In Chapter 4 the feasibility of skin reading using the overlapping spatiotemporal patterns is studies which contribute to the Research Question 2. Next, Chapter 5 investigates Research Question 3 by exploring the possibility of conveying continuous numbers using phantom sensation on a wearable vibrotactile display. Lastly, Chapter 6 takes the task of exploring the Research Question 4 and thus it proposes and investigates the incorporation of interaction for skin reading.

Chapter 2

Background and Related Work

Skin is considered to be the largest organ of the human body with dimensions of $1.5 - 2m^2$ in adults. The receptors within it are responsible for perceiving tactile sensations such as pressure, texture, puncture, temperature, softness, wetness, shape, edges and other details of the environment or the objects we interact with including vibrations. Given such sensing capabilities, researchers have seised the opportunity to use the skin as input medium thus utilise the haptic perception for human computer interaction purposes and to provide haptic interfaces.

This chapter initially provides the basic background information related to the tactile sensation starting from somatic sensory system and then moving to higher abstractions of perception, its properties and limitations. Additionally, it describes the related work and state of the art relevant for skin reading including construction of vibrotactile patterns, conveying information through means of vibrotactile patterns as well as reading patterns for other means of reading.

2.1 Somatic Sensory System

The somatic sensory system is a complex part of the sensory system, and it is responsible for the sensation of touch, pressure, pain, heat, position (limbs), movement, and vibration, which arise receptors within the skin or muscles [Purves et al., 2008]. Receptors differ in many aspects such as the type of stimuli they specialise, the receptive field, the sensation dynamics, etc...

For the sense of touch, **mechanoreceptors** fire and notify the nervous system when the skin is deformed. However, the different nuances of skin deformation will be captured by different receptors as they differ in temporal dynamics and the stimulation response. When stimulated, some mechanoreceptors fire rapidly but then discontinue even in the presence of continuous stimulation. They are known as rapidly adapting mechanoreceptors and are very good at perceiving continuous stimulus such as a movement across the skin. On the other hand, slowly adapting ones fire continuously as long as the stimulus is present. They can be very efficient at the perception of spatial aspects of stimulus including size and shape [Purves et al., 2008]. An illustration of the reaction of both slowly adapting and rapidly adapting mechanoreceptors in the presence of stimulus is given in Figure 2.1.

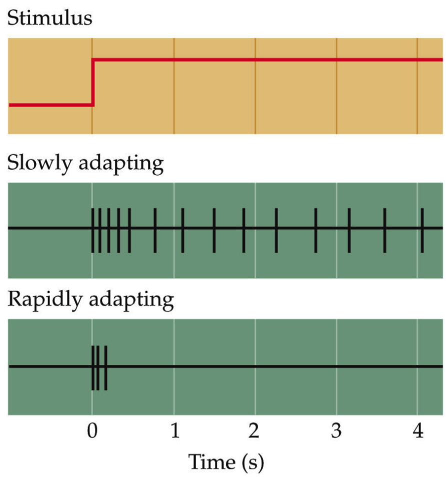

Figure 2.1: The response of slowly adapting mechanoreceptors versus the rapidly adapting ones when presented to a continuous stimulus [Purves et al., 2008].

Besides the adaptations, mechanoreceptors differ on other aspects such as receptive field and their location within the skin. Thus, there are four classes of mechanoreceptors that are specialised in receiving tactile information:

- **Merkel cells** are slowly adapting mechanoreceptors that are essential for light touch sensation. They represent about 25 % of the mechanoreceptors in hand and are unusually dense in the fingertips or similar sensitive skin regions. They have the highest spatial resolution of all mechanoreceptors enabling the sensing of spatial details down to 0.5 mm. Thus, they are very well suited to form sensing shapes and textures [Purves et al., 2008].

- **Meissner corpuscles** are rapidly adapting mechanoreceptors that lie very close to the surface of the skin. Given the close proximity to the skin, Meissner corpuscles are very sensitive, four times more sensitive compared to Merkel cells. However, their receptive fields are larger compared to Merkel cells, making them less effective in sensing the spatial aspect of stimulus. Meissner corpuscles are very efficient in sensing low-frequency vibrations (3-40 Hz) [Purves et al., 2008], making them a good fit for detecting textured objects moving across the skin. They are also tailored for sensing light touch.

- **Ruffini endings** are slowly adapting mechanoreceptors located in the cutaneous tissue and the least understood mechanoreceptors. They are very responsive to internally generated stimuli such as limb movements. Thus, they are well suited to provide kinaesthetic sensation and control of finger position and movement.

- **Pacinian corpuscles** are rapidly adapting mechanoreceptors especially sensitive to vibrations and high pressure. They are able to sense even centimetres away from the stimuli. Pacinian corpuscles have a lower response threshold than Meissner corpuscles and adapt even more rapidly. The most sensitive Pacinian corpuscles can sense even small skin displacements of 10 nanometers. However, they often have large receptive fields with overlapping boundaries. Their optimal sensitivity is at 250 Hz. Due to their properties, they are very effective on detecting vibrations transmitted through objects that contact the hand e.g grasping. Due to high sensitivity on vibrations, they are commonly used on vibrotactile related applications (by transmitting vibrations at their optimal frequency).

Figure 2.2 illustrates the four types of mechanoreceptors. It also shows their adaption properties. Additionally, Table 2.1 provides detailed properties of each mechanoreceptor. Note that even when two types of receptors are rapidly adapting they differ a bit on nuances of adoption (FA1 and FA2) as presented in Figure 2.2. The same holds for slowly adapting receptors (SA1 and SA2). In addition to four types mechanoreceptors, there are receptors that lack any specialisation and thus are referred to as **free nerve endings**. They play an important role in the sensation of pain.

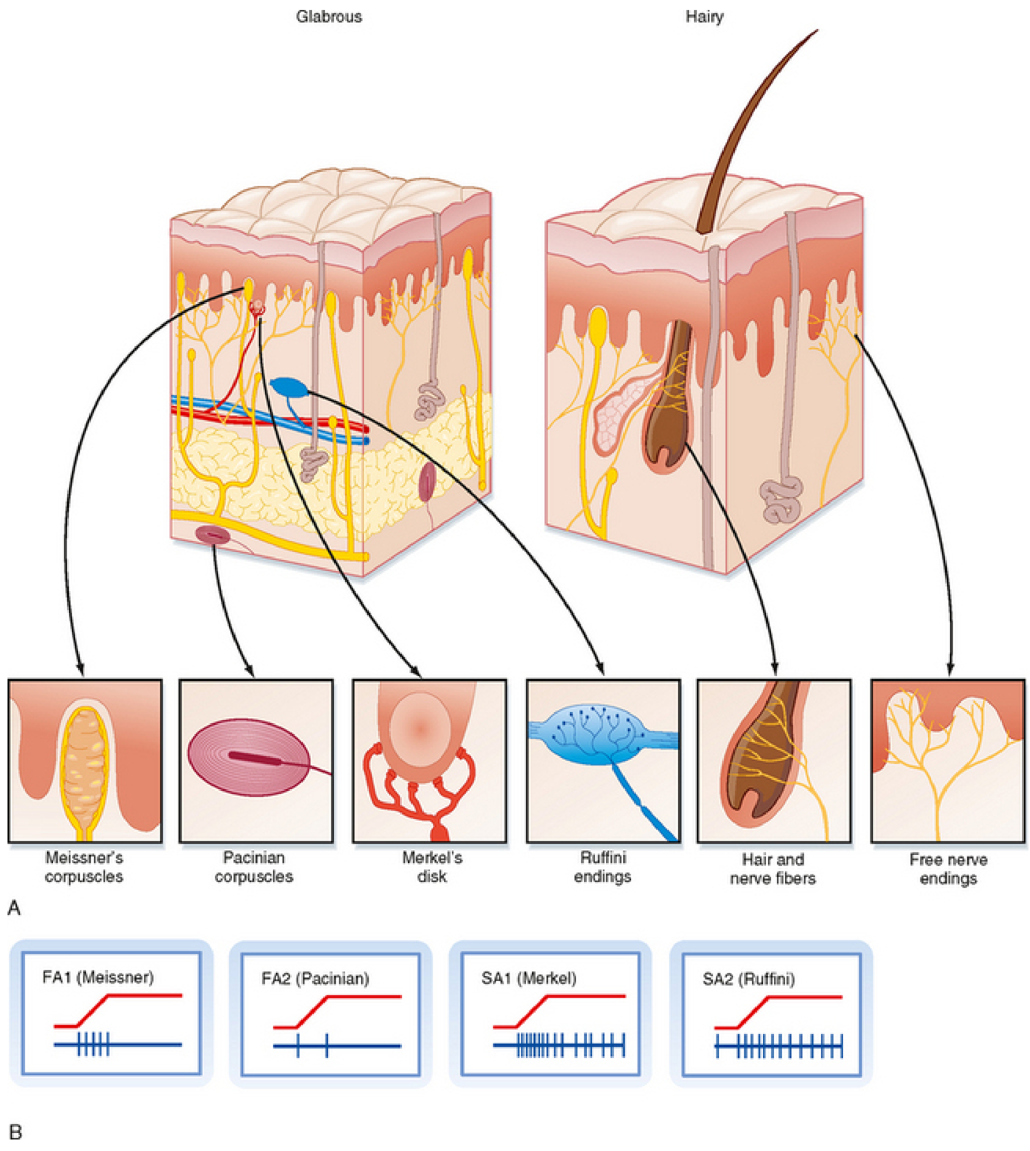

Figure 2.2: Different types of mechanoreceptors and their details [Purves et al., 2008].

	Small receptor field		**Large receptor field**	
	Merkel	**Meissner**	**Paclnian**	**Ruffini**
Axon diameter	7-11 μm	6-12 μm	6-12 μm	6-12 μm
Sensory function	Form and texture perception	Motion detection; grip control	Perception of distant events through transmitted vibrations; tool use	Tangential force; hand shape;
Effective stimuli	Edges, points, comers, curvature	Skin motion	Vibration	Skin stretch
Spatial acuity	0.5 mm	3 mm	10+ mm	7+ mm
Response	slow adaptation	rapid adaptation	rapid adaptation	slow adaptation
Frequency range	0-100 Hz	1-300 Hz	5-1000 Hz	0-? Hz
Peak sensitivity	5 Hz	50 Hz	250 Hz	0.5 Hz

Table 2.1: Properties of different receptors based on [Purves et al., 2008].

2.2 Tactile Perception

Tactile perception can be provided by both passive or active sensing which sometimes are referred to as active and passive touch. Active sensing is provided by exploratory touching and it utilises such a movement to direct the sensory system, in a useful way for the given task, to maximise the information gain [Prescott et al., 2011]. Such active sensing, where the moment is integrated, is crucial for determining object's properties such as shape, texture, hardness, etc... [Lederman and Klatzky, 1993, Robles-De-La-Torre and Hayward, 2001]. On the other hand, passive sensing is defined to be the stimulation produced by external elements. Both active and passive touching have been used for information transmission.

A very successful application of the active touch sensing is Braille reading [Braille, 1829], which is standard form of reading for visually impaired individuals. There, the characters have rectangular blocks called cells composed of up to six raised dots. For reading, users scan the written text by moving the finger horizontally to precieve the lines of text. There, the movement is essential to reading [Millar, 2003, Hughes et al., 2011, Millar, 2004] as it enables users to scan the text, control the speed and occasionally re-scan the text by moving the finger backwards to revisit the information [Millar, 2003, Hughes et al., 2011, Millar, 2004].

Passive touch is used in most of the **skin reading** applications [Luzhnica et al., 2017, Zhao et al., 2018, Chen et al., 2018b, Reed et al., 2018, Jiao et al., 2018, Cauchard et al., 2016, Zhao et al., 2018, de Jesus Oliveira and Maciel, 2014], where information is presented in the form of vibrotactile patterns stimulated by actuators. Such systems are very suitable for wearable tactile displays as the information can be conveyed dynamically. One disadvantage of using passive touch for presenting information is that a user has no control over the transmission of the information or its flow. The patterns of vibrations are stimulated by the device from start to end without any action by the user. This might be an issue in cases where the user may not understand parts of the information due to lack of concentration or training. Thus external means of interaction should be provided to enable the user to control the information flow.

When using tactile sensation for conveying information through tactile displays, it is crucial to consider perception properties and also limitations of the skin in the design process of such a display. Therefore in the following part, I will first provide a summary of perception properties and then describe some of the important limitations of the somatic sensory system. They will be considered throughout this thesis for the design and implementation decisions of the wearable haptic display prototypes that used to conduct the user studies.

2.2.1 Perception Properties

When providing a tactile stimuli, many properties of stimuli need to be considered as they affect the perception. Stimuli can be provided and perceived by means of various aspects such as intensity, frequency, time/duration (temporal, location (spatial) and a combination of space and time/duration (spatiotemporal).

Frequency

As show in Table 2.1 humans can perceive vibrations up to 1000 Hz with a peak sensitivity is at 250 Hz [Gunther, 2001]. Goof [Goff, 1967] investigated the discrimination threshold on the finger and it showed that the threshold depends on the stimulation frequency. For lower frequencies (less than 25 Hz), the threshold was 5 Hz whereas in the higher frequencies (greater than 320 Hz), the discrimination threshold increases. Along the same lines, Brewster et al. [Brewster and Brown, 2004] found that discriminating frequencies is more accurate when presented in relative way rather than absolute. Due to such nonlinearity and the interaction with amplitude, it is difficult to estimate the discriminable levels, but Gill [Gill, 2003] suggest that no more than nine levels should be used.

Intensity

Different parts of the body have different sensitivities. Thus, for the same stimuli, it varies from the location to location whether the stimuli will be perceived at all and the magnitude of the perception. Sensitive locations such as fingertip are estimated to require only about 10 microns of indentation [Kaczmarek et al., 1991] in order to be perceived as a stimulus. The perception threshold also depends on the frequency of vibration. The minimal vibration thresholds are found to be around the frequency of 250 Hz [Lofvenberg and Johansson, 1984], whereas the highest detection thresholds are found to occur in lower frequencies of vibration [Lofvenberg and Johansson, 1984]. An important concept of intensity perception is the so called *just noticeable difference* (**JND**) which defines the minimal change in amplitude that is noticeable by the tested subject. Gunther [Gunther et al., 2002] reported values ranging from 0.4dB to 3.2dB. Therefore, Grill [Gill, 2003] recommends that only up to four intensity values be used when presenting a stimuli in order to maintain the discrimination of them. However, the location of stimuli, the frequency and absolute values of amplitude will have a huge effect on this due to the interaction of such parameters.

2.2.2 Time/Duration

Humans can recognise two consecutive stimuli separated by only 5 ms apart as found by Cholewiak et al. [Heller and Schiff, 2013] which is 5 times faster than vision (25 ms) [Sherrick and Cholewiak, 1986]. Gescheider [Gescheider, 1966] reported that the threshold for detecting *clicks* on the fingerprint is 10 ms. The duration of stimuli has also an impact on the nature of perception. The vibrotactile stimuli shorter than 100 ms is perceived as jabs or taps whereas longer stimuli is perceived as smoothly flowing tactile phrases [Gunther, 2001].

2.2.3 Location

Different body parts have different sensitivity and spatial acuity. This can be seen also to the proportion of the cortex dedicated to processing their sensory input presented in Figure 2.3. Finger tips are regarded as body parts with highest vibrotactile sensitivity due to the high sensitivity to small amplitudes and their high spatial acuity [Craig, 1982]. However, in vibrotactile applications, they are usually discarded as they are very important for object manipulation and interaction with real world objects. That is why this thesis mainly focuses on the back of the hand (including fingers) and forearms.

Among the fingers (not only finger tips), they also differ in sensitivity. Different studies [Duncan and Boynton, 2007, Vega-Bermudez and Johnson, 2001, Hoggan et al., 2007] have shown that the sensitivity decreases from the index finger towards the little finger: the index finger is more sensitive than the middle, ring, and pinky finger. The thumb has the lowest sensitivity [Sterr et al., 2003]. For the forearms the sensitivity is higher near anatomical reference points such as elbow and wrist compared to the middle part of forearm [Cholewiak and Collins, 2003].

Figure 2.3: The cortical homunculus: representing the human body where the body areas are enlarged to correspond to the proportions cortex dedicated to processing sensory and motor functions[2].

Besides the amplitude thresholds and JND, there are other important measures that are important for the spatial aspect of stimulation such as point localisation and two point discrimination. **Point localisation** defines the accuracy to localise the stimulation point. **Two point discrimination** defines the minimal distance required for two stimulations to be recognised as separate stimuli. The thresholds of point localisation and two point discrimination for different body locations are presented in Figure 2.4. It is no coincidence that that the larger areas depicted in the cortical homunculus presented in the Figure 2.3 correspond to smaller point localisation and two point discrimination (in Figure 2.4).

The properties mentioned above of the skin should be considered very carefully when designing systems that use any spatial patterns. For instance, when using different levels of amplitude to encode information, then amplitude thresholds and JND should be an essential guide on designing patterns. On the other hand, when using a spatial encoding (as it is the case with the work in this thesis), two-point discriminations should be considered in deciding the design and layout of the actuators. Otherwise, different simultaneous stimulations might not be perceived as intended (see Section 2.3).

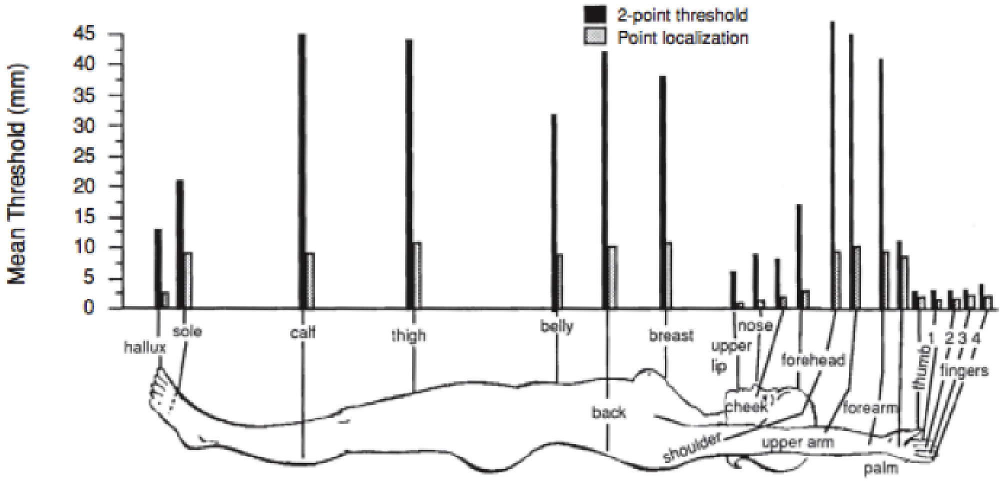

Figure 2.4: Point localisation and two point discrimination thresholds for different parts of the body [LEDERMAN, 1991].

2.3 Sensation Phenomena

Humans haptic perception has, depending on body parts, a relatively low spatial resolution. Due to such limitations of the skin, especially when multiple stimuli are applied, there occur some unexpected sensation phenomena. Such limitations and phenomena should be carefully considered when designing vibrotactile systems, so either to avoid or leverage them for creating special effects. Such phenomena are described in the following.

2.3.1 Masking

Masking effects occurs when two stimulations are presented concurrently and one stimuli (masker) hinders the perception of the other (masked) one [Cholewiak and Craig, 1984, Craig, 1982, Craig, 1983]. Delaying one of the stimuli, increasing the duration or increasing the spatial distance between stimuli can prevent such effect from occurring [Cholewiak and Craig, 1984].

2.3.2 Adaptation

Continuously applying a vibrotactile stimulus beyond the perception threshold can lead to reduced perception which is referred to as **adaptation** [Guyton, 1991].

This is just a temporary effect and after a short duration, it passes if proper delays between stimuli are provided. The adaption rate for fast adapting mechanoreceptors such as Pacinian corpuscle (responsible for vibration) is very rapid [Guyton, 1991] as also suggested by the name.

2.3.3 Phantom Sensation

Simultaneous stimulation of two or more locations in close proximity may only be perceived as a single stimulation somewhere *in between* tactors. This haptic illusion is typically referred to as phantom sensation [Alles, 1970] (sometimes also refered to as funnelling). The exact location of the perceived stimulus depends, among other factors, on the stimulation amplitudes [Alles, 1970, Schneider et al., 2015, Park et al., 2016]. Thus researchers have created models to control the location and the intensity of this perceived phantom for using it in various applications. When using amplitudes of tactors to create the illusion, there are three main rendering models to control the location of sensation:

- Linear model [Alles, 1970, Schneider et al., 2015, Park et al., 2016] which assumes that the perceived position and amplitude is a linear combination of the amplitudes of two tactors,
- Power model [Alles, 1970, Schneider et al., 2015, Park et al., 2016] where the location is depended on the powers (square) of amplitudes and
- Logarithmic model [Alles, 1970, Park et al., 2016] where the relationship is of logarithmic nature

The phantom sensation has been studied in the literature primarily with a focus on users' perceived quality of temporally dynamic continuous stimuli [Schneider et al., 2015, Seo and Choi, 2010, Cha et al., 2008, Eid et al., 2015, Israr and Poupyrev, 2011]. These studies indicate that the log model provides a qualitatively good movement sensation, as the vibration intensity is perceived as more stable towards the middle between two tractors compared to the linear model. Results of [Seo and Choi, 2010] suggest, however, that the linear model may yield higher localisation precision between two tactors compared to the log model. Similarly, the results of [Barghout et al., 2009] show that linear model could be a good choice for localisation of perceived stimuli. The power model introduced byIsrar [Israr and Poupyrev, 2011] also maintains intensity across space and participants had no clear preference between the power model and the log model [Schneider et al., 2015].

The phantom sensation is mainly used for tactile animation where the perception is moved from one place to the another [Schneider et al., 2015, Israr and Poupyrev, 2011] and it has been proposed to be used for enriching experience while consuming multimedia content [Schneider et al., 2015], creating an immersive experience in games [Israr et al., 2012] and storytelling [Yannier et al., 2015]. The wearable displays presented in Chapter 5 utilise the phantom effect to encode arbitrary continuous-valued quantities. Hence, the goal such proposed displays is to use the phantom effect on more than two tactors to encode numerical values.

2.3.4 Apparent Movement - Sensory Saltation

The apparent movement [Kirman, 1974a, Kirman, 1983] is sometimes referred to as sensory saltation [Geldard and Sherrick, 1972] is another tactile illusion and it is created when two points with distance apart are stimulated with a gap in between them. The perception is a moving stimulation from the first to the second point, sometimes referred to as 'cutaneous rabbit' phenomenon [Geldard and Sherrick, 1972]. In this case the gap between two stimuli is called inter-stimuli-interval (ISI) and it is essential on control the saltation effect [Cholewiak and Collins, 2000, Geldard and Sherrick, 1972, Geldard, 1975]. The saltation effect increases with the decrease of ISI. However, for ISI greater than 200 ms, the saltation effect disappears [Geldard and Sherrick, 1972, Geldard, 1975].

2.4 Vibrotactile Patterns for Encoding Information

Given that our somatic sensory system is able to perceive different properties of vibration, varying such properties can be used to encode information. The information can be delivered in the form of **tactons** which are defined to be vibrotactile patterns representing abstract meanings [Brewster and Brown, 2004].

Vibrotactile patterns are crucial when conveying information as they will affect many properties of the system including and the design of the display and encoding of the information. They dictate not only how fast the information can be conveyed but also how many actuators are needed to convey the information. To encode information, vibrotactile patterns need to be discriminative. in most of the cases (e.g. conveying text), they should also be delivered as fast as possible. Typically a combination of variations in amplitude [Summers et al., 2005, Ternes and MacLean, 2008, Xu et al., 2011], frequency [Summers et al., 2005, Ternes and MacLean, 2008, Xu et al., 2011], duration [Gunther, 2001, Geldard, 1957] (**temporal**) and body locations [Geldard, 1957, Xu et al., 2011, Nicolau et al., 2013, Seim et al., 2014b] (**spatial** or **spatiotemporal**) have been used. Sometimes more than one parameter is used for creating patterns. For instance, Geldard [Geldard, 1957] in his Vibratese work used five locations, a variation of three durations and three intensities to provide patterns for encoding the desired symbols. Reed [Reed et al., 2018] used a layout containing 24 (6×4) tactors on the forearm to construct patterns composed of multiple tactors (2-8). The underlying phoneme patterns were generated by varying location, frequency, duration, waveform, movement and number of tactors involved.

Grill [Gill, 2003] has shown that some levels of both frequency and amplitude can be discriminated. However, encoding information goes beyond discriminating between two patterns as it requires users to precisely identify the pattern in order to map the pattern to the encoded information. This aspect seems to be difficult for frequency or amplitude modulated patterns and in general spatial and temporal patterns are better dimcriminatable than frequency and intensity based patterns [Brown et al., 2006, Geldard, 1960]. Temporal patterns have been used in encoding information in the past. The best example is morse code [Chang et al., 2002, Tan et al., 1997], but they result in longer patterns compared to the spatial patterns. Therefore spatial patterns or a variation of them (spatiotemporal) are preferable when encoding information [Luzhnica et al., 2017, Zhao et al., 2018, Chen et al., 2018b, Reed et al., 2018, Jiao et al., 2018, Cauchard et al., 2016, Zhao et al., 2018, de Jesus Oliveira and Maciel, 2014].

However, while easy to discriminate and remember once they are perceived, spatial patterns are prone to masking effect if more than one location is used in patterns and the locations are close [Novich and Eagleman, 2015]. This is the case with Braille inspired vibrotactile patterns [Nicolau et al., 2013, Nicolau et al., 2015]. Therefore, researchers try to prevent such effect by not stimulating all points at once but rather by incorporating the temporal aspect into it. Such patterns are typically referred to as **spatiotemporal patterns**. Given that any combination of location and time/duration is considered spatiotemporal, there exist many variations of spatiotemporal patterns. A common type are the sequential spatiotemporal patterns [Novich and Eagleman, 2015] where vibromotors in a pattern are turned on and off sequentially one after the other and only one vibromotor is active at a time. Recently, Novich [Novich and Eagleman, 2015] showed that such sequential spatiotemporal patterns result in significantly better discrimination than the spatially encoded patterns where all vibromotors in a pattern onset simultaneously. Liao [Liao et al., 2016] utilised such a spatiotemporal encoding to encode symbols on the wrist. Although such encoding works well in terms of being identified by participants [Liao et al., 2016, Novich and Eagleman, 2015], it is many times slower than the spatial encoding as the total duration of a pattern is a multiple of number of vibromotors within the pattern. Zhao et al. [Zhao et al., 2018] and also Chen [Chen et al., 2018b] et al. [Chen et al., 2018b] used sensory saltation to construct spatiotemporal effects where patterns composed of more vibromotors are perceived as movements.

When using spatial or spatiotemporal patterns the number of actuators varies a lot (5-24) [Geldard, 1957, Nicolau et al., 2013, Nicolau et al., 2015, Reed et al., 2018] due to the use of different patterns and encodings.

In this thesis, for encoding discrete information (letters), I constructed prioritised overlapping spatiotemporal [Luzhnica and Veas, 2017] patterns where vibromotors are activated in sequence after a gap, and they stay on until the pattern is finished. As I will demonstrate later (see Chapter 3), this method velds in patterns that result in better recognition accuracy compared to spatial patterns as they avoid masking effect. Moreover, such patterns are faster than sequential spatiotemporal patterns, as vibromotors share most of the activated time. Also, their duration does not vary as much in contrast to sequential spatiotemporal patterns. As for encoding continuous numbers, I use phantom sensation to provide a spatially encoded continuous value (see Chapter 5).

2.5 Conveying Textual Information

Starting with Braille's invention of the Braille coding in 1824, tactile displays have long been widely used by people with visual impairments. Research on tactile displays equipped with actuators has been ongoing since at least 1924 [Gault, 1924], where Gault [Gault, 1924] used a piezoelectric unit to convert entire recorded speech to touch. Similarly, Kirman [Kirman, 1974b] transmitted speech streams to the palm using a 15×15 vibromotor matrix. Six participants learned to differentiate the patterns of 15 different words. Following similar methodologies of encoding speech directly to tactile, there are numerous works [Yuan et al., 2005, Rönnberg et al., 1998, Reed and Delhorne, 2003, Scott and Filippo, 1977, Phillips et al., 1994] concentrating on providing hearing aids for auditory impaired users. A more recent work on sound to vibrotactile is provided by Novich et al. [Novich, 2015]. They used the real-time recorded audio signal, processed and mapped to a vibrotactile vest. Seven users were trained to recognise 50 unique words for 12 days. Participants were able to recognise trained words with an average user accuracy of $35\% - 65\%$. Additionally, participants were able to recognise 50 other untrained words with an accuracy of above chance. Another less common technique of conveying information is imprinting a shape directly on the skin of the user [White et al., 1970, Bliss et al., 1970, Xu et al., 2011].

A more successful approach of transmitting information through haptics was provided by Geldard [Geldard, 1957] in 1967 through symbols. The device was named Vibratese and used five vibromotors placed on the chest to encode 45 symbols (letters, numbers and most frequent short words). For each symbol only one vibromotor was active. The system was capable of transmitting letters at 0.12 s on average. After 65 hours of training, one subject was capable of receiving 38 wpm (words per minute). Since then, researchers have proposed numerous ways of encoding symbols through vibrotactile cues. Typically a combination of variations in amplitude [Summers et al., 2005, Ternes and MacLean, 2008, Xu et al., 2011], frequency [Summers et al., 2005, Ternes and MacLean, 2008, Xu et al., 2011, Reed et al., 2018, Jiao et al., 2018], duration [Gunther, 2001, Geldard, 1957, Cauchard et al., 2016, Reed et al., 2018, Jiao et al., 2018], body locations [Geldard, 1957, Xu et al., 2011, Nicolau et al., 2013, Seim et al., 2014b, Novich, 2015] and haptic illusions [Zhao et al., 2018, Chen et al., 2018b, Reed et al., 2018, Jiao et al., 2018] have been used. Stimuli has been presented in different locations such as users back [Novich and Eagleman, 2015, Cholewiak and Collins, 2000, Novich, 2015], chest [Geldard, 1957], fingers [Bliss et al., 1970, Cholewiak and Collins, 1995, Cholewiak and Craig, 1984, Cholewiak and Collins, 2000, Nicolau et al., 2015], palms [Kirman, 1974b, Cholewiak and Collins, 1995, Cholewiak and

Craig, 1984, Cholewiak and Collins, 2000], with a recent focus (due to wearable devices) on the back of the hand [de Jesus Oliveira and Maciel, 2014, Luzhnica et al., 2016b, Luzhnica and Veas, 2017, Luzhnica and Veas, 2018a, Luzhnica et al., 2018], wrist [Liao et al., 2016] and forearms [Zhao et al., 2018, Chen et al., 2018b, Reed et al., 2018, Jiao et al., 2018, Cauchard et al., 2016].

Recently, much research concentrated on finding patterns to encode symbols such as letters, numbers or phonemes which then are used to construct complex messages such as words or sentences. Encoding needs to provide patterns that are discriminative, easy to learn and deliver them as fast as possible. Usually, there is a tradeoff between those dimensions. Typically, more than one vibromotor is used to encode a symbol.

As already discussed, spatial patterns where all vibromotors in a pattern onset and offset simultaneously have no time overhead when using more vibromotors and could be made very short but are prone to masking [Novich and Eagleman, 2015] and thus researchers use longer durations to compensate for it. Nicolau et al. [Nicolau et al., 2013] used six vibromotors on the fingers of both hands (index, middle, finger) to convey letters spatially encoded in a braille Alphabet. In an initial experiment, eleven blind participants could correctly identify letters encoded by $2000ms$ with an accuracy of 82%. Seven participants continued in a second study where they were tested for word recognition. They achieved an accuracy of 32.86% when symbols were encoded by a stimulus of $250ms$, 64.29% for $500ms$, 88.57% for $1000ms$ and 92.86% for $2000ms$ (the gap between symbols was the same duration as symbol stimuli). Later, Nicolau et al. [Nicolau et al., 2015] used the same braille encoding on the fingertips for encoding characters. Twelve blind participants could recognise letters stimulated by 2000ms duration with an accuracy of 73%.

Sequential spatiotemporal patterns can avoid masking [Novich and Eagleman, 2015]. Liao [Liao et al., 2016] used such an encoding where actuators are activated for 500 ms and a gap of 100ms is used between them. The alphanumeric symbols are encoded with 2-6 (resulting in a duration of 1100ms - 3000ms) sequential activated locations where the majority of them were encoded by 4 locations on the wrist. Twenty-four participants were able to identify symbols with an accuracy of $85.6\% - 88.6\%$ after on hour of training.

More complex spatiotemporal patterns and encodings have been proposed and successfully evaluated. Zhao et al. [Zhao et al., 2018] used a 6 (2×3) tactors display on the dorsal part of the forearm to encode phonemes either by a single tactor or by an apparent motion between two of them. In a user study, the authors found that although it was not feasible to generate 36 necessary distinguishable patterns, 9 patterns (6 with single and 3 with two vibromotors) were easy to discern. The authors trained 11 users for 45 minutes to recognise 9 phonemes transmitted by a base duration of 150 ms resulting in a duration of 180-770.5 ms for a phoneme, and also trained them on 10 words transmitted as a series of phonemes with a 500 ms gap. Participants were able to recognise phonemes with an accuracy of 82%. Participants were also able to identify seen words with an accuracy of $64.3\% - 74\%$ and ten new (unseen) words with $26.7\% - 46\%$ (depending on phonemes within). An additional study trained 9 participants only on words and not phones at all [Zhao et al., 2018] . They found that participants were able to recognise trained words with an accuracy of 88.6%, unseen words with 55.7% and interestingly phonemes with 90%. Furthermore, the authors tested the recognition of new (pseudo) words with higher speeds (base duration 50-150ms, gap 100-500 ms) which resulted in a recognition accuracy between $21.4\% - 45.7\%$ depending on the speed configuration. Similar encoding, position and layout but with 8 (2×4) tactors was used by Chen et al. [Chen et al., 2018b]. Authors trained participants to recognise words encoded by phonemes of duration 126-267 ms with a gap of 200 ms between them in words. After 65 minutes of training (in 3 days), 19 participants achieved an accuracy of 85.7% for words that were trained using guided training and 71.5% for the ones trained using self-guided training.

Reed [Reed et al., 2018] used a layout containing 24 (6×4) tactors on the forearm to encode 39 phonemes. The authors provided an encoding based on articulatory properties of the phonemes where each phoneme was encoded by multiple tactors (2-8). The underlying phoneme patterns were generated by varying location, frequency, duration, waveform, movement and number of tactors involved. The stimuli duration of a phoneme varied between 100ms to 480 ms with a majority of them being 400ms and 480ms. Such a system was used to train 10 participants on phonemes for 50-230 minutes (depending on the progress) after which they achieved an average accuracy of 85.77% on phoneme recognition. The very same device, encoding and stimuli properties were used by Jiao et. al. [Jiao et al., 2018] to conduct longer studies. The authors trained 12 participants over a course of 10 days (10 min per day) to recognise 39 phonemes and 100 words encoded as a series of phonemes with a 300ms gap in between. The participants achieved 92% accuracy on phonemes. While not all of them were tested in words as they had to achieve a certain performance on phonemes to be eligible, the eligible ones achieved an accuracy of 80% on words.

Dunkelberger [Dunkelberger et al., 2018] introduced a multi-sensory approach by building a haptic display composed by three types of haptic actuators: a vibromotors band containing four vibromotors, a radial squeeze band, and a haptic rocker. Such diverse actuators were able to produce concurrent sensations of vibration, radial squeeze and lateral skin stretch. The haptic display was placed on the upper arm, and it was able to produce 48 unique haptic patterns of 350 ms. The authors conducted a user study where 10 participants were trained for 100 minutes (in 4 different days) to recognise phonemes and words. Even though the device could encode 48 different messages (phonemes), the authors used only 23 phonemes in the study which then formed 150 words. The participants were tested in words as a sequence 350 ms encoded phonemes. Within a word, phonemes were not separated by a fixed gap but rather participants controlled when to proceed to the next phoneme which in turn resulted in an average of 3.5 s gap between phonemes. At the end of the study participants were able to recognise words with an accuracy of 86.6%.

Table 2.2 summarises the details of related work on skin reading that use a symbol (e.g. letters or phonemes) based approach of encoding the information. There, some papers are included in more than one row as they performed many testings with different parameters (e.g. different transmission speed, the gap between symbols or different body positions) which resulted in different recognition accuracies. The work developed as part of this thesis is added to the table and marked by *. Moreover, Figure 2.5 illustrates the symbol and words recognition accuracies in relation to the time it takes to convey a symbol or word. Note that, as within the same method, sometimes different symbols are encoded by different durations. Thus for the representation in the figure, the average of such durations is estimated. In addition, the duration of a word is estimated by the given formula:

$$w_d = \#l \times s_d + (\#l - 1) \times g_s \tag{2.1}$$

where w_d represents the average duration of words, $\#l$ represents the average length (number of symbols) of words, s_d represents the average duration of conveying one symbol and g_s represents the gap between symbols within words which is sometimes referred to as inter-symbol duration. Note that as different methods use either letters or phonemes as symbols, the average word length differs. Thus, for methods that use letters as their basic unit, $\#l$ is set to be 4.79 which is deduced from the Google Books Corpus [Michel et al., 2011]. For authors that use phonemes as basic unit, $\#l$ is set to be 3.34 as [Lamel et al., 1989] reveals.

Figure 2.5 also uses the size of the markers to represent the vocabulary size used by authors on their studies. This value is calculated as the number of used symbols divided by the number of symbols that would be needed for that language to be complete. Considering that all studies use English language, for letters the complete number would be 26 whereas for phonemes it would be 39.

To convey textual information for skin reading, the work accomplished throughout this thesis uses prioritised overlapping spatiotemporal (OST) patterns where vibromotors are activated in sequence after a gap, and they stay on until the pattern is finished. Such patterns are used as they deliver a better recognition accuracy than spatial patterns, and they are faster than sequential spatiotemporal encoding, as vibromotors share most of the activated time (see Sections 3.1, 3.3 and [Luzhnica et al., 2016b, Luzhnica and Veas, 2017]). This thesis also proves empirically that prioritising the activation of actuators based on the spatial acuity as increases the perception (see Section 3.3 and [Luzhnica and Veas, 2017]). In addition, initially, two wearable layouts with six vibromotors, one in forearms and the other on the back of the hand are designed and evaluated in this thesis. Furthermore, an initial letter frequency based encoding is designed to encode letters of the English Alphabet with OST patterns. An initial study (see Section 4.1) trains 16 participants for 5 sessions (5h in total). While the first sessions use lower speeds to train participants, each letter is encoded by a duration of $100ms - 110ms$ in session 4 and $70ms - 80ms$ in session 5. On those sessions, participants achieve a letter recognition accuracy of 90% in session 4 and 94% in session 5. Additionally they are able to recognise words with an accuracy of $86\% - 93\%$ depending on the gap between letters ($100ms - 250ms$). To improve the recognition accuracy, another layout of seven vibromotors and an optimised encoding of the letters is proposed in this thesis. The new layout and encoding aim at avoiding masking issues when letterers are transmitted in sequence to form words. A second study (see Section 4.2) evaluates such optimised encoding and layout, and it demonstrates that such optimisations drastically improve the accuracy of letters (97%) and word (97%) recognition.

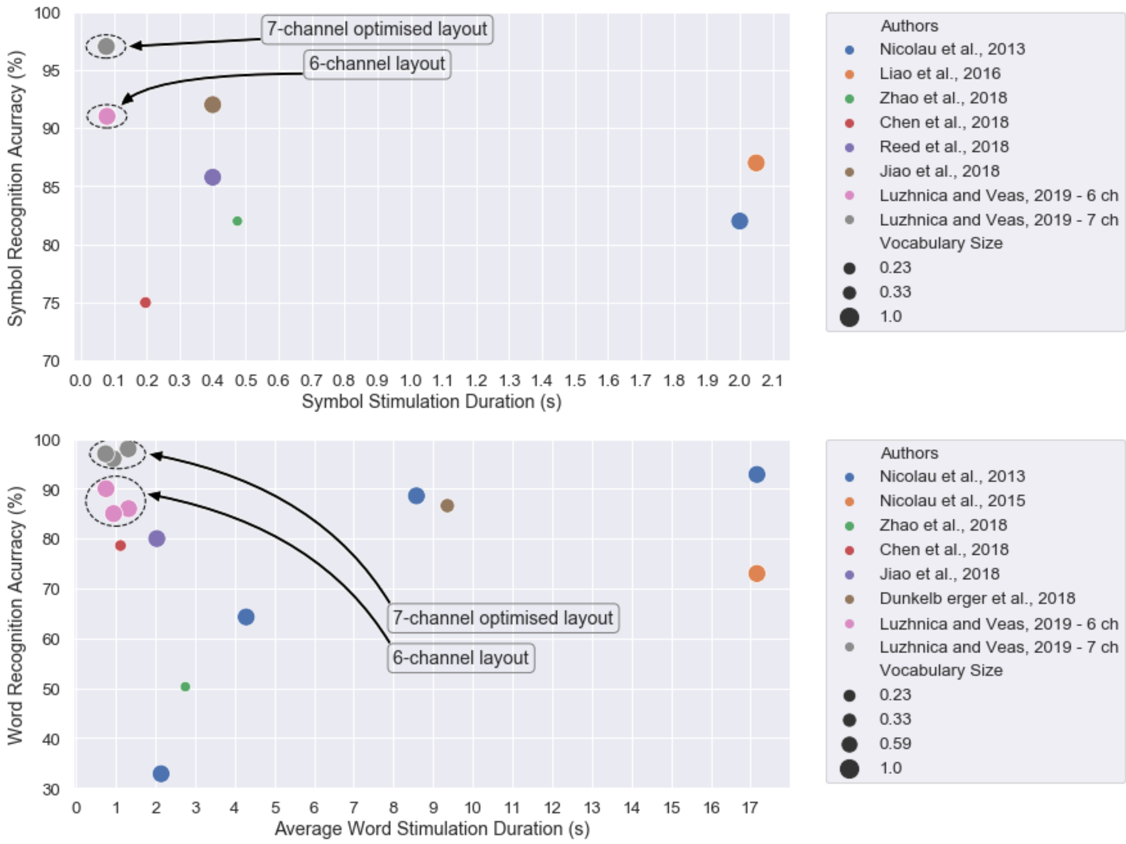

Figure 2.5: Different systems used for skin reading from a related work (from more details see Table 2.2). The figures present the average duration of conveying the information (symbols or words) in relation to symbol (top) and word (bottom) recognition accuracies from different systems. The size of the markers represents the number of symbols used in the author's vocabulary divided by the number of symbols needed to complete the entire vocabulary (26 for letters, 39 for phonemes). Note that some authors only test one of the recognition accuracies (either symbol or word) and thus they do not appear in both figures. In addition to the related work, two systems that are part of this thesis are included for comparison. The first system uses six channels (vibromotors) wearable display either on the back of the hand or forearms. The second system uses the improved layout of seven channels and an optimised encoding which increases recognition accuracy. Both system will be presented in Chapter 4 (Sections 4.1 and 4.2) and have already been published in [Luzhnica and Veas, 2019b].

Table 2.2: An overview of the related work in symbol based skin reading. The details include: Authors - the papers the work was published, ACT - number of used actuators, Train - training time, Users - number of participants, Encoding - essential details of encoding, SB - the type of symbols encoded (L- letters or P - symbols), #SB number of symbols, Duration - stimulation time for a symbol, Gap - gap between symbols in a word, LA - letter recognition accuracy and WA - word recognition accuracy. Other notation: NA - not applicable, NR - not reported, NC - not clearly reported, v/s - vibromotors per symbol. Note that some papers are included in more than one row as they performed many testings with different parameters (e.g. different transmission speed, a gap between symbols or different body positions) which resulted in different recognition accuracies. This table only includes the symbol (e.g. letters or phonemes) based encoding of information. The items marked by * represent the work that was developed from the author of this thesis and it is a part of it. Such work will be described in Sections 4.1 and 4.2 and is has vibromotors been published in [Luzhnica and Veas, 2019b].

Authors	**ACT**	**Location**	**Train**	**Users**	**PTN**	**Encoding**	**SB**	**#SB**	**Duration**	**Gap**	**LA**	**WA**
[Nicolau et al., 2013]	6	knuckles	NA	11 (blind)	S	braille	L	26?	2000 ms		82%	
[Nicolau et al., 2013]	6	knuckles	NA	7 (blind)	S	braille	L	26?	2000 ms	2000ms	NA	92.86%
[Nicolau et al., 2013]	6	knuckles	NA	7 (blind)	S	braille	L	26?	1000 ms	1000ms	NA	88.57%
[Nicolau et al., 2013]	6	knuckles	NA	7 (blind)	S	braille	L	26?	500 ms	500ms	NA	64.29%
[Nicolau et al., 2013]	6	knuckles	NA	7 (blind)	S	braille	L	26?	250ms	250ms	NA	32.86%
[Nicolau et al., 2015]	6	fingertips	NA	12 (blind)	S	braille	L	26?	2000 ms	2000 ms		73%

[Liao et al., 2016]	1	wrist	NC < 60 m	24	SST	2-6 v/s	L	26	1100-3000 ms	NA	85.6%-88.6%	
[Zhao et al., 2018]	6 (2×3)	forearm	45 m	11	SAM	1-2 v/s	P	9	180ms-770.5 ms	500 ms	82%	26.7%-74%
[Chen et al., 2018b]	8 (2×4)	forearm	65 m	19	SAM	1-2 v/s	P	13	126-267 ms	200 ms	NC 65%-85%	71.5%-85.7%
[Reed et al., 2018]	24 (6 × 4)	forearm	50m - 230 m	10	STM	2-8 v/s	P	39	100-480 ms	NA	85.77%	NA
[Jiao et al., 2018]	24 (6 × 4)	forearm	100 m (10 d)	12	STM	2-8 v/s	P	39	100-480 ms	300ms	92%	80%
[Dunkelberger et al., 2018]	6	upper-arm	100 m (4 d)	10	MST	vibration + rocker + squeeze	P	23	350 ms	3500ms	NR	86.6%
Sections 4.1*	6	back of hand & forearms	5 h	16	OST	1-3 v/s	L	26	70-90 ms	250ms	90-92%	86%
Sections 4.1*	6	back of hand & forearms	5 h	16	OST	1-3 v/s	L	26	70-90 ms	150ms	90-92%	85% - 86%
Sections 4.1*	6	back of hand & forearms	5 h	16	OST	1-3 v/s	L	26	70-90 ms	100ms	90-92%	89% - 90%

Sections 4.2*	7	back of hand	5 h	8	OST	1-2 v/s	L	26	70-80 ms	250ms	97%	98%
Sections 4.2*	7	back of hand	5 h	8	OST	1-2 v/s	L	26	70-80 ms	150ms	97%	96%
Sections 4.2*	7	back of hand	5 h	8	OST	1-2 v/s	L	26	70-80 ms	100ms	97%	97%

2.6 Reading Patterns

When providing means for any form of reading, as I intend to do with vibrotactile skin reading, it is important to consider the nature of the reading process itself, draw parallels and then consider the elements that support such process. The most common form of reading is visual, where users use vision to read information represented by symbols and a combination of them. For the individuals with vision impairment, Braille is a standard way of reading. In Braille, the information is represented by symbols which are conveyed as blocks constructed by raised dots. To read it, one must scan such printed information with the fingertip.

A common belief that reading is a sequential task, where eyes *glide smoothly across the page*, is merely an illusion [Rayner, 1998]. At the word level, well-established research postulated that words are recognised as units [Larson, 2004, Fisher, 1975, Reicher, 1969, Cattel, 1886] and they are even recognised before individual letters [Cattel, 1886]. Reading depends on the mechanics of the visual system to stop at fixed spots in the text (fixations) and jump quickly to other spots (saccades, covering about 8 letter spaces) [Rayner, 1998]. Skilled readers fixate on about 2/3 of the words in a text. Beside forward movements to advance in reading, they reread nearby material backwards in the text about 10 to 15% of the time, occasionally driven by breakdowns on comprehension. Conversely, beginning readers fixate every word (often more than once), perform shorter saccades, and up to 50% of their eye movements are regressions, as they rely more on context to identify words [Rayner, 1998]. Obtaining meaning from printed words is not sequential; it depends on processing words as units and uses backward jumps at word level to aid understanding.

In Braille, it is not possible to form a global shape recognition of the entire word, so the text has to be processed character by character [Daneman, 1988, Millar, 2004, Millar, 2003]. The perception and flow of information in Braille are controlled by moving the hand forward and occasionally backwards to revisit information [Millar, 2003, Hughes et al., 2011]. Thereby, Braille readers control reading speed, focus on particular letters or re-scan entire words.

Moreover, Braille readers have full control on the perception flow of information as by moving the hand back and forth, they can control the reading speed, can decide to focus more on particular letters of words or re-scan entire words. People performing visual reading have similar control over the information flow. On the vibrotactile skin reading, users are presented with the information and they have no control over transmission. Thus, we argue that it necessary to provide means of interaction such as re-transmitting different part of text (certain words, certain letters of words). Users might not understand particular parts of the text due to the lack of concentration or training. Additionally, users might want to pause, resume the transmission or change the speed of transmission to account for the progress in their reading skills. Thus this thesis (see Chapter 6) also investigates what interactions are needed for efficient skin-reading, what interaction modalities are preferred by users and how to enable such interactions using wearable sensors.

Chapter 3

Sensitivity Prioritised Overlapping Spatiotemporal Patterns and Wearable Display Layout Design

The work described in this chapter aims at finding vibrotactile patterns to represent discrete information and also design a wearable vibrotactile display layout to convey such patterns. The purpose such wearable vibrotactile display and patterns is to transmit generic discrete *tactons*, which would transmit letters of the English alphabet, and it should be possible to combine them to form words and sentences. Such process of conveying textual information through vibrotactile messages is referred to as *skin reading* throughout this thesis. The wearable device should be able to communicate messages of different types and find application in different scenarios for general purpose use as well as for users with visual or auditory impairment. Thus the work described in this chapter lays the foundations for the *skin reading* which is addressed on the Chapter 4. Both chapters have a common goal of providing methods for stimulating, encoding, and conveying information for skin reading but they address different aspects of it.

Given that the vibrotactile patterns are required to be used in the context of skin reading, they should fulfil two critical requirements:

- *Perceivability.* Patterns should be perceivable and discriminable. This way users should be able to easily identify them when stimulated and not confuse them with other patterns.

- *Throughput.* To be efficient, patterns should be stimulated with shortest possible duration so that they can *transmit* tactons at highest possible (given the restrains in perception) speed. This way when combining consecutive patterns into messages it will result in shorter transmission speed.

Note that optimising patterns for both requirements is a challenging task as such properties have an inverse relationship. Typically the shorter the duration the less perceivable are the patterns. Thus finding suitable patterns is a tradeoff between the two aspects as mentioned above.

On the other hand, there should also be design considerations for wearable vibrotactile display regarding transmission and wearability, where the following requirements should be met:

- *Encoding capacity.* The wearable display should be able to generate enough unique patterns to encode the set of all required symbols, which in this case are the letters of English alphabet. Thus, it is important for the layout to consider the number of vibromotors needed to provide such unique patterns.

- *Convenience.* It should not hinder user's normal activities which rules out some locations such as fingertips or the palm as they are required for object manipulation and everyday interaction.

- *Wearability.* It should be easy to put on and take off and be worn comfortably with different types of clothing.

Although the requirements above are categorised either for patterns or wearable display, some of them depend on both. For instance encoding capacity depends on the number of vibromotors on the layout but also the number of vibromotors within one pattern. The more vibromotors present in one pattern, the more combinations can be created, and thus the less total vibromotors are needed in the layout.

The choices of patterns and the vibrotactile displays designed in this thesis are guided by four user studies summarised in the following:

1. **Study 1** investigates both patterns and wearable layouts. It proposes three layouts for vibrotactile wearable displays based on the above-defined requirements. Additionally, it proposes the **overlapping spatiotemporal patterns (OST)** as a good choice for perception and throughput. Finally, it evaluates

the three layouts and patterns for identification and discriminability of such patterns.

2. **Study 2** proposes **insensitivity varying patterns (IV)**, where several actuators are activated simultaneously but with different intensities. Moreover, it investigates the effects of the sensitivity of the location (spatial acuity) on varying the intensity of vibromotors for the insensitivity varying patterns.

3. **Study 3** proposes to use the sensitivity of the location (spatial acuity) on prioritising the activation of vibromotors for the overlapping spatiotemporal patterns. Thus, it investigates the effect onset prioritisation based on sensitivity or reverse sensitivity.

4. **Study 4** investigates extending the hand based layout proposed in study 1 by three more vibromotors and it evaluates the identification of the overlapping spatiotemporal patterns for those locations.

While each of the studies target various research questions related to vibrotactile patterns and wearable layouts, the main research question of this chapter is:
RQ1: Do the overlapping spatiotemporal patterns result in better identification accuracy than the baseline spatial patterns on the hands and forearms?

The results of all four studies have already been published in two peer reviewed scientific papers [Luzhnica et al., 2016b, Luzhnica and Veas, 2017] (**P1** and **P2**) and their findings enable the scientific contributions **C1 (Vibrotactile Patterns)** and **C2 (Wearable Vibrotactile Display Design)** listed in Section 1.2.

Hardware. The same hardware device is used in the four user studies. The devices consists of an Arduino-Duo board coupled to a power regulator (LM2596S) controls 3.4 mm vibrotactile motors of type ROB-08449 (Voltage range: 2.5V 3.8V ; Amplitude vibration: 0.8G).

3.1 Study 1: Overlapping Spatiotemporal Patterns and Wearable Vibrotactile Display Layout Design

The aim of this user study is first to propose and then evaluate wearable layouts and stimulation patterns with regards to how accurately users identify the locus of vibrotactile stimuli. For this, first three wearable layouts are designed and then evaluated along with proposed vibrotactile patterns.

3.1.1 Wearable Designs

It is clear that the spatial acuity of the skin limits the relative number of vibrators that can be used. To increase the number of symbols, they have to be encoded in combinations of vibromotors. Hence, when designing a layout, the vibromotors have to be sufficiently separated, so that they can be discriminated when stimulated together. Taking into account the *convenience*, *wearability* and *encoding capacity* requirements in contrast with the spatial acuity of the skin in different body parts, three layouts were designed: hand and forearm and two-forearms:

- *Hand layout (H).* Although the skin of the palm has a high resolution in terms of spatial acuity, positioning the device on the back of the hand makes it unobtrusive. For this layout, the vibromotors were placed inside a glove, in the back of the hand and fingers, as shown in Figure 3.1, so as to avoid interfering with grasp and hand interactions typically performed with the palm. On the fingers, vibromotors were placed on the middle phalanx leaving the fingertips free. Using such a layout, fingers can remain uncovered by utilising a partially finger-less glove.

- *Forearm layout (F).* Vibromotors were fixed within a sleeve. The relative size of the vibromotors in relation to the spatial acuity in the forearm limited their number to six. Three vibromotors were placed on the outer and three on the inner side of the forearm, as shown in Figure 3.1. Due to the concern over the insufficient distance between vibromotors, a ”two-forearms” layout was introduced.

- *Two forearms layout (F2).* On each forearm, two vibromotors were placed on the outer side (extremes), and one in the middle on the inner side of the forearm, see Figure 3.1. A drawback of the two-forearms layout is that it requires wiring, e.g., across the back to the controller. Alternatively, the device could be implemented using two controllers connected wirelessly to provide a wireless wearable vibrotactile display.

A sketch of the wearable layouts with the locations of vibromotors is illustrated in Figure 3.1.

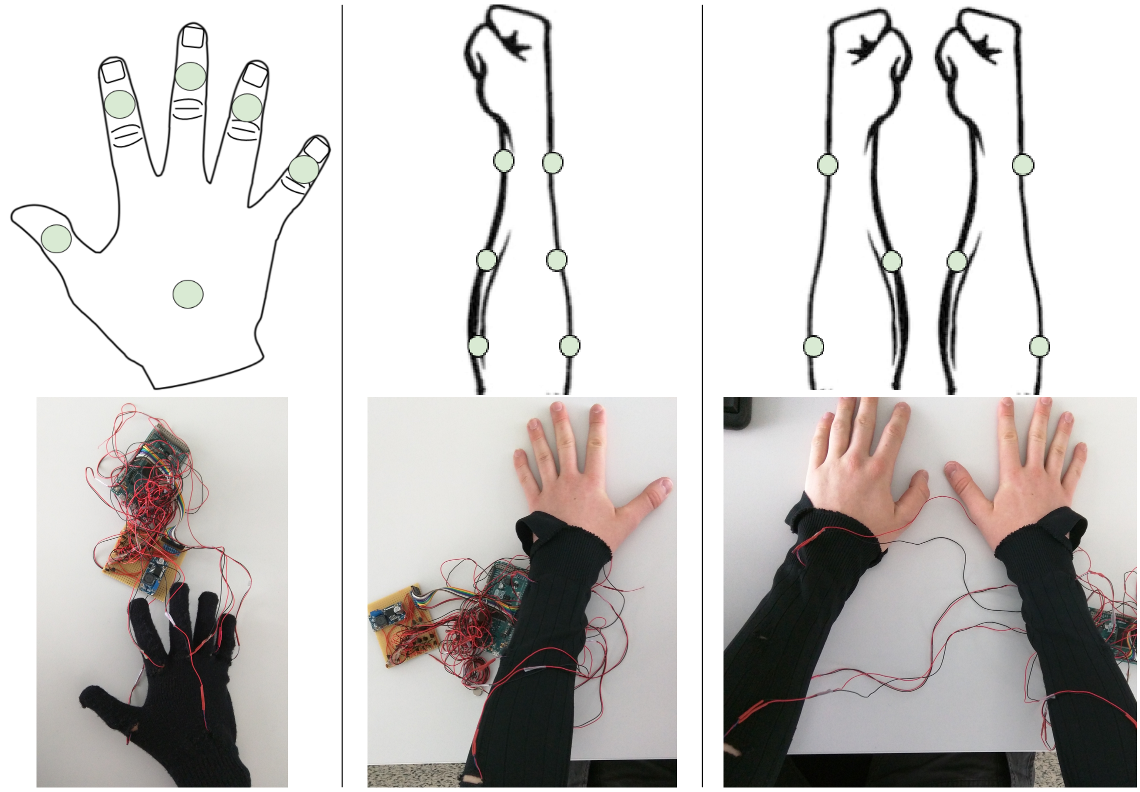

Figure 3.1: Wearable layouts (hand, forearm and two forearms): positions of the vibrotactile motors and pictures form actual wearable prototypes.

3.1.2 Overlapping Spatiotemporal Patterns

Tactons can be encoded using different patterns which are constructed by varying different parameters such as spatial (location of stimulus), temporal (duration of stimulus), varying amplitude or even frequency (of vibration). To fulfil the *throughput* requirement, the stimulation patterns need to ensure a short time of stimulation.

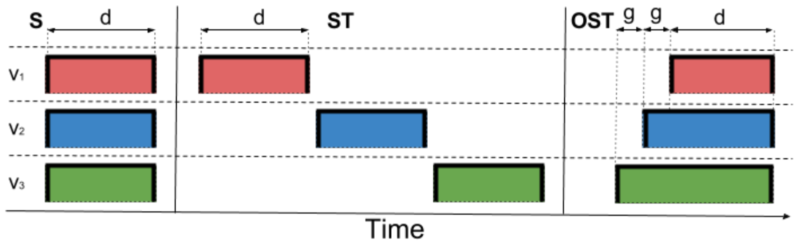

Figure 3.2: Patterns composed of three vibromotors/locations: spatiotemporal (ST), overlapping spatiotemporal (OST), spatial (S). Base duration (d) represents the activation time of a vibromotor (v_1, v_2 and v_3). The gap between the activation of vibromotors is denoted by g.

With *spatial patterns*, all motors in a pattern are activated concurrently for the same duration of time. Hence, the transmission time remains constant regardless of the number of motors involved (see Figure 3.2). However, several studies have shown that when participants are stimulated simultaneously, they fail to recognise some of them [Luzhnica et al., 2016b, Luzhnica and Veas, 2017, Cholewiak and Collins, 1995, Craig, 1982, Novich and Eagleman, 2015]. This problem is known as masking, as one stimulus decreases the dectectablility (masks) of another one [Cholewiak and Collins, 1995, Craig, 1982]. Novich [Novich and Eagleman, 2015] found out that spatiotemporal patterns result in much better discrimination compared to spatial ones. In spatiotemporal patterns each motor is turned on only after the previous one has been turned off, hence yielding a higher transmission duration and lower throughput.

As a trade-off, I used a combination of them, in which the vibromotors are activated in sequence but their vibration time overlaps (see Figure 3.2). That is, a pattern started with a single tactile stimulus (one active motor); after a *gap* time the next motor was activated, continuing so until all motors in the pattern were active. Hence, the total duration of a tacton equalled the duration time of a single active motor (*base duration*) plus the sum of in-between gaps (see Figure 3.2). I used a gap of 10ms; twice the minimum suggested value (5ms) [Gescheider et al., 2010]. This method yielded in patterns with just a bit longer duration (when using more than one stimuli) compared to spatial patterns, but it was expected to result in a better receptivity. Given that the activation of vibromotors was shared, I refer to such patterns as overlapping spatiotemporal patterns (OST) [Luzhnica et al., 2016b]. Moreover, the onset order of such OST patterns was prioritised based on location sensitivity. The least sensitive locations were assigned a higher priority. When a tacton was composed of several stimuli, the order of the stimulus was determined by the assigned priority. In the case of the hand layout the priority order was: pinky, ring, middle, index and thumb. The priority was chosen following output of many research studies [Duncan and Boynton, 2007, Vega-Bermudez and Johnson, 2001, Hoggan et al., 2007] converging to the same conclusion. For the forearm, the middle part was assigned the highest priority followed by the upper part; the lowest priority was assigned to the wrist. Priorities for the forearm were assigned in accordance with the sensitivity levels of the forearm [Cholewiak and Collins, 2003].

3.1.3 Research Objectives of the Study

Given the proposed wearable display layouts and OST patterns, this study targets two research questions:

- **Do the overlapping spatiotemporal patterns result in better identification accuracy than the baseline spatial patterns?**
- **Are the proposed layouts (hand, forearm and two-forearms) suitable position for recognising OST stimulated patterns?**

3.1.4 Procedure

The study compares overlapping spatiotemporal (OST) patterns with spatial (S) ones and the three designed layouts by evaluating how accurately users identify the locus of vibrotactile stimuli. Precisely, the study counted three independent variables: layout (F=forearm, F2=two-forearms, H=hand), stimulation (S=spatial, OST=overlapping spatiotemporal), and active vibration motor count (1,2,3), and random with constraints variable duration with values 100ms, 80ms, 50ms.

Every participant was subject of 162 patterns, where for every possible combination of variables (in random order) 3 random tactons were chosen ($3 \times 3 \times 2 \times 3 \times 3 = 162$). After each stimulation, the participant was asked to localise the stimuli, i.e. to point to the active motors involved.

	Partial			Absolute		
#M	S	OST	p	S	OST	p
2	.88 (.23)	.89 (.23)	1.0	.77(.42)	.79(.41)	.29
3	.80 (.24)	.83 (.22)	<.01	.53(.50)	.59(.49)	<.01

Table 3.1: Localisation partial accuracy and absolute accuracy depending on number of used motors (#M) for both stimuli types, p: pair-wise comparison.

Participants

Twelve (four males and 8 females) persons aged between 24 and 33 years participated in this experiment.

3.1.5 Results

Let us define two depended variables partial accuracy and absolute accuracy. The partial accuracy value is defined to be the number of correctly identified stimuli (active motors) versus the number of all stimuli that compose the given pattern. On the other hand, the absolute accuracy is computed as having the value of 1 only if the entire pattern (all active motors) is identified correctly, 0 otherwise.

First, pattern types across all layouts are analysed. Particularly the patterns involving 2 and 3 vibromotors, as both stimulation methods yield the same stimulation for single vibrator patterns. Factorial ANOVA indicates significant effects both in accuracy ($F(1) = 17.9, p < .01$) and absolute accuracy ($F(1) = 34.67, p < .001$). Pairwise comparisons using Wilcoxon signed-rank, shown in Table 3.2, indicates that for patterns with three motors the OST performs significantly better. Hence, OST will be used in further analysis.

Furthermore, Table 3.2 shows the localisation performance for layout and vibrator count when using the OST. A factorial ANOVA indicates significant effects in both partial accuracy ($F(2) = 34.05, p < .001$) and absolute accuracy ($F(2) = 12.7, p < .001$). Table 3.2 shows the result of paired-wise comparisons (Wilcoxon signed-rank) for "two-forearms" and "Hand". All differences between "hand" and "forearm" as well as between "forearm" and "two-forearms" layouts are significant ($p < .001$). "Two-forearms" layout is the most accurate, followed by the "hand".

	#M	F	F2	H	p_{H-F2}
Partial	1	.92 (.27)	1.0 (.00)	.99 (.07)	.32
	2	.78 (.30)	.97 (.12)	.91 (.20)	<.01
	3	.74 (.26)	.90 (.18)	.86 (.18)	.03
	All	.81 (.29)	.96 (.13)	.92 (.17)	<.01
Absolute	1	.92 (.27)	1.0 (.00)	.99 (.07)	.32
	2	.61 (.49)	.94 (.24)	.83 (.38)	<.01
	3	.40 (.49)	.75 (.44)	.61 (.49)	<.01
	All	.0.65 (.48)	.90 (.31)	.81 (.39)	<.01

Table 3.2: Localisation partial accuracy and absolute accuracy for number of used motors and layout when using OST stimulation. F: forearm, F2: two-forearms, H: Hand, p_{H-F2}: pair-wise comparison (between H and 2A)

3.1.6 Discussion

Comparing results for stimulation types (see Table 3.1) showed that OST performed better. Results for the different configurations indicate that "two-forearm" is the layout with maximum discrimination, followed by "hand". The "forearm" layout had undoubtedly the worst performance. This may be due to the small distance between the stimulus on each side of the forearm.

In terms of wearability, during the study I realised that sleeves could be rotated, resulting in the undesirable effect that a participant may feel the stimulus in a different area of the arm. Although for this study, it did not matter as the two sleeves were worn only once by each participant. But for everyday use, the vibromotors are expected to be at the same locus each day a user wears it, which might introduce issues. Aforementioned technical aspects of wiring contribute to this issue; making it difficult to wear the device as required for everyday use.

3.2 Study 2: Intensity Varying Spatial Patterns

The previous user study (see Study 1 in Section 3.1) evidenced that overlapping spatiotemporal (OST) patterns are easier identified than spatial patterns as they can better avoid masking. However, they introduce a gap which causes them to be longer ($(n-1)10ms$ for a given n -number of vibromotors). Since masking is expected to mask the location with lower sensitivity, one could try to avoid such phenomena by stimulating all involved locations at the same time but with different

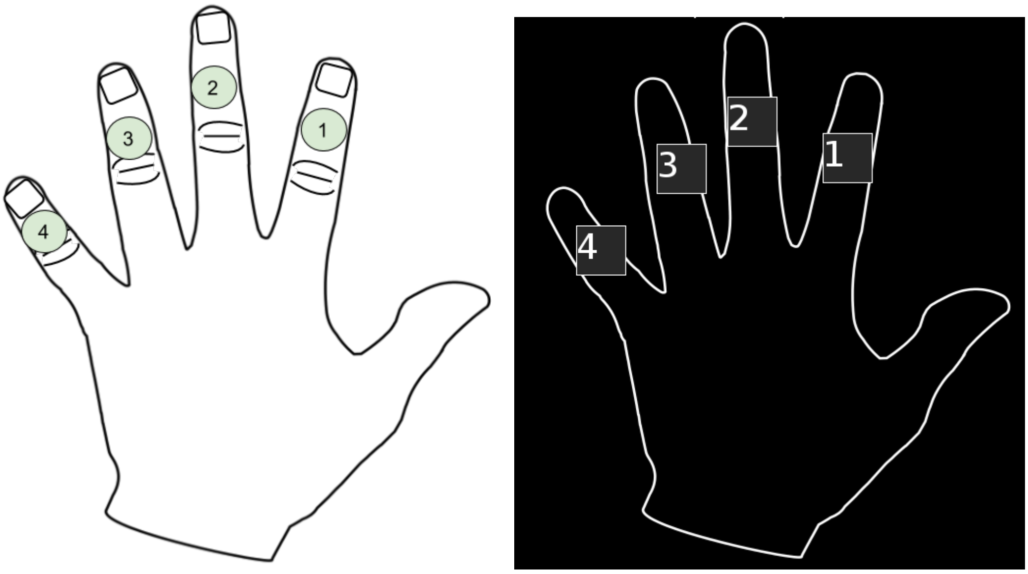

Figure 3.3: The haptic display investigated in the Study 2 and the used user interface for interaction with the study application.

intensities. This way, one could perhaps avoid masking and yet activate the motors for the same duration which would, in turn, result in a shorter total duration than OST patterns as it omits the gap between vibromotors.

The assumption is that stimulating less sensitive locations with higher intensity yields a higher accuracy in recognising a pattern. In other words, different intensities would be used for each vibromotor, with the vibromotor on a more sensitive location being stimulated with a lower intensity than the vibromotor in a less sensitive location. In such patterns, the transmission time remains constant. Thus this study addresses the given research question:

Does the simultaneous activation of vibromotors with different intensities result in higher identification accuracy compared to using the same intensity in all vibromotors?

To answer this question, this study composed patterns that differ on vibration intensity for each vibromotor and then investigated the effects of such variations. The study used only four vibromotors as it aimed to keep participants interested and at the same time gather enough data for statistical analysis. It concentrated on the fingers as locations, because of their known sensitivity order [Duncan and Boynton, 2007, Vega-Bermudez and Johnson, 2001, Hoggan et al., 2007].

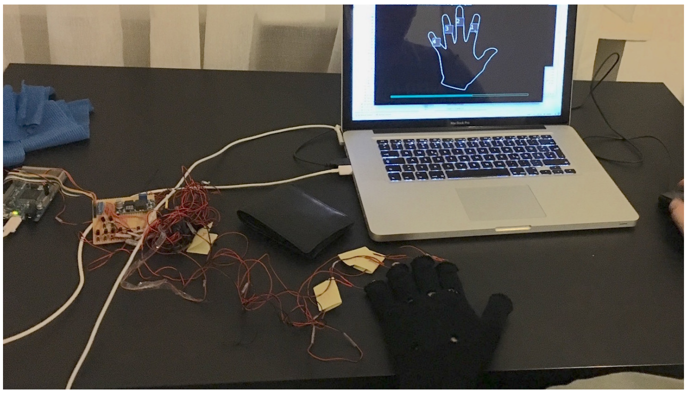

Figure 3.4: A percipient performing the second user study.

PT	PWD_1	PWD_2	g (ms)	d (ms)
S	1	1	0	100
IV1	1	0.75	0	100
IV2	1	0.50	0	100

Table 3.3: Pattern types (PT) used on the Study 2. PWD_1 and PWD_1 represent the duty cycles (vibration intensities) of the first and second vibromotor of the pattern. The base duration is denoted by d and the gap between activation by g.

3.2.1 Procedure

This study used 4 vibromotors as shown in Figure 3.3 and 3.4. For each permutation of the vibromotors, a set of patterns with two vibromotors was generated for spatial (S) and two (IV1, IV2) types of intensity varying patterns (IV). In intensity varying patters one of the vibromotors was activated with a lower intensity than the other one (see Table 3.3 and Figure 3.5). The spatial patterns (S) used the same intensity on both vibromotors, and will served as a baseline to compare with other IV patterns. The two types of IV patterns differ in the intensity of vibration used on the second vibromotor (see Table 3.3). Thus, in total three sets of patterns (S, IV1, IV2) were used. Figure 3.5 illustrates the patterns used in the study. Note that as an Arduino device is used to control the vibromotors, it was technically unable to set the intensity of the vibromotors (Arduino devices do not have analogue output). Nevertheless, the effect of lower intensity was achieved by setting a lower duty cycle of pulse width modulation (PWD). A duty cycle of 1 produced the highest vibration intensity.

Since for each permutation of vibromotors a pattern was generated, for spatial (S)

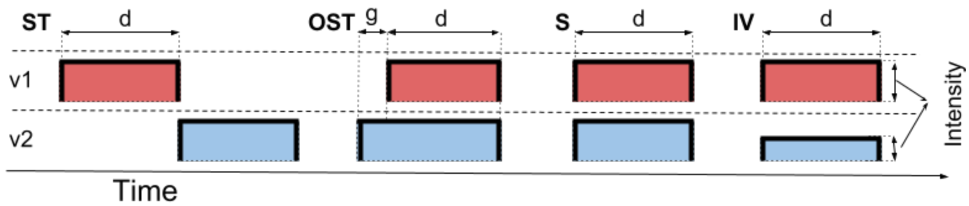

Figure 3.5: Pattern types composed of two vibromotors/locations: spatiotemporal (ST), overlapping spatiotemporal (OST), spatial (S) and intensity varying (IV). Base duration (d) represents the activation time of a vibromotor (v1 and v2). The gap between the activation of vibromotors is denoted by g. The height or rectangle represents the intensity of the vibration.

patterns, each pattern was included twice on the set (as the pattern with vibromotors $1-2$ is the same as $2-1$). In the case of IV type (IV1 and IV2), for every two vibromotors, two patterns with an opposite order of activation were included (e.g. $1-2$, where 1 was activated with higher intensity and 2 with lower one, and then $2-1$ where the order was reversed). Additionally, each set included a pattern with a single vibromotor (with max intensity) for each of the available vibromotors. In total each of the three sets included 16 patterns (12 with two vibromotors and 4 with one vibromotor). The main reason to include single vibromotor patterns is to prevent the cases where users would feel only one vibromotor, but being aware that there are only two-vibromotor patterns, would motivate them to guess one they did not feel. Each participant was tested twice for each of the three sets (S, IV1 and IV2) of patterns. Therefore each participant was tested for 72 ($2 \times 3 \times 12$) probes with two vibromotors and 24 ($2 \times 3 \times 4$) probes with single vibromotor.

The entire experiment was controlled by a Python-based application, which for each pattern in the probes, stimulated participants in a randomised order and then asked them to select the vibromotors in the user interface, by selecting the rectangles representing vibromotors using the mouse (see Figure 3.4). Participants could repeat the stimulation once if they were distracted while the stimulus was applied (e.g. if they were making a comment or a question).

Participants

Eleven participants (six male, five female) took part in the study. All of them were right handed, and we used the left hand for stimulation. The right hand was used to operate the mouse.

3.2.2 Results

Initially, let us define a variable called accuracy which represents the pattern identification accuracy and it is defined to be 1 if participants recognised all locations of stimuli, 0 otherwise. Additionally, let us introduce a new variable called order for pattern types IV1, IV2. We define the pattern to be ordered if the index of the first vibromotor is smaller than the index of the second vibromotor. Otherwise, the order is reversed. If the pattern is ordered, then it is prioritised to stimulate the higher sensitive place with higher intensity than the lower sensitive place. If it is reversed, the least sensitive place is stimulated with higher intensity. As presented in Figure 3.6, both ordered, and inverse variants of IV1 and IV2 result in worse accuracy than the spatial patterns (S). Chi-square comparisons reveal:

- S vs IV1 ordered: $\chi^2(2, N = 396) = 5.89, p = 0.015$,
- S vs IV1 reversed: $\chi^2(2, N = 396) = 3.4, p = 0.065$,
- S vs IV2 ordered: $\chi^2(2, N = 396) = 108.44, p = 0.0$,
- S vs IV2 reversed: $\chi^2(2, N = 396) = 83.76, p = 0.0$

For IV1, when comparing ordered vs reversed, a chi-square comparison reveals the differences are not significant $\chi^2(2, N = 264) = 0.14, p = 0.71$. Similarly, for IV2 the changes between ordered and reversed are not significant $\chi^2(2, N = 264) = 1.71, p = 0.19$. When comparing S with IV1 (both ordered and reversed) the changes are significant $\chi^2(2, N = 528) = 6.91, p < 0.01$. Also, for S and IV2 $\chi^2(2, N = 528) = 146.85, p = 0.0$ the changes are significant. Similarly, the changes between IV1 and IV2 are significant $\chi^2(2, N = 528) = 94.07, p = 0.0$.

3.2.3 Discussion

This study investigated whether varying the intensity of vibration of parallel vibromotors increases the accuracy of identifying the patterns and all involved vibromotors. This way one could provide an encoding which results in better accuracy than the baseline (S) without using any gap between the activation of vibromotors. At least with intensities that we investigated (which were controlled by a duty cycle of PWD), such patterns did not even achieve the same accuracy as the baseline S

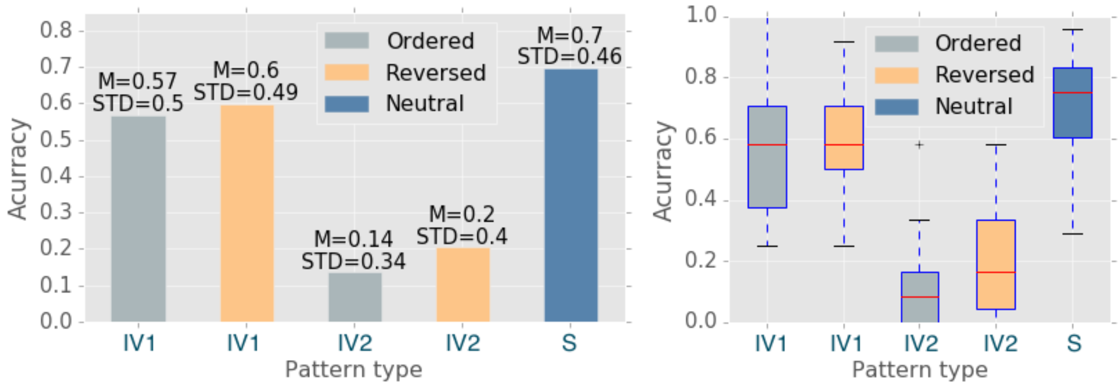

Figure 3.6: Correct identification (accuracy) of patterns for each pattern types (left) used during the Study 2. The box plot (right) presents the results averaged per user.

(spatial patterns), let alone exceed it. Nevertheless, this should not immediately discourage other researchers to investigate the same technique with actuators that offer a more accurate intensity control. It is entirely possible that by tuning the intensities (examining different levels of intensities), this might bring better results. Within the frame and settings of this study, such an encoding technique did not prove to be useful. Therefore such patterns will not be further utilised or promoted within the work of this thesis.

3.3 Study 3: Sensitivity Prioritised Overlapping Spatiotemporal Patterns

Th Study 1 (see Section 3.1) introduced the overlapping spatiotemporal (OST) patterns, where onset occurs in sequence after a *time gap* for each vibromotor after the first one and demonstrated that it can better avoid masking compared to spatial patterns. Moreover, it prioritised the activation of motors based on lower to higher sensitivity of the location. The main idea is that prioritising the least sensitive location and activating it for a gap while the more sensitive one is not active, users would perceive it. Then later (after the gap), when both vibromotors are activated, even if the least sensitive place is masked by the more sensitive, participants already are aware of its simulation (during the gap). However, the effects of such onset prioritisation were not investigated in the previous user study. Therefore, this study

PT	PWD_1	PWD_2	g (ms)	d (ms)
S	1	1	0	100
OST1	1	1	10	100
OST2	1	1	20	100

Table 3.4: Pattern types (PT) used in the Study 3 PWD_1 and PWD_1 represent the duty cycles (vibration intensities) of the first and second vibromotor of the pattern. The base duration is denoted by d and the gap between activation by g.

investigates in details how the prioritisation affects the perception and in which order should the prioritisation be applied for highest perception.

This study assumes that a sensitivity prioritised onset of stimulation using OST leads to higher recognition accuracy. When using OST, the vibromotors in a pattern are activated in sequence after a *gap*. The sequence of activation is given by the sensitivity of skin in the vibromotor location. All vibromotors remain activated for the duration of the pattern. Thus this study addresses the given research question: ***RQ2. Does the prioritisation of activation of vibromotors have an effect on the accuracy of identification of each vibromotor when using an overlapping spatiotemporal (OST) encoding? How should we prioritise, least sensitive to most sensitive locations or vice-versa?***

To answer this question, this study composed overlapping spatiotemporal patterns consisting of one or two vibromotors. Patterns differed on the gap between the activation of vibromotors and their order. Sensitivity prioritisation guided the onset of vibromotors in a pattern. Its effects were analysed combined with gap duration. Again, the study used only four vibromotors to keep the study concise and at the same time gather enough data for statistical analysis. The fingers were used as stimuli locations, because of their known sensitivity order [Duncan and Boynton, 2007, Vega-Bermudez and Johnson, 2001, Hoggan et al., 2007].

3.3.1 Procedure

Initially a set of patterns of various types was created which included a set of spatial patterns (S) and two (OST1, OST2) sets of overlapping spatiotemporal (OST) patterns where a gap between activation of vibromotors was used (see Table 3.4 and Figure 3.5). The rest of the procedure was identical to the Study 2 (Section 3.2) including the wearable layout (location fo vibromotors) and the user interface for the study. Each participant was tested twice for each three sets (S, OST1, OST2) of patterns. Therefore each participant was tested for 72 ($2 \times 3 \times 12$) trials with two vibromotors and 24 ($2 \times 3 \times 4$) trials with a single vibromotor.

Participants

Twenty participants (eleven male, nine female) took part in the study.

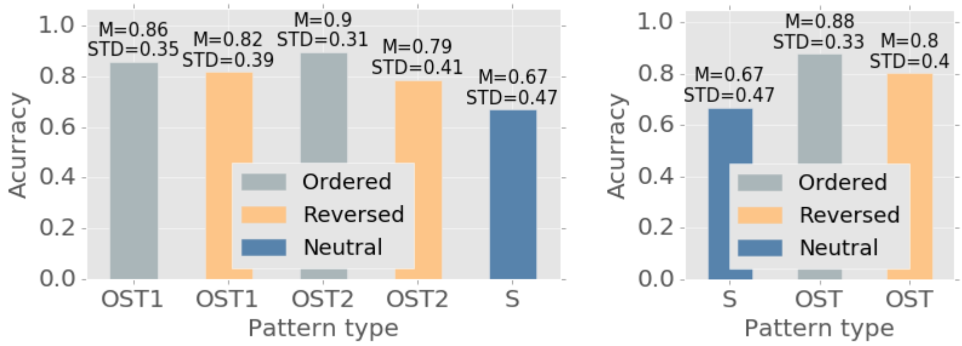

Figure 3.7: Correct identification of patterns for each pattern types (left) used during the Study 3.

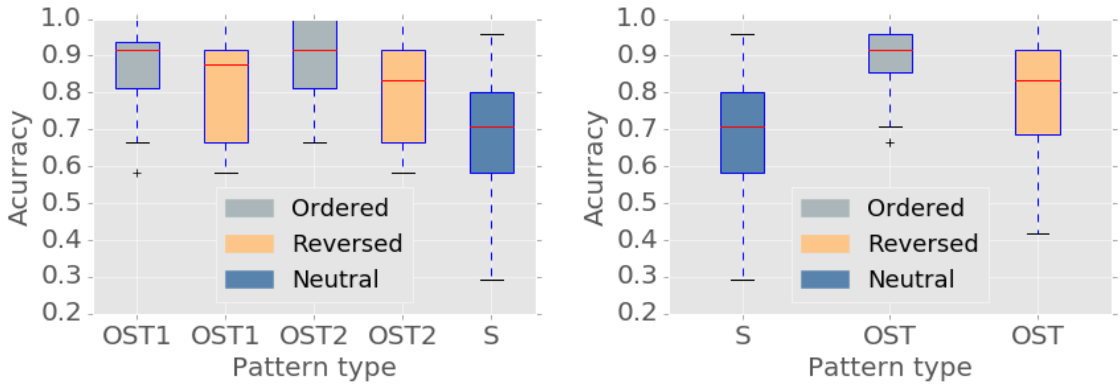

Figure 3.8: Correct identification of patterns for each pattern types (left) used during the Study 3. The results are averaged per user.

3.3.2 Results

Let us introduce order as a variable for pattern types OST1, OST2. If the pattern is ordered, then the location with higher sensitivity is stimulated first, and if it is

	t1/t2	1	2	3	4
S	**1**		.79 (.41)	.29 (.46)	.70 (.46)
	2			.32 (.47)	.67 (.47)
	3				.43 (.50)
	4				
	t1/t2	**1**	**2**	**3**	**4**
OST1	**1**		.94 (.24)	.62 (.49)	.96 (.20)
	2	.96 (.20)		.75 (.44)	.83 (.38)
	3	.50 (.51)	.50 (.51)		.69 (.47)
	4	.85 (.36)	.85 (.36)	.65 (.48)	
	t1/t2	**1**	**2**	**3**	**4**
OST2	**1**		.85 (.36)	.77 (.42)	.94 (.24)
	2	.92 (.28)		.88 (.33)	.98 (.14)
	3	.62 (.49)	.65 (.48)		.60 (.49)
	4	.81 (.39)	.90 (.31)	.71 (.46)	

Table 3.5: Results of the Study 3 for each combination of two vibromotors. The row defines the first activated vibromotor whereas the column defines the second. In the case of S both vibromotors are activated in parallel, therefore, the results are displayed together. Color coding: ○ - spatial, ● - ordered (OST), ● - reversed (OST).

reversed, then the lowest sensitivity location is prioritised. The average identification accuracies of ordered, reversed and neutral patterns are presented in Figure 3.7 and Table 3.5. Additionally a boxplot of averages for each user is presented in Figure 3.8. Figures 3.7 and 3.8 reveal that ordered OST performed better than reversed (for both OST1 and OST2) and all combinations of OST performed better than S. Nevertheless, for determining significance chi-square tests are used.

Comparing S with ordered and reversed of OST (combined OST1 and OST2) reveals that changes between S and both ordered and reversed OST were significant $\chi^2(2, N = 960) = 57.56, p = 0.0$; respectively $\chi^2(2, N = 960) = 24.54, p = 0.0$. Additionally also the changes between ordered and reversed OST were significant $\chi^2(2, N = 960) = 7.23, p = 0.0072$. On the other hand, all combinations of OST and ordering performed significantly better than baseline S [1]:

- OST1 ordered vs S : $\chi^2(2, N = 720) = 27.28, p = 0.0$,

[1]as significance a threshold of $\alpha = 0.0125$ is used according to following Bonferroni correction

- OST1 reversed vs S :$\chi^2(2, N = 720) = 19.92, p = 0.0$,
- OST2 ordered vs S : $\chi^2(2, N = 720) = 42.64, p = 0.0$ and
- OST2 reversed vs S : $\chi^2(2, N = 720) = 11.24, p = 0.0008$

Additionally, the baseline (S) seems to have performed significantly worse than both OST1 $\chi^2(2, N = 960) = 38.33, p = 0.0$; and OST2 $\chi^2(2, N = 960) = 39.39, p = 0.0$ regardless of order. When comparing the ordering within OST1 and OST2, for OST1 the differences were not significant between ordered and reversed $\chi^2(2, N = 480) = 0.4, p = 0.527$; whereas for OST2 the changes were significant $\chi^2(2, N = 480) = 9.31, p = 0.002$.

Interestingly, the differences between OST1 and OST2 do not seem to have been significant $\chi^2(2, N = 960) = 0.1, p = 0.75$. Also the differences between ordered OST1 and ordered OST2 were not significant $\chi^2(1, N = 480) = 0.0, p = 1.0$. Similarly, the differences between reversed OST1 and reversed OST2 were not significant $\chi^2(1, N = 480) = 0.0, p = 1.0$.

3.3.3 Discussion

This user study reveals that participants identified the stimuli significantly better using overlapping spatiotemporal patterns (OST) than Spatial ones (S). Participants also performed significantly better when the order of vibromotors was from smallest to the highest index (for the OST). Since that order was the exact order of sensitivity of the locations [Duncan and Boynton, 2007, Vega-Bermudez and Johnson, 2001, Hoggan et al., 2007], this suggests that prioritising the onset of vibromotors based on sensitivity in OST encoding significantly increases the accuracy of identification of patterns. Surprisingly, the increase in accuracy was achieved by prioritising the most sensitive locations first, which is the opposite of what was assumed to be the case in previous research [Luzhnica et al., 2016b]. Intuitively, one would expect that by prioritising the least sensitive location while the more sensitive one is not active (gap), users would perceive it. Later when both vibromotors are activated, even if the least sensitive place is masked by the more sensitive, participants already are aware of its simulation (during the gap). Although such patterns were significantly better than the baseline (spatial patterns), the opposite, prioritising the most sensitive location, worked significantly better. Perhaps exactly the kick of the second activation is much more efficient mechanism against masking. It is also interesting that the gap (10 ms vs. 20 ms) used between activation of vibromotors in OST did not have a significant effect on identification accuracy. In the used settings (base duration of 100 ms), 20 ms gap increases the total duration of patterns for 9% (110 ms vs. 120 ms) over the 10 ms gap. Despite the overhead which will affect the throughput when encoding symbols, it still did not result in a significant gain in accuracy.

3.4 Study 4: Extending the Hand Based Layout

When using OST to provide patterns, Study 1 (see Section 3.1) revealed that encoding patterns with more than two vibromotors resulted in significantly lower accuracy than the cases with one and two vibromotors [Luzhnica et al., 2016b]. Similarly, when encoding information (e.g. letters of English Alphabet) with such patterns the comprehension of encoded information suffers for information represented by patterns encoded by three vibromotors (see Section 4.1). Thus when using OST patterns, it would be recommended to encode information by maximum 2 vibromotors. However, this limits the number of patterns and thus the vocabulary of information that could be encoded by them. To increase the size of the vocabulary that can be encoded, it becomes necessary to add vibromotors in different locations. The number of symbols that can be encoded with one or two vibromotors is:

$$n = \binom{m}{2} + m = \frac{m(m-1)}{2} + m = \frac{m(m+1)}{2} \tag{3.1}$$

where m represents the number of vibromotors in the haptic display. A display with $m = 7, n = 28$ can encode the entire English alphabet plus two other characters (e.g. space and period). With $m = 8, n = 36$ and with $m = 9, n = 45$ it would be sufficient to also encode most of the punctuations and symbols.

This user study aims at extending the "hand" based layout proposed in Section 3.1 by three vibromotors for the purpose of increasing the number of unique OST patterns that can be generated. The vibromotors are added in such a way that they still could be placed inside a fingerless glove. No vibromotor is placed on the palm, to avoid interference with everyday interactions. As shown in Figure 3.9, two of vibromotors (6 and 7) are placed on the back of the hand, at acceptable discrim-

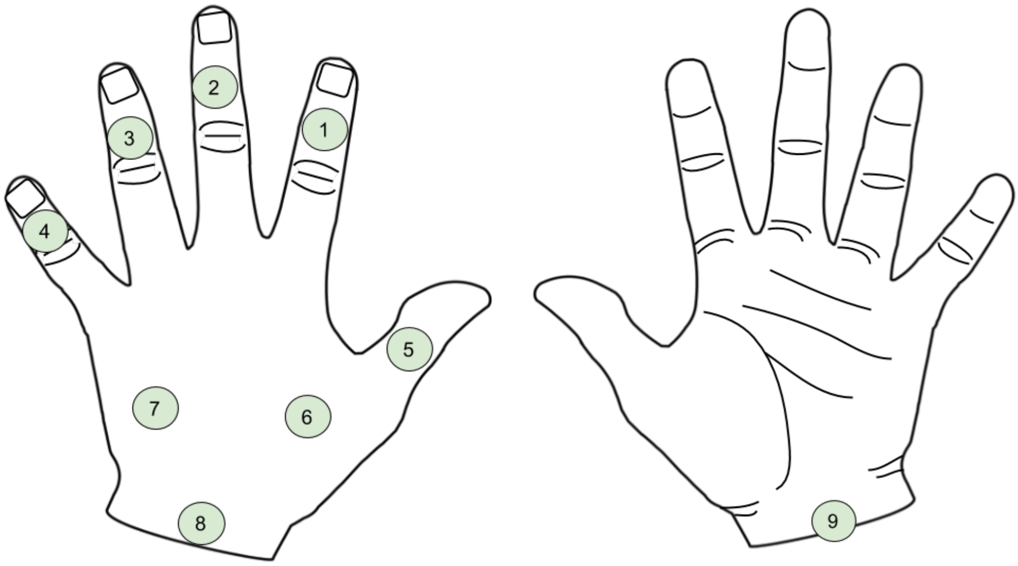

Figure 3.9: The proposed wearable haptic display with 9 vibromotors which extends the "hand" layout presented in Figure 3.1. Study 4 uses the back of the hand and wrist locations $(6-9)$.

ination distances as given by the cutaneous sensitivity of the hand. Additionally, there are two vibromotors near the wrist, one on the back and one on the front of the hand. The primary concern about this design is whether combinations of vibromotors 6, 7 and 8 can be used within the same pattern, considering the small distance between them. Additionally, as the vibromotor 9 is on the opposite side of vibromotor 8, it raises the question whether their combination would be recognisable as well. Combining fingers with a single vibromotor on the back of the hand was tested in Study 1 and Study 5 (see Section 4.1) using a "hand" based layout with a similar distance between the hand and finger vibromotors which is assuring that the distance between hand vibromotors and fingers is enough to avoid masking. Thus, to keep the study procedure within a manageable time, this study does not study patterns combining finger vibromotors $(1-5)$ but hand vibromotors $(6-9)$ to identify whether such positions are suitable for OST patterns. Thus this study addresses the following research question: ***RQ3. Can stimulation with high throughput and accuracy be achieved in less sensitive parts of the hand such as the back of the hand and wrist area?***

3.4.1 Procedure

Three vibromotors were placed on the back of the hand (one of them near the wrist) and one on the palm side near the wrist. The exact positions are given on Figure 3.9 (vibromotors 6-9) of the new design. Apart from the position of vibromotors, the rest of the study was organised in the same manner as the previous user study (Section 3.3). Three sets of probes (S, OST1 and OST2) were used in a randomised order to test the participants for identification. In total each user was tested for for 108 ($3 \times 3 \times 12$) probes with two vibromotors and 36 ($3 \times 3 \times 4$) probes with one vibromotors.

Participants

In this study participated 15 people (seven male, eight female).

vibromotor	6	7	8	9
Accuracy	.99 (.08)	.99(.12)	.90(.30)	.99(.12)

Table 3.6: Average correct identification of patterns composed of only one vibromotor.

3.4.2 Results

The accuracies for each combination of vibromotors are presented in Table 3.6. The table shows that all the patterns that involve vibromotor 8 performed worse than the others. To elaborate on this result, Figure 3.10 illustrates the accuracies of patterns grouped by vibromotors that they contain. On the top, it visualises the accuracies of all patterns, whereas, on the bottom, it visualises only ones that do not involve vibromotor 8. Note that groups are not exclusive as each pattern belongs to two groups (e.g., patterns 6-7 belongs to both groups of vibromotor 6 and vibromotor 7) and hence inaccuracies of patterns involving vibromotor 8 affect other groups as well. Here, when comparing pairwise, each group is statistically significant compared to group 8 (6 vs 8: $\chi^2(2, N = 1728) = 62.55, p = 0.0$; 7 vs 8: $\chi^2(2, N = 1728) = 73.15, p = 0.0$; and 9 vs 8: $\chi^2(2, N = 1728) = 64.25, p = 0.0$).

In addition, the accuracies of patterns composed of only one vibromotor are presented in Table 3.6. When looking at the comparison between accuracies for patterns with one vibromotor only, the differences between all other vibromotors

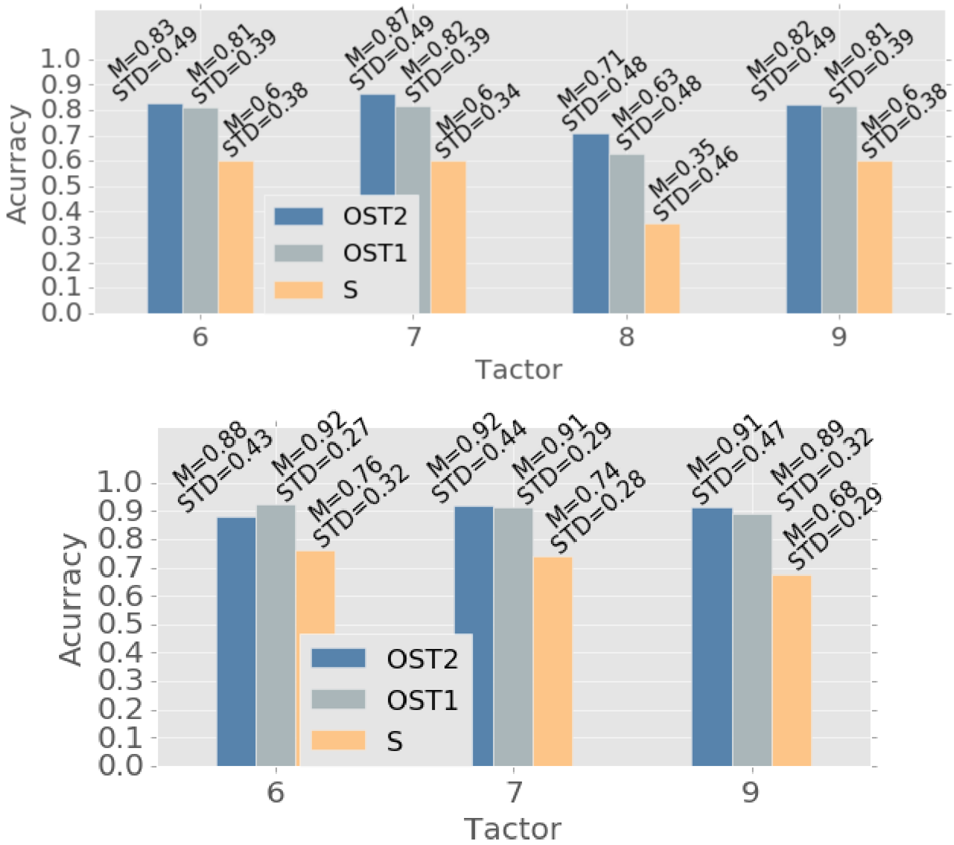

Figure 3.10: Correct identification of patterns for each pattern that involves the vibromotor. Please note that each pattern is included in two categories as it contains two vibromotors. Therefore, when removing vibromotor 8, the accuracies increase in other groups.

and vibromotor 8 are significant (6 c 8 :$\chi^2(2, N = 288) = 10.13, p = 0.0015$; 7 vs 8: $\chi^2(2, N = 288) = 8.01, p = 0.0047$; and 9 vs 8: $\chi^2(2, N = 288) = 8.01, p = 0.0047$), whereas the differences between vibromotors 6, 7 and 9 are not. Both comparisons (Figure 3.10 and Table 3.6) point out that the location for vibromotor 8 is not a good choice for a haptic display.

Following the second study where onset prioritisation resulted in higher accuracy, let us define the order of activation for positions on hand as well. While for the first and second studies the sensitivities are well known and studied [Duncan and Boynton, 2007, Vega-Bermudez and Johnson, 2001, Hoggan et al., 2007], for the positions chosen in this study, to the best of my knowledge, there is no evidence comparing their sensitivities. For this, Figure 3.11 presents the average accuracies

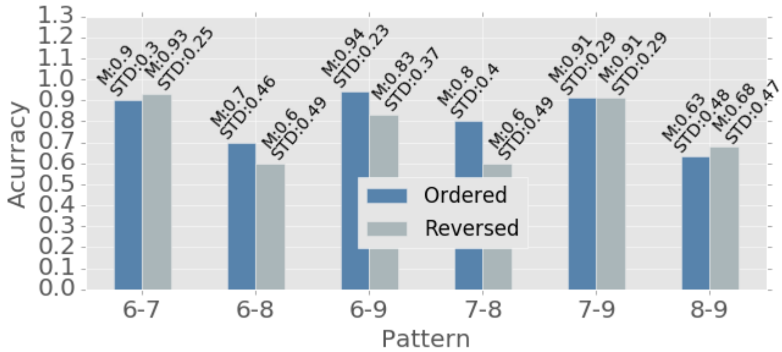

Figure 3.11: Correct identification of patterns for each pattern that involves two vibromotors (only for OST patterns) in the Study 4.

for all combinations of vibromotors including the order for OST pattern types (both OST1 and OST2). Comparing each of them (e.g 6-7 vs 7-6) reveals that the differences between 6-9 vs 9-6 ($\chi^2(2, N = 192) = 5.35, p = 0.0208$) and 7-8 vs 8-7 ($\chi^2(2, N = 192) = 11.84, p = 0.0006$) are significant.

Based on this evidence, the priorities of stimulation for vibromotors 6 and 7 would be higher than 9. Between 6 and 7, one would prioritise 7 just because of the average accuracy, but either way, it would not make a major difference. Whereas for vibromotor 8, I would recommend to remove it from the design as long as it is not required to encode a vocabulary with more than 36 symbols.

3.4.3 Discussion

This user study introduces four vibromotors on the hand and tests all combinations of patterns composed of one and two vibromotors. The results revealed that vibromotor 8 is comparably poor for haptic stimulation as patterns that contain it were identified significantly worse than patterns that do not. Other proposed locations seem to provide comparably good accuracy with both OST encoding types.

Please note that even though accuracy is not 100%, they are still a good fit for a haptic display as there is some learning effect to it. Numerous studies in neuroscience shed evidence that with exposure to stimuli there is a tactile learning effect [Reuter et al., 2014, Godde et al., 2000, Dinse et al., 2005, Pleger et al., 2001, Pleger et al., 2003, Hodzic et al., 2004] and thus with the usage of the vibrotactile displays perception might improve. For instance, in [Luzhnica et al., 2016b], participants after some training time performed much better in identifying the symbol associated with the pattern (98% accurate within one hour of training) than they performed in a pre-study where they were asked identify the location of the stimulus (83% on the hand for patterns with two vibromotors). This suggests that as users are exposed more to the stimulus, they can identify more accurately the stimulus.

Considering the results of the third study, the final design of the glove-based haptic display is composed on eight vibromotors. Seven ($1-7$ in Figure 3.9) of them placed on the back of the hand whereas one is placed on the wrist of the palm side of the hand (9 in Figure 3.9). With eight vibromotors, the new proposed wearable display would able to encode 36 different symbols using prioritised OST, which is enough for the entire English alphabet and most important punctuations.

3.5 Summary

The main objectives of the work reported in this chapter are to construct vibrotactile patterns that are optimised for perception and transmission speed as well as design wearable vibrotactile displays that are suitable for stimulating such constructed patterns. This work designed wearable vibrotactile layouts and constructed vibrotactile patterns with the goal of using them later for skin reading. Thus, it was also essential to consider the number of unique patterns that could be generated for the proposed pattern types and the designed wearable layout. The design of patterns and wearable layouts was achieved through the process of four user studies each examining some aspects of them.

The first user study proposed three wearable display considering a thorough discussion of perceptual factors and design considerations. Moreover, it proposed overlapping spatiotemporal patterns with the goal of increasing the identification accuracy compared to spatial patterns. The study revealed that two layouts: a hand based and two-forearms based could be a good fit for identifying the proposed OST patterns. Additionally, it showed that the proposed OST patterns are a good choice as they result in better identification accuracy compared to the spatial patterns.

The second user study showed that using different vibration intensities between vibromotors does not contribute to a higher accuracy (even when they are prioritised by sensitivity) than the baseline spatial encoding where the intensities are kept constant for both vibromotors.

The third user study examined whether the order of activation of vibromotors in overlapping spatiotemporal patterns has an effect in correctly identifying the stimulus. The results suggested that prioritising the activation of vibromotors based on highest sensitive place towards lowest significantly increases the accuracy. Such results are surprising and exactly the opposite of what it was assumed in the first user study. Prioritising the vibromotors suggest, will contribute to an increase in perception accuracy when using a wearable vibrotactile display.

Moreover, the fourth user study extended the investigation on sensitivity to new hand locations. It experimented with four additional locations and kept three for a proposed final design with eight vibromotors. The results showed that the proposed layout with eight vibromotors is suitable for stimulating OST patterns and thus an excellent choice for skin reading. Such a wearable vibrotactile display can be used to encode up to 36 characters in one or two vibromotors with prioritised OST, which is more than necessary for the entire English Alphabet.

Overall the work reported in this chapter shows that the proposed overlapping spatiotemporal patterns prioritised by sensitivity (highest to lowest) are a good tradeoff for both perception (better than spatial) and speed (faster than sequential spatiotemporal) and thus present an excellent choice for skin reading. They result in better identification accuracy than the baseline spatial patterns and could be used in the back of the hands and forearms (**RQ1**). The results also propose a layout with enough vibromotors to encode 36 symbols. Thus the foundation for encoding text messages is laid out in terms of wearable device and stimulation methods.

Chapter 4

Conveying Textual Information through Vibrotactile Wearable Displays

Wearable devices are already a part of our everyday life. They provide assistance to daily activities and enrich them with additional information collected by the sensors within them. The primary feedback modalities of wearables devices are visual and auditory. Although, most of them include vibrotactile capabilities, the primary utilisation of vibrotactile feedback is to provide additional support to visual interaction. On the contrary, the vibrotactile feedback on wearable devices has a lot of potential to be used on its own. A prominent application and the subject of this chapter is the so called *vibrotactile skin reading* (*VSR*) which is sometimes referred to as simply *skin reading*.

Vibrotactile skin reading uses vibrotactile patterns to encode symbols [Geldard, 1957, Luzhnica et al., 2016b, Liao et al., 2016, Zhao et al., 2018] (letters or phonemes) which then can be combined to convey complex messages such as words and phrases [Geldard, 1957, Luzhnica et al., 2016b, Zhao et al., 2018]. While traditionally, perceiving information through skin has been applied for visually impaired users (e.g Braille reading) and skin reading presents some good opportunities for them [Luzhnica and Veas, 2018b, Chen et al., 2018a], users with normal vision in multi-tasking scenarios can also benefit from a means of perceiving messages that do not recruit the visual or auditory senses. Depending on the method, the duration of conveying symbols varies from hundreds of milliseconds to few seconds. Performance varies as well, with most studies reporting recognition accuracies between 80% and 90% on symbols and words with a few hours of training. Considering the short training time, such performances are quite promising but yet not at a state to be considered for real-world applications. Performance could potentially improve with more training and practice. However, limitations to design of wearable devices, locations of simulations and encoding of information could also cause systematic performance errors.

The work on this section explores the use of high speed overlapping spatiotemporal patterns (OST) and also wearable display layouts which were developed and introduced in Chapter 3 for skin reading. In addition to encoding information and evaluating the capabilities of OST patterns and wearable layouts, it investigates other aspects of skin reading such as optimising encoding, training methods, human capability of perceiving information through skin reading in multi tasking scenarios, transferability of training and the decay effect of training through time. The work is guided by four user studies, each investigating different aspects of skin reading:

- **Study 5**: investigates skin reading using overlapping spatiotemporal patterns on two layouts (hand and forearm) designed and evaluated in Section 3.1. It first proposes a training program and a letter frequency-based encoding to encode letters of English Alphabet using OST patterns, which then are combined to form words. This study enables to analyse the performance and more importantly explore whether there are any systematic errors that could be avoided. Given that two layouts are used, it can be considered that these any found errors are not bound to a particular layout or body location and contribute a generalisation of the findings. The data produced during the study allows to successfully identifies issues on the recognition of individual letters as well as sequences of letters (words). Given such findings, a two-step optimisation of the alphabet encoding and layout is proposed to prevent issues in communicating sequences of letters and avoid issues with single letters. The optimisation process results in a new, optimised layout and alphabet encoding, which is then later evaluated in the next user study. The locations of the new optimised layout are based on the previous investigations (based on Study 4 four described in Section 3.4) on suitable positions of actuators.

- **Study 6**: evaluates the new optimised alphabet encoding and the optimised layout and it provides evidence on the impact of the proposed optimisation methodology on the aforementioned issues. It shows that such an optimisation has a major impact on the recognition of individual letters and words. In addition, results shed new evidence on the effectiveness of the proposed method for skin reading. First, it empirically demonstrates that participants can transfer the knowledge of encoding and letter recognition to an untrained body part (more specifically, to the untrained hand). Additionally, it provides evidence that such knowledge and letter recognition performance remains over time even without re-training or usage after 3, 10 and 19 days.

- **Study 7**: builds on top of Study 6 to investigate skin reading in multitasking scenarios. For the study, already trained participants were recruited to investigate whether vibrotactile encoded messages can be perceived as secondary while performing another primary task. Moreover, it evidences the effect of primary task on the secondary one (decoding vibrotactile messages) and vice versa.

- **Study 8**: investigates the use of passive haptic learning to train for skin-reading. Additionally, it explores the the effect of teaching structures by comparing word based training with letter based training. Moreover, it explores the effect of different transmission speed on the recognition of encoded information.

While each user study target different research questions related to skin reading, the main research question of this chapter is:
RQ2: Are overlapping spatiotemporal patterns suitable for vibrotactile skin reading on the hands and forearms? More specifically, what performance on the recognition of letters and words can participants achieve with few hours of training?

The work in this chapter has already been reported in peer reviewed scientific papers [Luzhnica et al., 2016b, Luzhnica and Veas, 2019b, Luzhnica and Veas, 2019a, Luzhnica et al., 2018] (**P1**, **P3**, **P4**, **P5**) and one peer reviewed poster (**P8**). Moreover, the findings of this work constitute for the scientific contributions **C3**, **C4**, **C5**, **C6**, **C7** and **C10** as listed in Section 1.2.

4.1 Study 5: Investigating Skin Reading using Wearable Vibrotactile Displays on the Hand and Forearms

The main objective of this study is investigate to what extent participants can learn to successfully decode vibrotactile encoded messages (skin reading) in English using the designed described vibrotactile displays and the provided encoding. Initially, it provides encoding of the entire English alphabet and a specific training program designed for our purpose. Participants are exposed to the encoding using a provided training program which trains them to associate the encoded information with their representative OST patterns. They are trained to recognise symbols and then tested on symbols and words stimulated on two different body locations: hand and forearms. Participants also are tested continuously to evaluate their performance in skin reading.

Although not a direct objective, the results of this study are used to identify issues that hinder performance in the task of skin reading in general. By using two different layouts, it ensures that the findings are not bound to body part or layout. This study uses the exact wearable layouts, and OST patterns designed and developed and evaluated during the Study 1 (see Section 3.1).

4.1.1 Wearable Layouts

Two layouts were used for stimulation in different body parts. Figure 4.1 and Figure 4.2 illustrate the location of vibromotors in each layouts. The wearables use an Arduino-Duo board coupled to a power regulator (LM2596S) which controls 3.4 mm vibrotactile motors of type ROB-08449 (Voltage range: 2.5V 3.8V; Amplitude vibration: 0.8G).

Hand layout. Six vibromotors are placed on the back of the hand and fingers, so as to avoid interfering with grasp and hand interactions. On the fingers, vibromotors are placed on the middle phalanx leaving the fingertips free and kept uncovered by utilising a partially finger-less glove as shown in Figure 4.1.

Forearms layout. On each forearm, two vibromotors were placed on the outer side (extremes), and one in the middle on the inner side of the forearm, see Figure 4.2 fixed by elastic adhesive bandages. In practice, the motors could be placed within two wearable sleeves in a wearable consumer product.

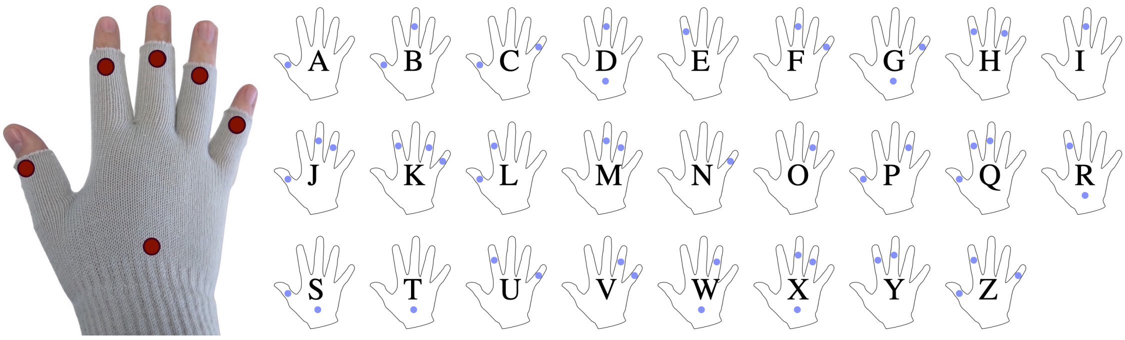

Figure 4.1: A finger-less glove for the "Hand" layout with the positions of vibromotors annotated and the encoding of letters [Luzhnica et al., 2016b].

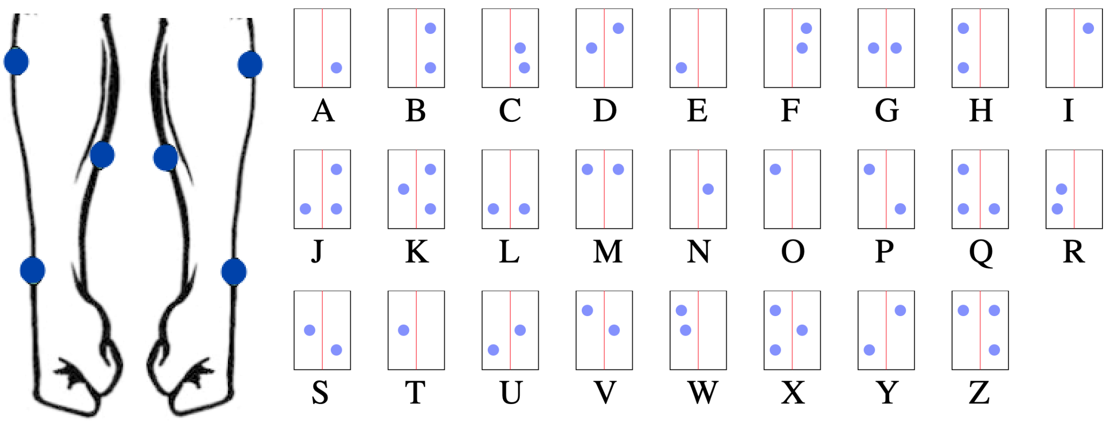

Figure 4.2: Letter encoding for forearms layout. Only active motors that are used for encoding the letters are displayed.

4.1.2 Patterns and Encoding

Each letter is encoded in patterns of one, two or three vibromotors using the overlapped spatiotemporal stimulation. OST delivers patterns that are very short and yet can be discriminated with a high accuracy (see Section 3). The activation of vibromotors is done in sequence after a gap (g) of 10 ms, but they share most of the activation time, as illustrated by Figure 4.5. When encoding words, there is a between-letter gap (see Figure 4.5) to separate them. The encoding considers the frequency of letters [1] to encode the most frequent letters each with a single vibromotor, and increases the number of vibromotors with decreasing frequency of letters (see Figure 4.1 and Figure 4.2). Thereby, common letters are transmitted faster, contributing to a higher throughput. This encoding results in 6 single-vibromotor letters, 15 two-vibromotor and 5 three-vibromotor letters.

4.1.3 Procedure

The study was split into five sessions as proposed by Luzhnica et al. [Luzhnica et al., 2016b]. Each session consisted of several different rounds, each serving different purposes:

Letter Training (LT) introduced letters in a number of repetitions for participants to learn to associate them with vibrotactile patterns. Each letter would be stimulated with the vibrotactile pattern (haptic cue), displayed on the screen (visual cue), and spelled via speakers (audio cue). A training round would repeatedly show letters in a predefined sequence.

Letter Reinforcement (LR) allowed participants to test if they recognised a letter. They were stimulated with a (randomly ordered) pattern and asked to enter (keyboard) the letter associated with it. After entering the answer, the letter was spoken and displayed in green if correct, red otherwise. Participants could repeat the stimuli before answering by pressing space bar. LR rounds aimed to give feedback and served to record performance and progress.

Word Challenge (WC) stimulated full words, which participants had to type in. Participants could repeat the entire word but not single letters within the word.

Participants were only gradually introduced to single new letters at a rate of 16, 5, 5 for each session $1-3$– the full English alphabet of 26 letters. WC rounds were introduced from session 1, to keep participants engaged with a feeling of progress.

[1] `https://en.wikipedia.org/wiki/Letter_frequency`

#S	LT	LR	Letters	BD	Words/ unique	BL Gap	TD
1	8	4	16	100ms	16 / 4	800ms	45
2	10	5	21	100ms	26 / 6	800ms	70
3	5	3	26	100ms	82 / 40	500ms	75
4	3	2	26	100ms	48 / 48	250ms	45
					25 / 25	150ms	
					25 / 25	100ms	
5	3	2	26	70ms	48 / 48	250ms	45
					25 / 25	150ms	
					25 / 25	100ms	

Table 4.1: Training program: number of training rounds (LT), reinforcement rounds (LR), number of letters, number of words, stimulation base duration (BD), between letter (BL) gap and total session duration (TD) in minutes for each session (#S).

The goal of sessions $4-5$ was to train participants with shorter letter duration and time gaps between letters in a word. The words used in sessions 4 and 5 were composed of 2-5 letters with an average length of 3.5 letters per word. Table 4.1 presents the number of training and testing rounds in each session, the number of letters, words and the stimulation parameters (stimulus base duration for each letter; gap time between letters when transmitting a word). Additionally, Figure 4.3 elaborates on the training procedure for a single session.

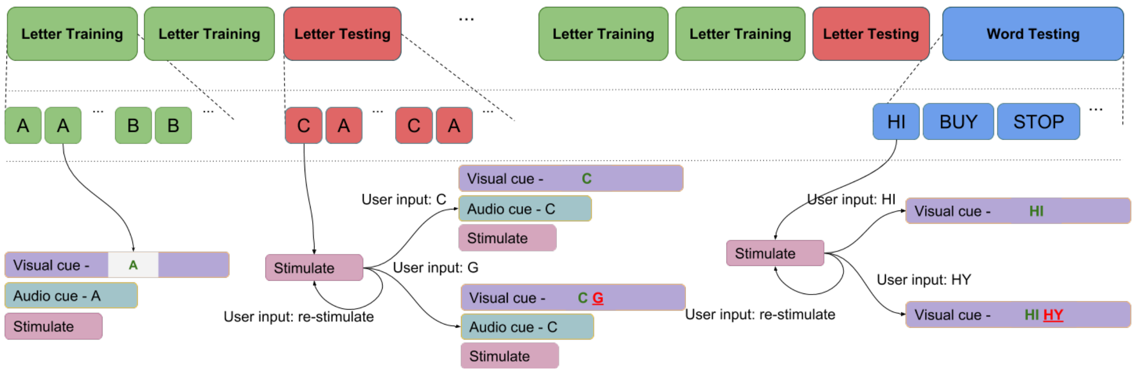

Figure 4.3: Training program process for a session during the studies 5 and 6.

Participants

Sixteen participants (12 males and 4 females) aged between 20 and 37 volunteered for this user study. Half of them (6 males and 2 females) used the hand layout and the rest used the forearms layout. None of them had any prior experience with skin reading, and none of them was a native English speaker. Regardless of their performance, they were rewarded for their participation with 30 euros in the form of a voucher. Figure 4.4 shows two participants, one using the hand layout and the other using the forearms layout.

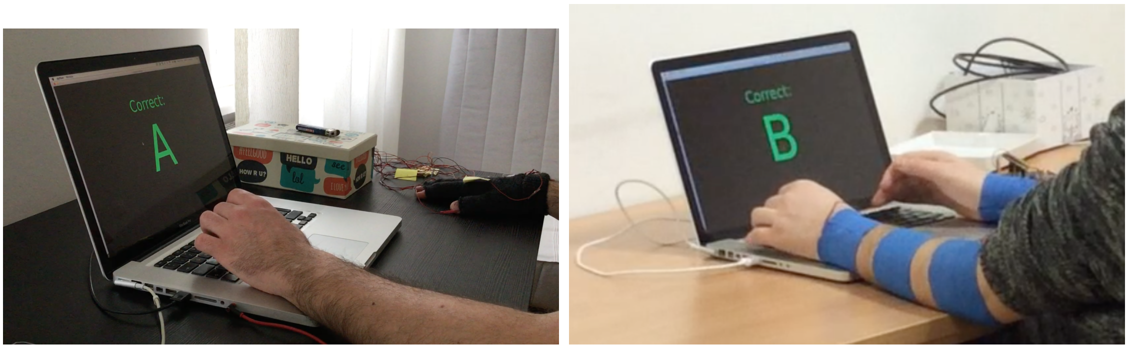

Figure 4.4: Participants during the Study 5 using the hand and forearms layout.

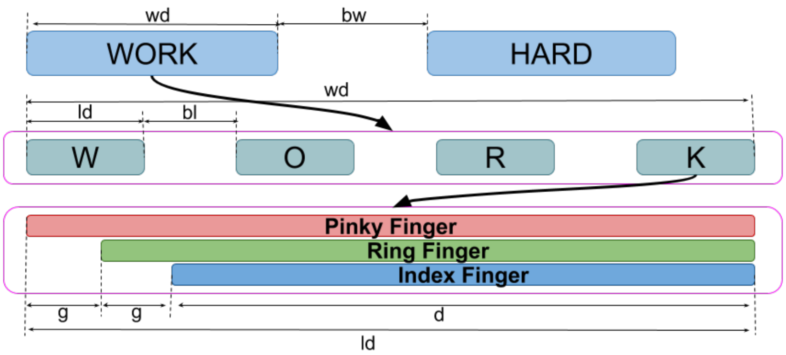

Figure 4.5: Stimulation of letters and words. Letters are encoded using OST patterns where vibromotors are activated in sequence with a gap in between (g). Within words, letters are transmitted in series with a gap in between (bl).

4.1.4 Results

For each session, the last LR and WC rounds were considered to indicate performance by:

- **Accuracy**: whether the user correctly recognised the letter or word. Letter accuracy is a binary value defined to be 1 when correctly recognised. For words, the accuracy is computed as $\sigma(a,s) = 1 - \frac{d(a,s)}{\#s}$, where d is a Levenshtein distance between a stimulated word (s) of length $\#s$ and user's answer (a), which is a common measurement to compare words.

- **Response time**: the difference between the response time-point and the first stimulation time-point including repetitions.

- **Re-stimulation**: whether participant repeated the stimuli.

	Hand			Forearms				
#S	Accuracy	TTR (s)	Re-stim	Accuracy	TTR (s)	Re-stim	UL	TL
1	.98 (.15)	2.8 (2.2)	.33 (.47)	.85 (.35)	2.6 (1.5)	.22 (.41)	18	54
2	.93 (.25)	2.7 (1.9)	.35 (.48)	.90 (.30)	3.1 (2.5)	.37 (.48)	23	92
3	.92 (.27)	3.4 (4.2)	.47 (.50)	.84 (.36)	3.9 (3.8)	.45 (.50)	26	52
4	.89 (.31)	2.9 (1.9)	.31 (.46)	.91 (.29)	3.5 (2.9)	.46 (.50)	26	52
5	.92 (.28)	2.6 (2.3)	.28 (.45)	.90 (.30)	3.5 (3.5)	.44 (.50)	26	52

Table 4.2: Letter recognition accuracy, time to respond (TTR), the re-stimulation rate (re-stim), number of unique letters (UL) and all tested letters (TL) for each session (#S). The values correspond to the average and standard deviation of the given variable.

	Hand			Forearms		
#M	Accuracy	TTR	Re-stim	Accuracy	TTR	Re-stim
1	.95 (.22)	2.0 (1.2)	.15 (.36)	.95 (.22)	3.0 (2.8)	.38 (.49)
2	.92 (.27)	2.9 (2.3)	.31 (.46)	.92 (.27)	3.5 (3.1)	.45 (.50)
3	.80 (.41)	3.3 (2.0)	.43 (.50)	.79 (.41)	4.4 (3.8)	.55 (.50)

Table 4.3: Letter recognition accuracy (sessions 4 and 5), time to respond (TTR) and the re-stimulation rate (Re-stim) depending on the number of stimuli used to encode the letter.

Letters

First, let us analyse the letter recognition accuracy across last LR rounds of all sessions. The results for both layouts are presented in Table 4.2. Already in the first day, participants were able to identify letters with an accuracy of 98% for the hand layout and 85% for the forearms layout. With increasing number of letters in the next two sessions, the recognition accuracy balances to $90\% - 93\%$ (depending on the layout) in session 2 and $84\% - 92\%$ in session 3.

Let us focus on the last two sessions when participants were already trained on the entire Alphabet and the information was transmitted at a very high speed (70-90 ms for a letter in session 5 and $100 - 120$ ms in session 4). In session 4, participants achieved an accuracy of 89% in hand layout and 92% in forearms layout. In session 5, they achieved an accuracy of 92% in hand layout and 90% in forearms layout.

For the rest of the analysis, the focus will be put on the last two sessions as they use a higher speed and also represent a state where users already are trained on the entire alphabet.

Since accuracy is a binary variable, McNemar's test (paired) will be used to compare the results of the two sessions which revealed no significant differences in accuracy between the two sessions; $\chi^2(1, N = 1664) = 0.064, p = 0.8$. Thus, in the further analysis the two sessions will be analysed together. Similarly, (unpaired) chi-squared analysis reveals that there are no significant differences in accuracy between the two layouts; $\chi^2(1, N = 1664) = 0.01, p = 0.93$. Thus both layouts will be combined for the further analysis.

Table 4.3 presents the recognition accuracy depending on the number vibromotors that encode a letter. To analyse its effect, let us first average the accuracy by user, session, layout and number of vibromotors as we deal with unbalanced sets. As the aggregated accuracy is not normally distributed (Shapiro-Wilk test: p=0.0) and we deal with paired samples, the Wilcoxon signed-rank test is used to determine the statistical differences. Thus a Wilcoxon signed-rank test reveals that there is no significant change in accuracy between one-vibromotor and two-vibromotor letters; $r = 101.5, p = 0.059$, but there is a significant difference between two-vibromotor and three-vibromotor letters; $r = 42.5, p = 0.0$.

Words

The accuracy depending on the session and gap time between consecutive letters (bl) when transmitting a word is presented in Table 4.4. Using a gap of $500ms$ in the third session results in an accuracy of $h : 90 - a : 91\%$. However as the gap time between the letters is reduced and also the stimulation time is decreased, the performance decreased as well. In sessions 4 and 5, three different gap times are used which yield three different results. Again, for further the analysis focus will be placed on the last two sessions because of the fast transmission speed and maturity of training. In session 4, depending on the gap, participants achieved an accuracy of $84 - 89\%$ with the hand layout and $84 - 87\%$ with the forearms layout. Similarly, in the last session, participants achieved an accuracy of $86 - 89\%$ with the hand layout and $86 - 88\%$ with the forearms layout.

Considering that the accuracy is a real value and it is not normally distributed (Shapiro-Wilk test: p=0.0), it will be relied on non-parametric tests to determine the significance levels. A rank-sum test reveals that there was no significant difference in accuracy between hand and forearms layout neither for session 5 ;$r = -0.8, p = 0.425$ nor for session 4; $r = 0.87, p = 0.385$. Additionally, there was no significant differences in accuracy between word recognition in session 4 and session 5; $r = -1.06, p = 0.29$.

		Hand		Forearms	
#S	BL Gap	Accuracy	Re-stim	Accuracy	Re-stim
1	800ms	.92 (.22)	.45 (.50)	.98 (.15)	.31 (.47)
2	500ms	.96 (.17)	.33 (.47)	.99 (.10)	.40 (.49)
3	500ms	.90 (.22)	.90 (.29)	.91 (.21)	.94 (.23)
4	250ms	.84 (.28)	.92 (.27)	.84 (.28)	.89 (.31)
4	150ms	.89 (.23)	.84 (.37)	.87 (.24)	.90 (.30)
4	100ms	.89 (.22)	.90 (.31)	.87 (.24)	.95 (.22)
5	250ms	.86 (.28)	.88 (.33)	.86 (.28)	.91 (.29)
5	150ms	.85 (.30)	.86 (.35)	.86 (.27)	.87 (.34)
5	100ms	.89 (.22)	.86 (.35)	.90 (.22)	.89 (.31)

Table 4.4: Word recognition accuracy depending on the session and the gap (BL) between subsequent letters.

Characters within Words Letter testing rounds identified problems that are related to individual letter patterns evidencing that letters encoded with three vibromotors are more difficult to recognise. However, the errors in word recognition can not be explained through them. Due to the letter frequency based encoding, the five letters encoded with three vibromotors appear rather infrequently in words.

According to the Google N-gram Corpus [Michel et al., 2011][2,3], their frequency in the text is 0.54% for K, 0.23% for X, 0.16% for J, 0.12% for Q and 0.09% for Z. So, all four combined have a frequency of 1.04%. To put that in perspective the most frequent letter E has a frequency of 12.49%. In the words used in sessions 4 and 5, 55.2% of the letters are encoded by one vibromotor, 41.4% by two vibromotors and only 3.4% of letters were encoded by three vibromotors. Moreover, the average word recognition accuracy of all used words in sessions 4 and 5 for both layouts is 87.3%. When discarding the words that contain any three vibromotor letters, the word recognition accuracy increases only slightly to 88.4%. Thus, the overall word recognition error rate cannot be explained by the poorer recognition of three vibromotor letters.

Nevertheless, when conveying words, patterns are transmitted in a sequence, and this might introduce additional problems. On sessions 4 and 5 where the transmission speed was higher, participants mentioned that sometimes, when two sequential letters (transmitted within a word) share one vibromotor, it became more difficult to distinguish whether that vibromotor was active in both letters or only in the first one. For instance, if when the bigram of WH (e.g. in the word WHAT) is transmitted using the hand layout, the letter W and H share a vibromotor in the ring finger. According to the claims, it would be hard to determine whether the shared vibromotor was active in both letters or the second letter was only composed of vibromotor located in index finger (which would then result in the letter E). Thus, let us investigate the claim by analysing whether the sharing of a vibromotor between the patterns of two subsequent letters within a word has an effect on the recognition of the letters. This could be achieved by examining the recognition of single letters within all bigrams of tested words. For that, let us exclude the ones that contain letters encoded with three vibromotors. Such letters have a high probability of not being correctly identified alone and also have the higher probability of sharing vibromotors (they have more vibromotors), thus they would skew the results.

For letters that share vibromotors, the number of vibromotors that encode the current letter will be taken into consideration. The hypothesise is that when the current letter is composed of only one motor, then it should not create confusion as it is easier to perceive one stimulus than two simultaneously. Thus the three types

[2] `storage.googleapis.com/books/ngrams/books/datasetsv2.html`

[3] `norvig.com/mayzner.html`

of sequential letters will be noted as:

- **S1** - current letter shares one vibromotor with previous letter and it is encoded by only one vibromotor (16% of bigrams),
- **S2** - current letter shares a vibromotor with previous letter and it is encoded by two vibromotors (14% of bigrams),
- **NS** - current letter shares no vibromotor previous letter.

Figure 4.6 presents the recognition accuracy of such sequential letters. To further analyse, let us first average the accuracy by user, session, layout and number of vibromotors as we have imbalanced sets. Given that the aggregated accuracy is not normally distributed (Shapiro-Wilk test: p=0.0) and we deal with paired samples, the Wilcoxon signed-rank test will be used to determine significance. A combined test for sessions 4 and 5 for both layouts reveals that there is no significant difference on accuracy between NS ($M = 0.89, MDN = 0.92, STD = 0.08$) and S1 ($M = 0.89, MDN = 0.91, STD = 0.1$) ; $r = 215.0, p = 0.36$ but there is a significant difference between NS and S2 ($M = 0.7, MDN = 0.69, STD = 0.15$); $r = 0.0, p = 0.0$. It is worth noting that also when separately analysing the sessions 4 and 5, and the layouts, the results of the significant differences remain the same.

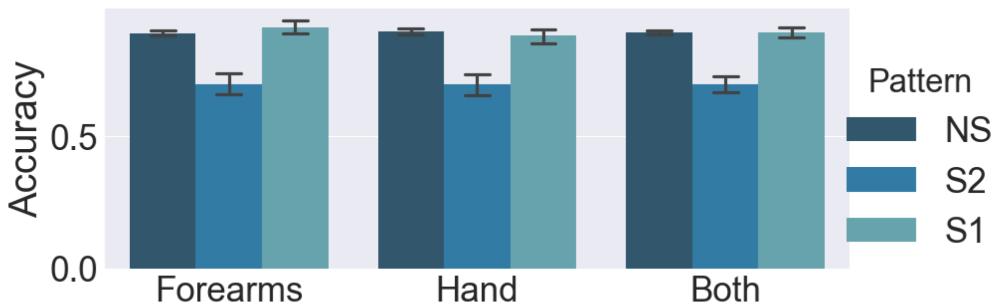

Figure 4.6: Letter recognition accuracy within words depending on whether the encoding of current letter shares a vibromotor with the encoding of the previously transmitted letter and the number of motors that encode the current letter.

4.1.5 Discussion

This user study used two layouts of vibrotactile displays: hand and forearms; the state of the art OST patterns [Luzhnica et al., 2016b] with frequency based letter encoding [Luzhnica et al., 2016b] and five sessions of training/testing. For both layouts, participants achieved a comparable and relatively good accuracy in the letter ($89\%-92\%$) and word recognition ($85\%-90\%$). Nevertheless, in both layouts, systematic errors occurred during both letter and word recognition.

When testing for individual letters, most problems occurred with letters encoded with three vibromotors. Thus, I propose to avoid encoding the letters with more than two vibromotors. For the forearms layout, the upper forearms could be used to place six additional vibromotors (three on each upper arm). With 12 vibromotors, 78 symbols could be encoded each using at most two vibromotors. As for the hand, eight vibromotor layouts have already been proposed and evaluated in terms of OST pattern identification (see Section 3.4). Eight vibromotors could encode 36 symbols. Additionally, a similar layout using two hands (with eight vibromotors each) could encode 136 symbols each using at most two vibromotors. In both cases, such a large number of patterns could be used to encode letters, numbers, punctuations and even commonly used words. However, using only seven vibromotors would be enough the encode the entire English alphabet. It would also reduce the sharing of the vibromotors between the patterns of subsequent letters, which is a problem when transmitting words.

When testing for words transmitted as a sequence of letters, another problem was identified for letters encoded by more than one vibromotor. Participants had significantly more difficulties in correctly recognising letters that are encoded with two vibromotors and share a vibromotor with the encoding of the previous letter (within the same word). Such letters were correctly recognised with an average accuracy of only $60\% - 65\%$ (depending on the layout). I propose to use the internal structure and representation of the used language to develop an alphabet encoding that minimises occurrences of such cases. Moreover, I provide a complete methodology for this process. The methodology starts by defining a metric which can be used to measure how often such cases occur in the text by considering the probability distribution of bigrams for the given language. Then, I model finding of an optimal encoding by minimising the score of the defined metric. For such an optimisation process I provide a greedy algorithm, which then is used to find an optimal encoding for any given number of vibromotors which is enough to encode every symbol with at most two vibromotors.

4.1.6 Optimising the Encoding: Compensating Perception Problems

While both layouts deliver good results, the previous study evidence that both layouts share the same perception problems:

1. Letters encoded with patterns composed of three vibromotors result in poorer recognition accuracy when compared with letters encoded with one or two vibromotors. Thus, the overlapping spatiotemporal patterns (OST) with three vibromotors are more difficult to identify correctly.

2. When two letters are transmitted in sequence (within a word), the recognition accuracy of the second letter is negatively affected when their corresponding patterns share vibromotors. However, this problem only occurs if the second letter is encoded by more than one vibromotor. Thus when applying two OST patterns in the sequence, the second pattern is prone to more errors if this pattern uses more than one vibromotor.

Both problems are related to perception. Numerous studies in neuroscience shed evidence that with exposure to stimuli there is a tactile learning effect [Reuter et al., 2014, Godde et al., 2000, Dinse et al., 2005, Pleger et al., 2001, Pleger et al., 2003, Hodzic et al., 2004] and thus with the usage of the vibrotactile displays perception might improve. However, there is no assurance on how much it will improve. Thus, in this work, methods to avoid or minimise their occurrences will be proposed constructed following a two-step optimisation process.

Step 1: Letters Encoded with Three Vibromotors

The perception problem of patterns encoded with three vibromotors can be solved by simply adding more vibromotors to the layout. This would result in more patterns encoded with only one or two vibromotors. The set of all available OST patterns OP composed of m vibromotors can be expressed as the set of all combinations with one or two elements:

$$OP = \{op_i\} = \binom{V}{2} \cup \binom{V}{1}, i \in [1, n] \tag{4.1}$$

$$n = \binom{m}{2} + m = \frac{m(m+1)}{2} \tag{4.2}$$

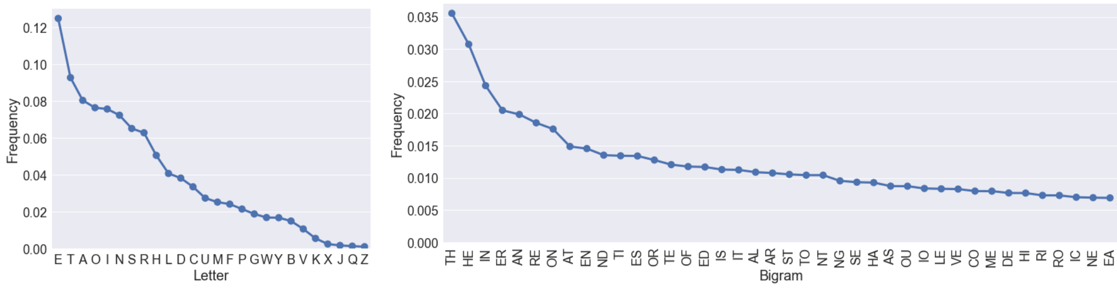

Figure 4.7: Letter frequency distribution and bigram frequency distribution (top 40 most frequent bigrams) for the English language based on Google Books N-gram Corpus.

where $V = \{v_i\}, i \in [1, m]$ is the set of vibromotors and n is the total number of patterns. A vibrotactile display with $m = 7, n = 28$ can encode the entire English alphabet. With $m = 8, n = 36$ and with $m = 9, n = 45$ it would be sufficient to also encode most of the punctuations and symbols. For the arms, there would be enough space for more vibromotors in the upper part of the arm. For the hand layout, more vibromotors could be placed on the back of the hand and the wrist to provide distinguishable OST patterns as already demonstrated in Chapter 3.

Step 2: Consecutive Letters with Shared Vibromotors

Totally avoiding consecutive patterns that share vibromotors is not feasible as long as patterns that use two vibromotors are used. However, their occurrences could be minimised by leveraging the morphological structure of the targeted language. For instance, in every language, some letters occur more often than others which is modelled by the letter frequency distribution[4]. Similar frequency distribution can also be observed for bigrams (a pair of consecutive letters). This information could be utilised to construct an alphabet encoding, such that pairs of letters that are more probable to appear in a sequence are encoded with patterns that do not share any vibromotor. Thus, the encoding of the alphabet could be treated as an optimisation problem which reduces the occurrences of bigrams with shared vibromotors. The distributions of both letters and (40 top) bigrams are presented in Figure 4.7.

To approach this problem, let us first introduce a cost function, which measures the proportion of consecutive letters with shared vibromotor occurrences, where the

[4] `https://en.wikipedia.org/wiki/Letter_frequency`

second letter is encoded by two vibromotors. The cost function assumes that each letter is encoded at most by two vibromotors. Let us denote the alphabet with k letters as $A = \{l_i\}, i \in [1, k]$ and the set of possible OST patterns as $OP = \{op_i\}, i \in [1, n], n \geq k$, where each pattern op_i is a set of one or two vibromotors as defined in Eq. 4.1. Also, let us define the bigram probability distribution: BF : $A \times A \to \mathbb{R}$ as $BF(l_1, l_2) = P(l_2|l_1)$ and a mapping function called encoding E : $A \to OP$. The cost function is defined as:

$$f(E, BF, A) = \sum_{l_1 \in A} \sum_{l_2 \in A,, l_1 \neq l_2} BF(l_1, l_2) \cdot |E(l_1) \cap E(l_2)| \cdot (|E(l_2)| - 1) \tag{4.3}$$

The purpose of $|E(l_1) \cap E(l_2)|$ is to count the shared vibromotors between patterns. The expression $|E(l_2) \setminus E(l_1)|$ ensures to discard bigrams that share vibromotors but the second letter in bigram is encoded by a single vibromotor. It also excludes accounting for bigrams composed of the same letters.

Given that the Study 5 (see Section 4.1) showed that letters encoded by three vibromotors are prone to recognition errors due to masking effect, it is strongly advised avoiding them, unless the reader in the future discovers any stimulation technique that bypasses masking. Therefore, a special case of the cost function when each letter is encoded by at most two vibromotors is given by:

$$f(E, BF, A) = \sum_{l_1 \in A} \sum_{l_2 \in A} BF(l_1, l_2) \cdot |E(l_1) \cap E(l_2)| \cdot |E(l_2) \setminus E(l_1)| \tag{4.4}$$

Having Equations 4.3 and 4.4 allows us to define the problem of avoiding occurrences of consecutive letters with shared vibromotors as a minimisation problem. For a given language with an alphabet A and a bigram probability distribution BF, the goal is to find a mapping E from letters to patterns (encoding) that minimises the objective function f:

$$E_o = \arg\min_E f(E, BF, A) \tag{4.5}$$

for a given alphabet A of length k and a set of n ($n \geq k$) patterns OP, a valid mapping E can be considered an element-wise pairing $(l_i, \hat{p}_i)$ of the alphabet A and any permutation $\hat{P} = \{\hat{p}_i\}, i \in [1, k]$ of OP with k elements. Therefore, the problem is reduced to a combinatorial optimisation of finding the best permutation for a given cost function. The aim is to find an optimal solution as finding the best solution would require the evaluation of every possible solution, which is not feasible. For the purpose of solving such optimisation, a greedy algorithm is developed. Existing approaches proven to provide optimal solutions in similar combinatorial optimisations problems such as genetic algorithms [Reeves, 1995, Chu and Beasley, 1997b] and simulated annealing [Chu and Beasley, 1997a, Osman and Potts, 1989] were also explored. However, in this case, both algorithms performed worse than the proposed greedy approach, and thus their details will not be discussed.

Greedy Approach

The algorithm starts with a random permutation $\hat{P}$ and employs an iterative process. At each iteration, it finds a pair of elements $\hat{P}[i]$, $\hat{P}[j]$ for which when being swapped, the resulting cost function score decreases the most. Then, it performs swapping until there is no room for improvement. The pseudocode is given in Algorithm 1. As with most of the greedy algorithms, it suffers from getting stuck at a local minimum, and the results highly depend on the initial permutation $\hat{P}$. To avoid getting a poor optimised local minimum, the algorithm should be executed many times (10000), where each time the initial permutation is selected randomly. The optimal solution is chosen as the permutation with the lowest resulting cost.

ALGORITHM 1: Greedy Encoding Optimisation

Input : $\hat{P}$ - an initial permutation of patterns, BF - bigrams probability distribution, A - alphabet, f - cost function as given in Eq. 4.3, $E(\hat{P}, A) : l_i \to p_i$, $\forall i \in [1, k]$, $l_i \in A$, $p_i \in \hat{P}$ - a function that maps element wise every i-th letter to i-th permutation, $swapped(\hat{P}, i, j)$ - a function that swaps the i-th element with j-th element

Output : $\hat{P}_o$ - a new permutation

$n \leftarrow |\hat{P}|$, $k \leftarrow |A|$, $\hat{P}_o \leftarrow \hat{P}$, $gain \leftarrow 1$

while $gain > 0$ **do**

  $i, j \leftarrow \arg\min_{i \in [1,n], j \in [1,n]} f(E(swapped(\hat{P}, i, j), A), BF, A)$

  $\hat{P} \leftarrow swapped(\hat{P}, i, j)$

  $gain \leftarrow f(E(\hat{P}_o, A), BF, A)$ - $f(E(\hat{P}, A), BF, A)$

  if $gain > 0$ **then**

    $\hat{P}_o \leftarrow \hat{P}$

  end

end

Bigram Probability Distribution

The Google Books N-gram Corpus [Michel et al., 2011][5] was used for the bigram probability distribution. It contains frequency distribution of all possible 2-5 n-grams. The data is constructed by scanning over five million books holding over a trillion words.

Optimisation Results

As the goal of this work is to provide an encoding and layout that avoids both issues occurring in letter and word recognition, seven vibromotors will be utilised. This allows to encode letters with at most two vibromotors as three vibromotor letters should be avoided. Although, when the five least frequent letters are encoded with three vibromotors they are expected to appear rarely, they will still appear in real-world usage. Moreover, since the new encoding will be optimised based on bigram sharability, there is no guarantee that exactly the least common letters will be encoded by three vibromotors.

The result of the greedy approach for seven vibromotors delivered an optimal permutation with a cost function score of 0.038. Note that there are many solutions with the same result (lowest score) as vibromotors can be arranged differently and still preserve the same relationship between patterns. Thus, one of the optimal encodings with the lowest score was chosen, which is presented in Table 4.5. Interestingly the most frequent letters are encoded with only one vibromotor without the cost function explicitly aiming for such an optimisation. Obviously, the most frequent letters form the most frequent bigrams. In comparison, the encodings used in Study 5, would get a score of 0.184 for the hand and 0.188 for the forearms layout.

To outline the effects of such optimisation, let us observe a text snippet from the "I have a dream" speech by Martin Luther King Jr. The text is encoded with the optimal encoding and with a randomly generated one. Clearly, when using optimal encoding, letters share very rarely a vibromotor with the previous letter (see Listing 4.1). Please note that the algorithm does not optimise for double letters (e.g. ll) as there it is impossible to avoid sharing vibromotors. The optimised encoding with seven vibromotors for the hand layout is visualised in the Figure 4.8. The vibromotors on the hand are positioned based on the recommendation of Luzhnica

[5] `storage.googleapis.com/books/ngrams/books/datasetsv2.html`

and Veas [Luzhnica and Veas, 2017].

I ha**v**e a dream that my four little chil**d**ren wil**l** one day live in a nation where they wil**l** not be jud**g**ed by the color of their skin but by the content of their character.

I h**ave** a d**re**am th**a**t my four li**ttle** c**hi**ld**ren** w**ill** o**ne** d**a**y liv**e** i**n** a n**a**t**i**o**n** whe**re** they w**ill** no**t** be judg**ed** by the co**l**or of thei**r** s**kin** but by the co**ntent** of thei**r** c**ha**ract**er**.

Listing 4.1: Illustrating the effects of different encodings. The underlined bold letters number of letters that are encoded with two vibromotors and share a vibromotor with the previous letter in the same word. The text is part of Martin Luther King Jr.'s "I Have a Dream" speech and it is encoded with the optimal encoding (top) and with a random encoding (down).

Letters	A	B	C	D	E	F	G	H	I
Vibromotors	1	3,6	5,6	5,7	3	1,6	3,7	1,2	4
Letters	J	K	L	M	N	O	P	Q	R
Vibromotors	2,3	3,4	2,7	6,7	2,6	5	3,5	2,5	2
Letters	S	T	U	V	W	X	Y	Z	
Vibromotors	7	6	1,4	1,7	4,7	2,4	4,5	1,3	

Table 4.5: The resulting encoding (letter, vibromotors) for seven vibromotors. The encoding can be applied to any layout.

4.2 Study 6: Evaluating Optimised Encoding

The previous study (Section 4.1) evidenced systematic errors on recognition of letters and words. Moreover, Section 4.1.6 proposes a two step optimisation process which promises to avoid such systematic errors and as a result it proposes a new encoding of letters. In order to observe the effect of the encoding optimisation, another user study is conducted where the setup is kept relatively close to the previous study, but the optimised encoding and layout is used.

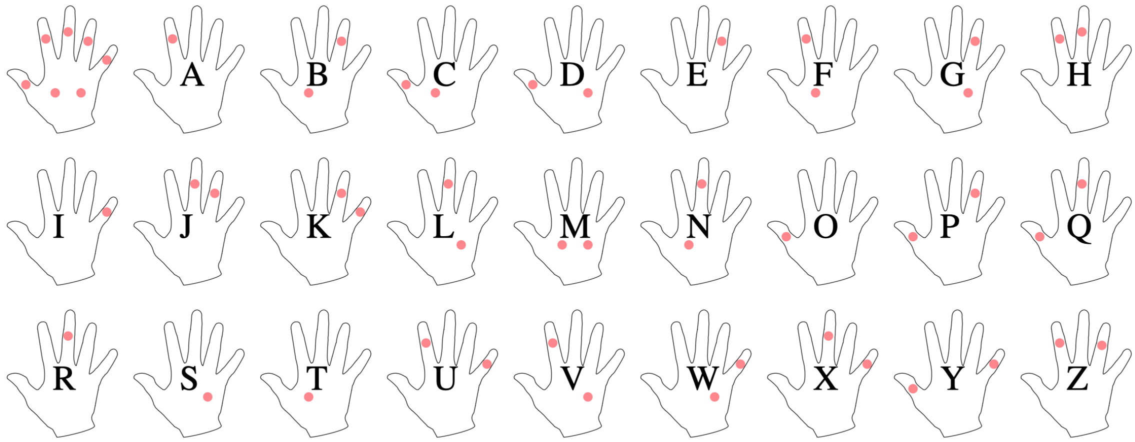

Figure 4.8: The optimised encoding for seven vibromotors. The positions of the vibromotors are selected following the recommendation of Luzhnica and Veas [Luzhnica and Veas, 2017].

4.2.1 Procedure

In this study, a hand layout with 7 vibromotors is used, which allows to encode every letter of English alphabet with at most two vibromotors. The locations of the new optimised layout are based on the previous investigations (based on Study 4 four described in Section 3.4) on suitable positions of actuators

The study setup, rewarding scheme, and procedure including the training and testing program (the letters and words) is borrowed from previous user study (Section 4.1). Nevertheless, in order to investigate other phenomena, some minor changes are applied in the study design without jeopardising the comparability with the previous user study (Section 4.1):

1. At the beginning of sessions 2-5 (before training), a recall test only with letters was performed, assessing how well participants recall the letters that they were trained on previously.

2. At the end of session 5, participants were tested with a round of letters on the left hand, to evaluate how well they are able to transfer the skill of decoding letters on an untrained body part. The layout was simply mirrored on the left hand.

3. After the last training session (5), participants were invited twice and exposed

to one round letter testing in order to evaluate how well they could still recall the letters when no training or usage is provided for a period. Such letter test rounds were assigned to be (i) three days and then (ii) ten days after the session 5. I will will refer to them as recall session 6 and 7.

4. Users filled an NASA TLX questionnaire for every session and also a questionnaire evaluating the wearability and comfort of the prototypes. However due to the focused scope, they will be not discussed in this section.

5. At the end of the session 5, in addition to NASA TLX, participants were required to answer a simple questionnaire containing one question and two statements to rate as follows:
   Question 1: Would you consider using such a wearable glove to perceive information in any activity/task of your daily life/work? If yes in what situations? The participants had to choose one answer between "YES", "NO" or "MAYBE" and provide text for situations.

   - *Statement 1: Vibrotactile glove was very uncomfortable to wear!* The participants had to rate the statement using a five levels Likert scale (from strongly agree to strongly disagree). Additionally, they were able to provide any comment regarding uncomfortability to justify their statement.
   - *Statement 2: The vibration in my hand did feel unpleasant!*. Participants had to rate the statement using a five levels Likert scale, and they could provide comments if they had any.

Participants

Eight individuals participated in the study (seven males and one female) aged between 21 and 34. None of the participants had any prior experience with skin reading, and none of them was a native English speaker. All of them used the left hand for stimulation and the right to interact with the computer.

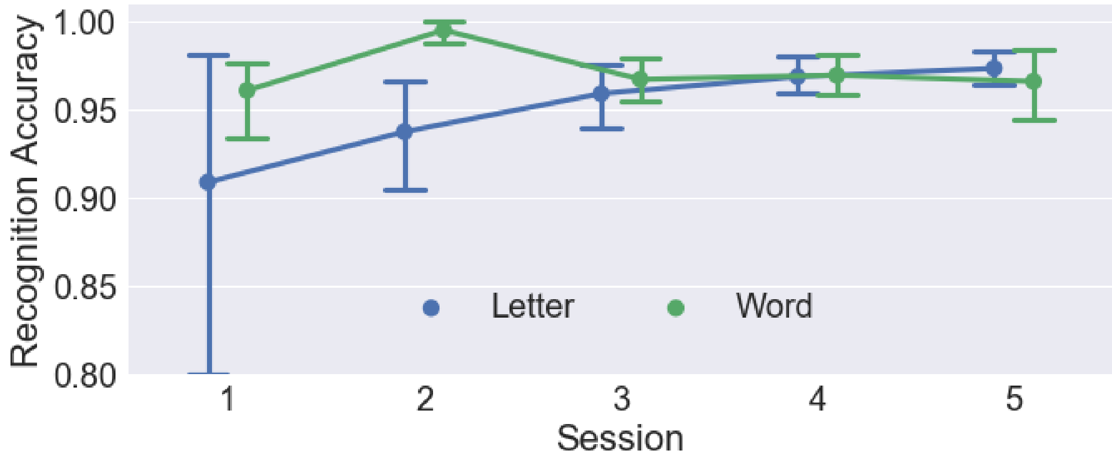

Figure 4.9: Letter and word recognition for each session. Data is initially averaged per user and session to emphasise the variation among users which is expressed via error bars representing the 95% confidence intervals.

4.2.2 Results

Letters

First, let us analyse the letter recognition accuracy across last letter testing rounds of all sessions. The results are presented in Table 4.6. Additionally, the results grouped by user and session are visualised in Figure 4.9. The grouped averaging is performed to emphasise the variations among users. When looking at the letter recognition (Figure 4.9) on the first day, participants on average achieve the lowest accuracy (among all sessions), and at the same time, the variation among users was the highest. As the training session followed, the accuracy increased and at the same time as the variation among users decreased, meaning that at the end they achieved a comparable performance (see Figure 4.12) but with some differences in the learning rate along the way.

For the rest of the analysis, analogously to the previous study, let us focus on the sessions 4 and 5 where the letter recognition accuracy is 97% which is a major improvement over six vibromotor hand layout used in previous user study (Section 4.1). When comparing those two layouts, chi-squared analysis reveals that the differences in the letter recognition accuracy are significant in both session 4 ($\chi^2(1, N = 832) = 16.95, p = 0.0$) and session 5 ($\chi^2(1, N = 832) = 12.17, p < 0.001$). This improvement is attributed to the absence of the three vibromotor encoded

	Recall			Post-Training		
#S	Accuracy	TTR (s)	Re-stim	Accuracy	TTR (s)	Re-stim
1				.91 (.29)	2.9 (2.7)	.41 (.49)
2	.92 (.27)	3.3 (4.5)	.27 (.44)	.94 (.24)	2.6 (1.9)	.28 (.45)
3	.95 (.22)	2.5 (1.8)	.16 (.37)	.96 (.20)	2.6 (1.9)	.22 (.42)
4	.92 (.27)	2.8 (1.7)	.20 (.40)	.97 (.17)	2.3 (1.3)	.17 (.38)
5	.98 (.15)	2.6 (2.0)	.17 (.38)	.97 (.16)	2.4 (1.8)	.22 (.42)
5L				.92 (.27)	3.6 (2.9)	.23 (.42)
6	.97 (.17)	2.3 (1.4)	.11 (.31)			
7	.94 (.24)	4.8 (8.1)	.26 (.44)			

Table 4.6: Letter recognition accuracy for each session using the optimised encoding on the hand layout. Post-training represents the results of the last testing round of the current session (at the end of the training). Recall tests of session 2-5 represent the testing round at the beginning of the session (before any training). Recall tests in sessions 6 and 7 represent letter testing 3 days respectively 10+ days after the session 5. The post-train of session 5 on the left hand is noted as **5L**.

letters in the optimised layout. A comparison of the results for both layouts is shown in Figure 4.10.

Additionally, the confusion matrix of letter recognition for last round of sessions 4 and 5 is visualised in Figure 4.11. It shows that from 32 probes per letter collected in the last rounds of session 4 and 5, most of the letters are confused once or at most twice. The letter L seems to be an exception, where it is confused four times, and two of them are with letter G which are very similar in terms of the patterns (see Figure 4.8).

Recall Tests

The recall tests of sessions 2-5, which were performed at the very beginning of each session, aim to investigate how much participants would remember from the previous days of training. The recall tests performed on sessions 6 and 7 investigate how the recognition of letters decays over time, as between session 5 and those recall tests, participants were not exposed to the usage of the vibrotactile device. The recall in session 6 was performed exactly 3 days after the last round of training and letter/word testing which corresponds to day 8 from the first session. The recall test in session 7 was initially planned to be exactly 10 days after the session 5 (15

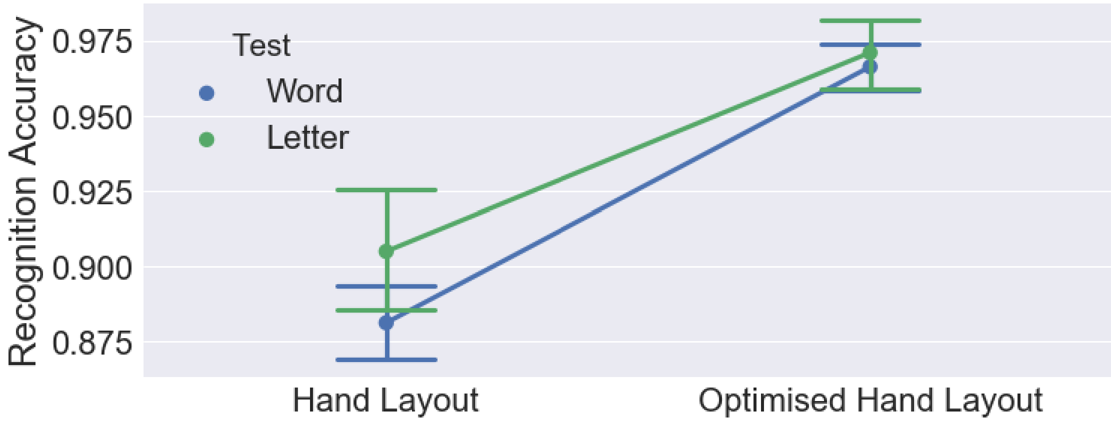

Figure 4.10: Letter and word recognition accuracy for both hand based layouts (with six vibromotors and the optimised with seven vibromotors) using the data from the sessions 4 and 5. Words that contain letters (K, J, Q, X and Z) are excluded.

days after session 1). However, in practice participants were not able to hold the plan. Thus, this test was performed between 10 and 19 days after the session 5 (15-24 days after the first session) for 7 participants whereas one participant did not perform it at all. Although initially unplanned, one participant performed an additional recall test (session 8) 45 days after the session 5.

The results are shown in Table 4.6 and the results of the recall for session 5 and onwards for each user are visualised in the Figure 4.12. Table 4.6 shows that for the next day recalls (sessions 1-4), except for session 3, in other sessions participants performed even slightly better on the recall tests on the next day compared to the last round of testing in the current session (post-training). On the recall of session 6 which was 3 days after the last training session, participants did as good (97%) as in the post-training test of the session 4 and 5. On the recall of session 7 which was performed 10-19 days after the last training session, the letter recognition accuracy dropped to 94%.The 8th recall session performed by only one user reveals that even after 45 days after the last training session, the participant was still able to correctly recognise 92% of the letters.

Left Hand

The results of letter recognition test on the left hand performed at the end of session 5 are presented in Table 4.6 and Figure 4.12. Table 4.6 shows that participants were

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	31	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B	0	31	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	31	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
F	0	0	0	0	0	31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
G	0	1	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
J	0	0	0	0	0	0	0	0	0	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	0	31	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
L	0	0	0	0	0	0	2	0	0	0	1	28	0	1	0	0	0	0	0	0	0	0	0	0	0	0
M	0	0	1	0	0	0	0	0	0	0	0	0	31	0	0	0	0	0	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	2	0	0	0	0	0	0	0	0	0
O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	32	0	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	31	0	0	0	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	32	0	0	0	0	0	0	0	0	0
R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	32	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	31	1	0	0	0	0	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	32	0	0	0	0	0	0
U	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	30	0	0	0	0	1
V	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	31	0	0	0	0
W	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	31	0	0	0
X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	2
Y	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	31	0
Z	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	31

Figure 4.11: The confusion matrix of the letter recognition using the last rounds of the sessions 4 and 5 of the Study 6.

able to recognise the letters with an average 92% accuracy. On the other hand, Figure 4.12 shows that the results might have been influenced by user 5 and user 7.

Words

The results of word recognition accuracy are presented in Table 4.7. Additionally, the results aggregated by user and session are shown in Figure 4.9. The grouped averaging is performed to emphasise the variations between users. It shows that the overall variability is quite low. On the sessions 4 and 5 where participants achieved a word recognition accuracy of $96\% - 98\%$ in both of the sessions. It is visible in Figure 4.9 that when looking at the aggregated data (by user and session), in session 5 the variation between different users (STD $= 0.031$) is a bit higher compared to session 4 (STD=0.017). Perhaps due to the shorter duration used in session 5 for conveying letters. However, the differences in accuracy between sessions 4 and 5 are not significant; $r = 0.05, p = 0.96$.

Let us compare the word recognition accuracy with the unoptimised six vibro-

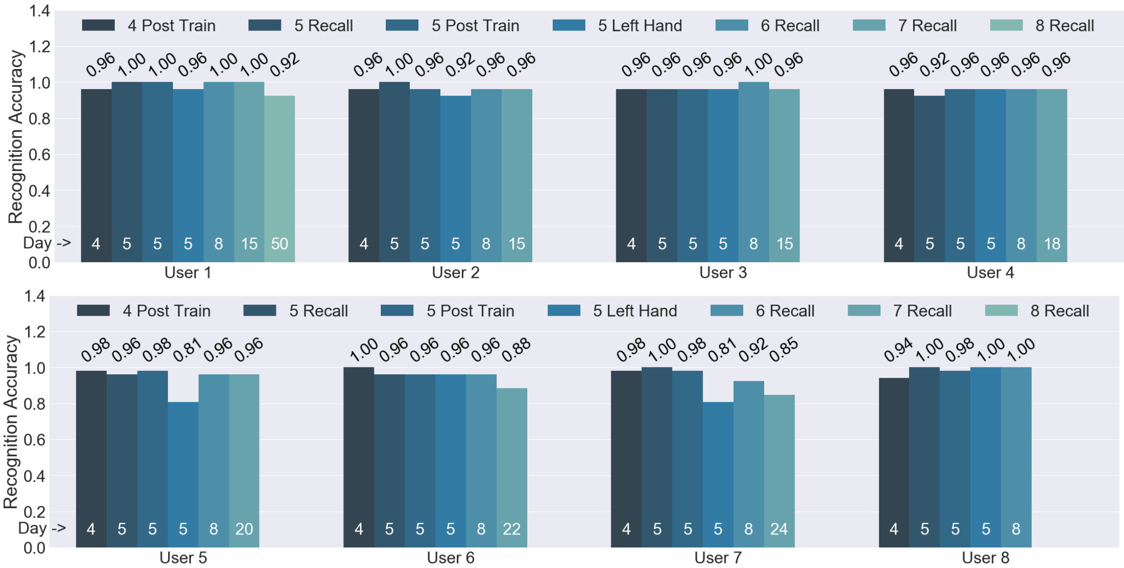

Figure 4.12: Letter recognition accuracy in post-training tests, recall tests and also the test on the left hand for each user. The white numbers on the bottom represent the day of the test it was performed relative to the first day of training (first session). Note that the last user did not perform a the 7th recall and the first user was the only one to perform the 8th recall test.

motor hand layout used in the previous user study (Section 4.1). The six vibromotor layout encodes five letters with three vibromotors whereas the optimal layout encodes all letters with at most two. Thus, to make the comparison fair, let us exclude from analysis all words that contain any of those five letters (K, J, Q, X and Z). A rank-sum test analysis reveals that participants recognised significantly better words (97% in both session) when using the optimised layout, compared to using the unoptimised six vibromotor hand layout (88% in both session) (see Section 4.1) in both sessions 4; $r = 5.31, p = 0.0$ and 5; $r = 4.72, p = 0.0$. Figure 4.10 shows the comparison for both sessions.

Questionnaire

The users rating on the how uncomfortable the wearable device was; and how unpleasant the vibrations felt are given in Figure 4.13. The overall ratings are very positive. From eight participants, all of them either disagreed or strongly disagreed that the vibrations felt unpleasant. Also, seven of them disagreed that the wearable device was uncomfortable to wear whereas one participant provided a neutral rating.

#S	BL Gap	Accuracy	Re-stim
1	800ms	.96 (.15)	.27 (.45)
2	500ms	1.0 (.05)	.12 (.33)
3	500ms	.97 (.11)	.80 (.40)
4	250ms	.98 (.12)	.74 (.44)
4	150ms	.96 (.14)	.73 (.45)
4	100ms	.97 (.12)	.80 (.40)
5	250ms	.98 (.13)	.67 (.50)
5	150ms	.96 (.18)	.68 (.47)
5	100ms	.97 (.13)	.73 (.44)

Table 4.7: Word recognition accuracy depending on the session (#S) and the gap (BL) between subsequent letters when using the optimised encoding in seven vibromotors hand layout.

Additionally, this participant noted that his rating was due to hand sweating. When asked whether they would consider using such a device in real life, four participants answered with yes, four others with maybe and none of them with no. Among use cases they would consider using it, they stated: while driving, while biking, while jogging, in presentations, for cheating in exams, for navigation, in meetings.

The results of NASA TLX for letter and word recognition tasks depending on the session (day) are depicted in Figure 4.14.In addition to the six metrics contained in NASA TLX, workload is calculated using the simplified R-TLX method (averaging all metrics where performance is inverted). In the case of letter recognition, the workload remained steady for the first three sessions as new letters were being introduced. However, the workload fell sharply in session 4 as they did not have to learn any new letters. In session 5, the workload increased compared to the session 4, as the transmission speed increased. When looking at word recognition, the workload was relatively low on the first two sessions where participants had to recognise only a few short words. Once the number of words (and their length) increased, the workload increased sharply on the session 3. However, once they got used to is, the workload gradually decayed on the two subsequent sessions.

Figure 4.13: User ratings on how unpleasant the vibrations felt and how uncomfortable was the prototype.

4.2.3 Discussion

The previous study found systematic erros in the skin reading. Section 4.1.6 proposed optimisation methods to improve the recognition acurracy of skin reading by avoiding such issues. To investigate the effects of such optimisation, this user study applied and evaluated the proposed encoding optimisations. The results show that participants were able to achieve a letter recognition accuracy of 97% in sessions 4 and 5, which is a significant improvement over the six-vibromotor layout used in Study 5 ($89\% - 92\%$). Besides, the encoding of the new layout is optimised to minimise the probability of two sequent letters in any word to share a vibromotor between them in order to improve the word recognition accuracy. This allows us to investigate the effects of such optimisation in word recognition by comparing its results with the ones from the hand layout from Study 5. The results show that, in sessions 4 and 5, participants were able to achieve an accuracy of $96\% - 98\%$ on the words which indeed is a significant improvement over the hand layout used on Study 5 ($85\% - 90\%$). The archived accuracies are outstanding considering only 5 hours of training. for comparison, sighted users have been reported to need six months of training to for recognising the Braille alphabet [Bola et al., 2016] and three additional months for reading words [Bola et al., 2016].

Note that, the optimal hand layout (in this user study) encodes every letter by at most two vibromotors whereas the unoptimised layout (in previous user study)

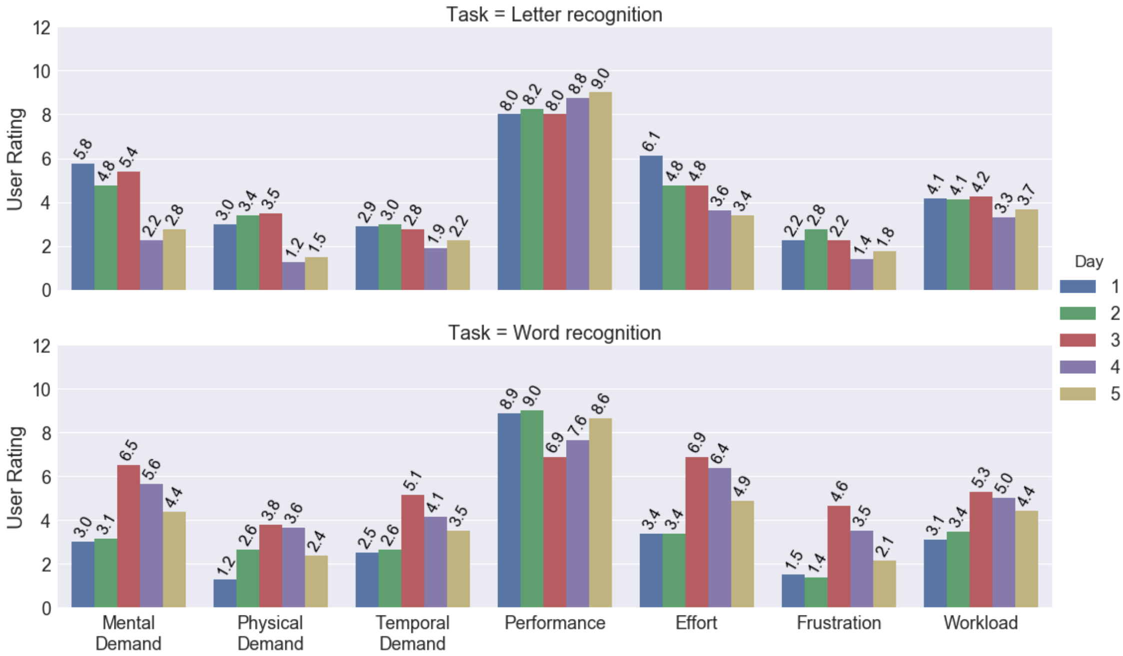

Figure 4.14: NASA TLX self-evaluation metrics for letter and word and recognition tasks per each session. Notation: SA - strongly agree, A - agree, N - neutral, D - disagree, SD - strongly disagree.

encodes five letters (K, J, Q, X and Z) by three vibromotors. However, when performing a statical comparison of word recognition, the words that contain any of those five letters are excluded. This ensures that only the optimisation effects of sharing among bigrams are taken into account and excludes the influence three vibromotor letters.

The second user study also tested how well participants would be able to transfer the recognition of letters on the body parts that they did not train on. In practice, this would mean that participants could easily change the body position of the wearable device and be able to perceive the information without re-training. In this user study, participants trained on the right hand and were tested at the end of the session 5 on the left hand. Participants were able to recognise letters on the left (untrained) hand with an average accuracy of 92%.

Besides post-training letter testing (at the end of a the session), recall tests at the beginning of a session (for sessions 2-5) were performed. Interestingly, when comparing the post-training letter accuracy with the letter accuracy of the recall session the next day, results show that except for one session (3), the accuracy in the recall was even slightly better than in the post-training. This suggests that the used training program might be unnecessarily extensive. Participants demonstrated that on the very next day, they had a very good recollection of letters learned on the previous day. Thus, re-training them on the letters of the previous session might not have been necessary in the performed extent. Additionally, participants were asked to do recall tests 3 days after and 10 days after the last training session. All participants completed the recall test scheduled 3 days after the last training. Unfortunately, due to conflicts with participants' private schedules, the 10 days after the last training session proved infeasible to hold. Nevertheless, seven participants were able to perform this recall session within 10 to 19 days after the last training session. In the recall test 3 days after the training, they performed with 97% accuracy, as good as on the day of the training. In the recall test 10-19 days after the last training, they were still able to recognise letters with an accuracy of 94%. Moreover, one participant performed an additional recall test 45 days after the last day of training and was able to recognise the letters with 92% accuracy.

The proposed encoding scheme is tailored for English language and layouts with seven vibromotors. The proposed bigram optimisation methodology can be used for any language and number of vibromotors. The methodology could be applied to other units of encoding such as phonemes [Zhao et al., 2018]. To use OST patterns for encoding phonemes, the same methodology could be applied, simply by constructing the bi-phoneme frequency (frequency of subsequent phonemes) distribution and then replacing the bigram frequency distribution with bi-phoneme frequency distribution in the cost function defined in Equation 4.3.

4.3 Study 7: Background Perception of Vibrotactile Encoded Messages

Previous user studies (see Sections 4.1 and 4.2) have demonstrated the feasibility of conveying vibrotactile encoded information efficiently using vibrotactile wearable devices. Users can understand vibrotactile encoded symbols and complex messages combining such symbols. Clearly, perceiving information through the skin using wearable devices can be beneficiary for a broad spectrum of applications for visual and/or hearing impaired users by utilising sensory substitution where a sensory modality (e.g. vision or auditory) is captured, processed and then transmitted to a user via vibrotactile stimuli.

Moreover, given the efficiency (perception accuracy, speed of transmission), vibrotactile skin reading using wearable devices could be useful beyond impaired sensory substitution, in general purpose applications to facilitate multitasking or reduce demands on the predominant visual displays. For instance, users would be able to perceive their phone notifications, SMS, emails etc... while performing other task e.g. driving, biking, working, etc... For most of such use case, multitasking is a key aspect. Nevertheless, for multitasking, it would be necessary for the perception and comprehension of vibrotactile information to be less attention demanding and not interfere with other parallel tasks.

This section presents a user study which investigates whether vibrotactile tactons which represent letters of English Alphabet can be concurrently perceived in the background (as secondary task) while performing a task that requires full attention. For the user study, the recruiting was limited to only participants who are trained in skin-reading and are proficient on understanding such tactons. The hypothesis is that the perception of vibrotactile encoded symbols should be handled by automatic processing and should not affect the performance of the other primary task which is designed to be challenging and requires controlled processing. Thus, this work targets the following research questions:

- **Can trained users perceive and decode high-speed vibrotactile encoded symbols in the background while performing another attention demanding primary task?**
- **How does the background perception of such vibrotactile encoded symbols affect the performance of the primary task?**
- **How does the primary task affect the performance of the background perception of vibrotactile encoded symbols?**

4.3.1 Background Processing of Information

Cognitive processing theory suggest that, when users are presented with multiple stimuli/tasks while multitasking, they prioritise or ignore some of them if attention bottlenecks occur [Schneider et al., 1982]. There are two categories of cognitive processes: controlled and automatic, which is determined by the amount attention needed. Automatic processes occur without the need for attention and process initiation, are considered effortless and do not draw general processing resources. As such, they do not interfere with other parallel occurring thought processes [Uleman and Bargh, 1989]. On the other hand, controlled processes are considered very flexible but very costly at the expense of the attentional resources available [Uleman and Bargh, 1989]. Schneider et al. [Schneider and Shiffrin, 1977] showed that the very same tasks could be processed using one or other processing model depending on the user experience and proficiency on the given task. In their short memory experiment, in a condition where parameters (search target) of the task changed, users needed constant attention for solving it, and thus the task required controlled processing. When performing it in parallel with another task that required controlled processing, the performance on both tasks declined. On the contrary, in a condition where the same parameters were kept constant, after users gathered enough experience and became proficient with the task, the processing became automatic. When combined with another task that required controlled processing, the performances of both tasks were not affected by each other.

Lee and Starner [Lee and Starner, 2010] tested the perception of three vibrotactile patterns presented on the wrist as a secondary task. They used non-overlapping spatiotemporal patterns containing three vibromotors activated in a sequence for a total duration of 1.5 seconds. Participants were not trained to associate patterns with the meaning but asked to build their own mental model instead. As a primary task, a visual search task with three levels of difficulty was used. The authors reported that the primary task and secondary task were not significantly affected by each other in terms of accuracy but they were affected in terms of reaction time. The study in this section uses the same primary task proposed and used by Lee and Starner [Lee and Starner, 2010]. As for the vibrotactile secondary task, first, the symbols representing letters of English alphabet are encoded and participants are trained to recognise the entire English alphabet prior to the study. Second, concise overlapping spatiotemporal patterns are used where each symbol is encoded with only 100-110 ms as aim is to maximise the throughput for information transmission. Both, the very short duration and the number of encoded symbols are expected to increase the difficulty of vibrotactile symbol identification which it is expected to be compensated by the pre-training where the recognition of tactons is formed and crystallised.

4.3.2 Procedure

This user study aims at evaluating how well users can perceive tactons in background while performing another attention demanding primary task. Prior to this study, participants were trained in 5 sessions to recognise all 26 letters of English Alphabet. So participants were recruited for this study upon finishing the Study 6 (see Section 4.2). Three days after the last training session (session 5), participants were invited to take part in study presented in this section. They were exposed to a recall session where they were tested for all 26 letters of English Alphabet. Additionally, they were tested how well they recognised letters of Alphabet in background while doing another primary task. For testing, while performing another task, only 10 letters encoded with one or two vibromotors were selected. Note that, participants were already trained to recognise all letters and the rationale to use only 10 letters for this test was simply to keep the study short and yet try different levels of difficulties in the primary task.

Pre Study Training

Prior to this study, participants were exposed to 5 sequent days (≈5h in total) of training and testing in letters and words. This training was done for the purpose of conducting the Study 6 (see Section 4.2) and participants were recruited for this study after completing the Study 6. During such training, participants learned the entire English Alphabet and were able to interpret words of 2-5 letters. For details of the training the reader is encouraged to review the Section 4.2. Similarly to the Study 6, the layout is hand based with seven vibromotors as illustrated in Figure 4.17. For simplicity, let us will refer to this training phase as pre-training throughout the section.

Patterns and Wearable Layout

Each symbol is encoded with one or two vibromotors using an OST (overlapped spatiotemporal) stimulation as introduced in Chapter 3. Symbol encoding uses a base duration (d) of 100 ms and a 10 ms gap (g) between the activation of vibromotors.

This means that the duration (ld) is 100 ms for one vibromotor symbols and 110 ms for two vibromotor encoded symbols.

A layout design with six vibromotors on the back of the hand is used identical to the Study 6 4.2(see Figure 4.17).

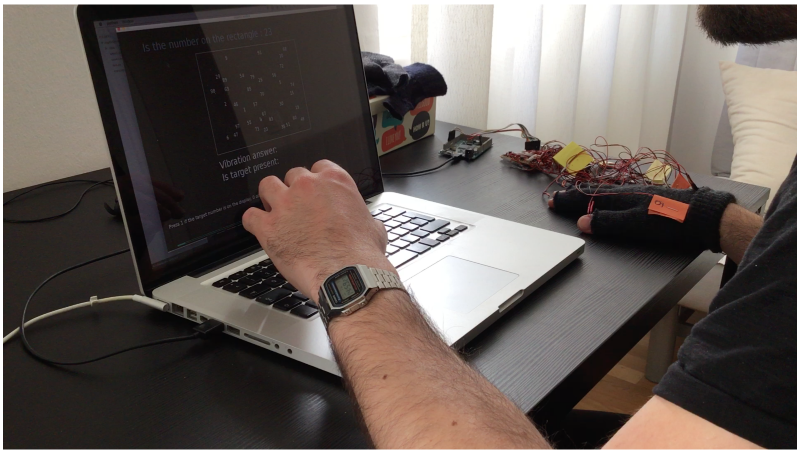

Figure 4.15: A participant performing the Study 7.

Participants

Seven participants (six males and one female) aged between 21 and 34 years old participated in this experiment.

Apparatus

The device consisted of an Arduino Due board which controls 3.4mm vibrotactile motors of type ROB-08449 (Voltage range: $2.3V \sim 3.6V$; Amplitude vibration: $0.8G$).

Study

Participants were equipped with the device. Initially, they were exposed to a round of vibrotactile only (**VBO**) testing with all 26 letters to measure the recall accuracy. Here, participants were stimulated with a vibrotactile pattern corresponding to a letter and asked to provide the answer, and they were not performing any other task during this procedure.

Figure 4.16: Visual search task user interface used in the Study 7. First, the search target was presented on the screen (left). Then the visual search screen was presented for 5 seconds (bottom). While this screen was active, participants were stimulated with a vibrotactile cue in 50% of the tasks. In the response screen (right), participants could still provide/change their answer.

After that, they continued to the multitasking (primary and background task) study where first they were exposed to a trial phase and then finally continued with the tasks. During the study, participants were asked to solve a visual search as a primary task. During this task, participants initially were presented with an integer representing a **search target**. Then a new screen was presented to them with a set of integers within a box which will referred to as **search set**. Their task was to determine whether the search target was within the search set. The screen with search set was visible only for 5 seconds during which participants could provide the answer. Additionally, in 50% of the cases, 0.5 seconds after the screen with search set appeared a vibrotactile stimulus representing one of ten selected letters

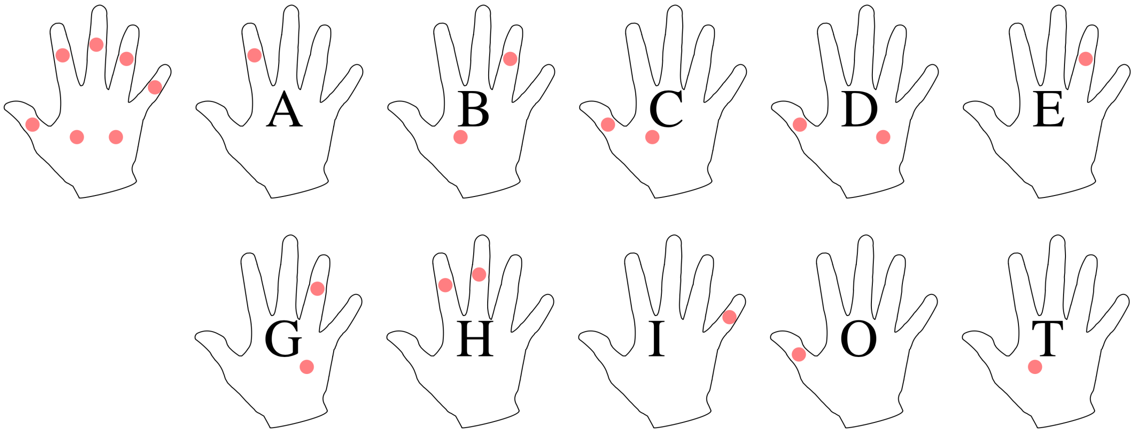

Figure 4.17: The hand based layout containing 7 vibromotors and the encoding of letters used for the study.

was presented to the participant and they needed to solve both visual search task and recognise the vibrotactile stimuli (**VSVB**) as a secondary/background task. In the other 50%, only visual search task was required (no vibrotactile stimuli) to solve (**VSO**). Participants were not informed priory whether the task contains a vibrotactile stimuli or not. Participants had the chance to repeat the vibrotactile stimuli by pressing the SPACE bar as long as the search set screen was visible and they could also provide the answer during this time. After that, the visual set screen was replaced with a new one, where participants had the chance to provide or change the answer, but they were not able to neither see the visual search set nor repeat vibrotactile stimuli. To provide the answer for visual search task, participants used keyboard numbers 0 (no) and 1 (yes), whereas to provide the answer for vibrotactile stimuli, they typed the letter representing the stimuli.

There were three difficulties of visual search task which was determined by the number of integers (size) within the search set. 9, 25 and 36 integers are used as such numbers have been used in the past [Lee and Starner, 2010] and suggested as three appropriate levels of difficulties. The position of the integers was randomly assigned. Additionally, the integers within the search set were unique and randomly selected from 1 to 99.

In the visual search and background vibrotactile tasks, each participant was exposed to 10 letters x 3 difficulties x 2 task type (VSO and VSVB) = 60 probes during the study. For seven participants, 420 probes in total are collected, 210 for each task type (70 for each difficulty). During the trial phase, they were exposed to 2 letters x 3 difficulties x 2 stimulation = 12 probes, but the responses were not recorded. In 50% of trials (balanced by difficulty and stimulation), the visual target was present in the visual search set. The entire probes appeared in random order. In the vibrotactile only test (VBO) 26 probes (each letter once) are collected for each participant. However, only the collected data only for the ten letters will be analysed as that are used in the visual search and background vibrotactile tasks. Thus for 7 participants, 70 probes (7 participants x 10 letters) will be analysed. The entire session took around 20 minutes (3-5 minutes vibrotactile only test, 3-5 minutes trail mode and 10-14 minutes the main study), although it varied as it depended on how fast participants responded. At the end, users filled a NASA TLX questionnaire where they were asked to self-assess the tasks in three different categories:

1. Vibrotactile only (**VBO**) - where there was no visual search task,
2. Visual only (**VSO**) - where there were no vibrotactile stimuli during the visual search task and
3. Vibrotactile and visual (**VSVB**).

Note that, this study uses only 10 letters to keep the session short and avoid fatigue. Given that, when the entire Alphabet (26 letters) is encoded, every letter is encoded by one or two vibromotors, 5 random letters with one motor and 5 random with two are selected for the study. Moreover, participants were not informed that only 10 letters would be used (or which ones). Thus, they were prepared to respond for 26 letters of the Alphabet during the VSVB tasks.

4.3.3 Results

Performance

The recall test had dual purposes. But for the scope of this section, let us include in the results only the ten letters that are used in primary/secondary task. Thus, the responses of the rest of 16 letters will be ignored. Furthermore, let us define the following independent variables for the analysis of the data:

- Task type which takes values: VBO (Vibrotactile only), Visual only (VSO) and Visual and vibrotactile (VSVB).

- Difficulty which is determined by the size of search set and takes values: 9, 25, 36 (for VSO and VSVB). Occasionally the value 0 (in visualisations) will be used and it will indicate that there was no visual search task and thus representing VBO.

Additionally, let us define the depended variables to be:

- Vibrotactile accuracy which represents the recognition accuracy letters represented by a vibrotactile stimulus.

- Visual accuracy which represents the accuracy on finding whether the search target was present in the search set during the visual search tasks.

In this examination, first, let us compare how the primary task affects the recognition of vibrotactile encoded symbols. For this, let us compare the performance of the users on recognising letters in vibrotactile only (VBO) rounds with the recognition of letters as a secondary task while they had to perform the visual search task (VSVB) as a primary task. The results are visualised in Figure 4.18 which show that recognition accuracy is quite the same regardless of whether users were performing another primary task and its difficulty. The recognition accuracy is a binary variable set to be 1 if the participant recognised the letter or 0 otherwise. Given that we are also dealing with repeated measurements, let us use McNemar's test in order to test for statistical significance between groups.

Overall, participants were able to recognise letters with a comparable and very high accuracy in both VBO ($\mu = 0.97, \sigma = 0.17$)[6] and VSVB ($\mu = 0.98, \sigma = 0.14$) conditions. When comparing particular difficulty levels of visual search task in VSVB with VBO, according to McNemar's tests, the differences in letter recognition are not significant in neither of the levels:

1. Search size set of 9: **VBO** ($\mu = 0.97, \sigma = 0.17$) vs **VSVB** ($\mu = 0.99, \sigma = 0.12$); $\chi^2(1, N = 140) = 0.0, p = 1.0$

2. Search size set of 25: **VBO** ($\mu = 0.97, \sigma = 0.17$) vs **VSVB** ($\mu = 0.97, \sigma = 0.17$); $\chi^2(1, N = 140) = 0.25, p = 0.62$

3. Search size set of 36: **VBO** ($\mu = 0.97, \sigma = 0.17$) vs **VSVB** ($\mu = 0.97, \sigma = 0.12$); $\chi^2(1, N = 140) = 0.0, p = 1.0$.

[6]The mean of a group is denoted by μ wheras standard deviation by σ.

Let us also examine the repetition of vibrotactile stimuli as this could be an indication of the difficulty of perceiving the vibrotactile message. Depending on whether there was a primary visual search task and the level of difficulty (search set size), between 4% and 6% of letters were repeated. McNemar's tests reveal that the differences are insignificant between any of the groups. The groups were compared similarly as in the case of the letter recognition accuracy (see above).

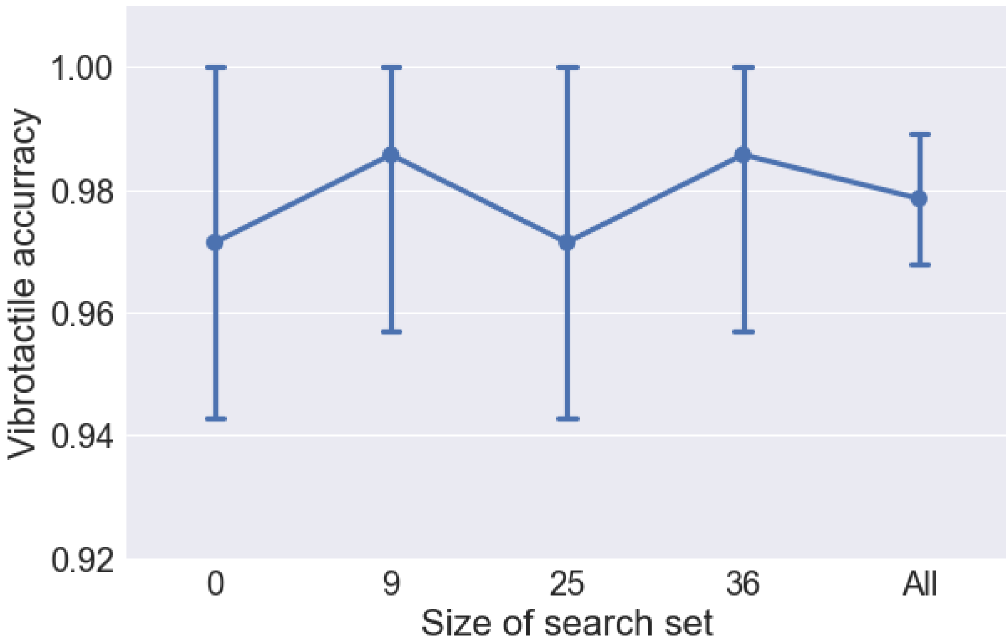

Figure 4.18: Vibrotactile letter recognition accuracy. The data is categorised depending on search set size and for all tasks combined. The set size of 0 indicates the VBO condition where no visual search task was present. Additionally, 'All' represents all set sizes combined (9, 25 and 36).

Moreover, let us analyse how the presentation of vibrotactile stimuli as a background/secondary task did affect the primary visual search task. Thus, let us compare the visual search task accuracy when participants were not stimulated with the vibrotactile message (VSO) with the tasks where they were required to solve both visuals search task and recognise the message/letter (VSVB). Overall, participants were able to achieve slightly better performance on solving visual search task when they did not have any other secondary task (VSO) ($\mu = 0.93, \sigma = 0.25$) compared to when they did have a vibrotactile secondary task (VSVB) ($\mu = 0.9, \sigma = 0.31$).

To determine the statistical significance, the differences among each level of

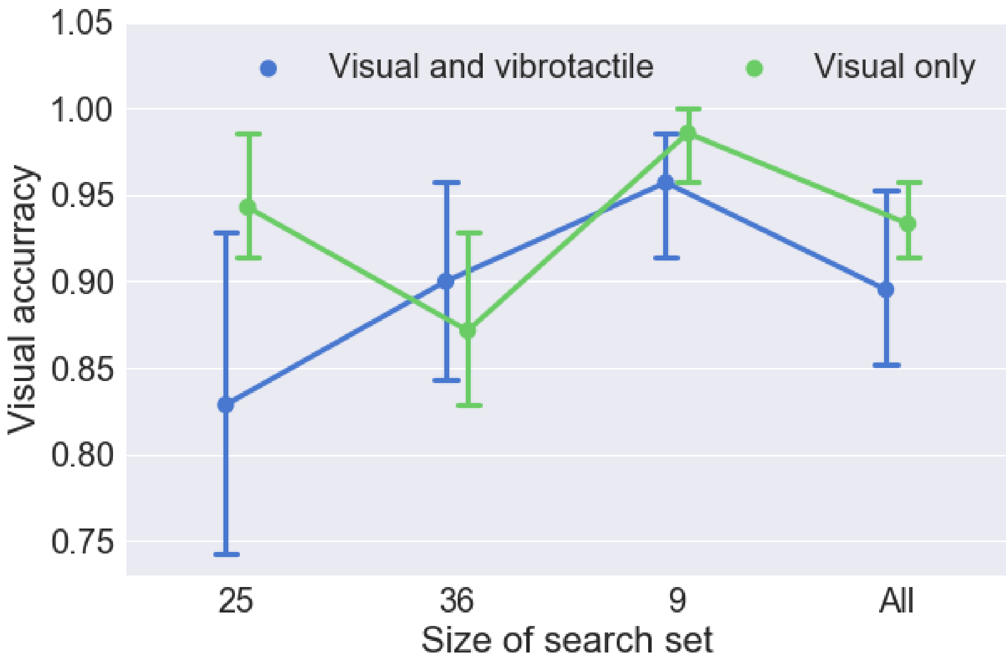

Figure 4.19: Visual search task accuracy depending on whether there was a vibrotactile task. The data is categorised depending on search set size and for all sizes combined ('All').

difficulty will be analysed. When looking at particular levels of difficulty, for the set size of 9 and 36 there are only slight differences (in both directions) whereas for the search size of 25 participants were able to solve visual search task better when they did not have a secondary vibrotactile task. Nevertheless, performing McNemar's tests between conditions (VSO vs VSVB) reveal that regardless of the level of difficulty, there are no statistically significant differences in the performance of solving visual search task between VSO and VSVB:

1. Search size set of 9: **VSO** ($\mu = 0.99, \sigma = 0.12$) vs **VSVB** ($\mu = 0.96, \sigma = 0.2$); $\chi^2(1, N = 140) = 0.25, p = 0.62$

2. Search size set of 25: **VSO** ($\mu = 0.94, \sigma = 0.23$) vs **VSVB** ($\mu = 0.83, \sigma = 0.38$); $\chi^2(1, N = 140) = 3.5, p = 0.061$

3. Search size set of 36: **VSO** ($\mu = 0.87, \sigma = 0.34$) vs **VSVB** ($\mu = 0.9, \sigma = 0.3$); $\chi^2(1, N = 140) = 0.06, p = 0.80$

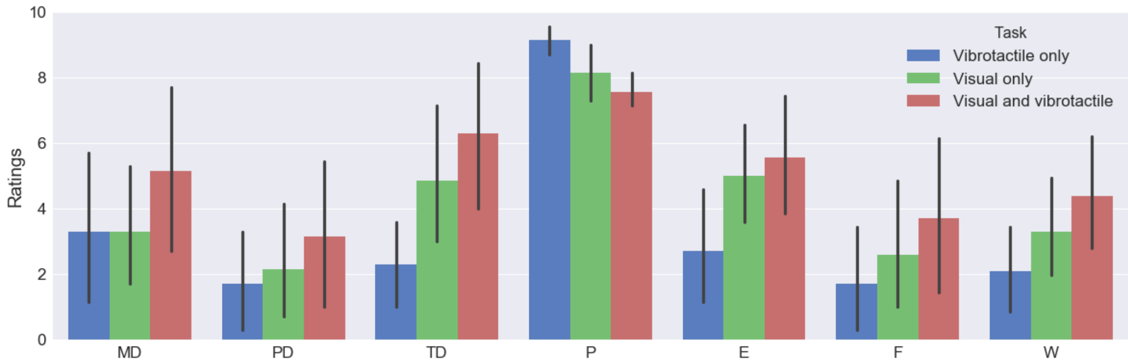

Figure 4.20: NASA TLX self evaluation metrics for the given tasks. Metrics notation : **MD** - mental demand, **PD** - physical demand, **TD** - temporal demand, **P** - performance, **E** - effort, **F** - frustration and **W** - workload.

4.3.4 Questionnaire

The results of NASA TLX are presented in Figure 4.20. In addition to the six metrics contained in NASA TLX, the workload is calculated using the simplified R-TLX method (averaging all metrics where performance is inverted). Let us compare the workload between three different task type conditions, namely VBO vs VSVB and VSO vs VSVB.

Considering that the workload of each task is normally distributed (Shapiro-Wilk: $p > 0.05$) and the variances of each compared groups are homogenous (Levene: $p > 0.05$) for each compared pairs, a paired t-test is used for determining whether there is a significant difference in workload. A paired t-test analysis reveals that the workload for VBO task ($\mu = 2.1, \sigma = 1.94$) was significantly lower compared to the workload of VSVB task ($\mu = 4.38, \sigma = 2.65$); $t(14) = -5.69, p = 0.001$. Additionally, the workload for VSO task ($\mu = 3.29, \sigma = 2.26$) was lower compared to the workload of VSVB task ($\mu = 4.38, \sigma = 2.65$); but not enough to be considered significant; $t(14) = -1.65, p = 0.15$.

4.3.5 Discussion

This user study was designed to investigate the perception and comprehension of tactons (representing letters of English Alphabet) in the background (as secondary task) while performing another attention demanding primary task. For encoding, the overlapping spatiotemporal patterns (see Chapter 3) are used. Such patterns use a very high speed of transmission where vibration patterns for a symbol are 100-110ms. The participants of this study were previously trained and proficient in recognising such tactons. Hereby, the hypothesis is that the recognition of such tactons would be processed using cognitive automatic processing and thus performed effortlessly and without interfering with the primary task. For the primary task, attention demanding visual search task with three difficulties is used, which is considered appropriate for this type of task.

Indeed, the results show that to be the case. First, the recognition of tactons was very accurate ($\geq 97\%$) regardless of whether this was done as a single task or as a secondary task along the visual search task (See Figure 4.18). Similarly, visual search task did not have any effect on the repetition of vibrotactile stimuli. Second, the performance of primary visual search task does not seem to be significantly deteriorated by the presence and recognition of tactons in the background. Although, on average the performance did slightly decrease. As shown in Figure 4.19, clearly such deterioration seem not to be accelerated with the increase of primary task difficulty. For instance, when the search set size was 36, which represents the highest difficulty, participants even performed better on the visual search task when having to simultaneously recognise a tacton in the background compared to the cases when they did not perform any task in the background. Additionally, the self-assessed workload (see Figure 4.20) did also not significantly increased when users performed the visual search task when the parallel background vibrotactile task was present.

The results show that transmitting information through vibrotactile wearable devices, even in very high speed is very efficient and can be used along with other user cognitive activities. As such, it provides many opportunities to support and facilitate multitasking especially considering that wearable devices are used. It suggests that, information conveyed using vibrotactile wearable devices can be effortlessly comprehended while performing other tasks as well, assuming that users are trained properly to recognise and associate the vibrotactile patterns. The used tactons represent letters of English alphabet. However, such tactons could represent any other abstract meaning such as commands, warnings, states etc.. as long as users are trained to recognise them. Additionally, only 10 tactons are used in order to keep the study short and avoid fatigue. However, participants were already trained in a pre-training period (as part of another study) to recognise the entire Alphabet (26 letters). During this user study, participants were not aware that they will be tested for only 10 symbols/letters and which of them will be used. Thus, I argue that they would have had the same cognitive load if they were tested on all 26 symbols/letters in multitasking experiment.

Although 10 or even 26 tactons might be limiting for all use cases, yet they can encode sufficient information in a lot of use cases (commands, warnings, states, etc..). On the other hand, when perceiving vibrotactile encoded information, it has already been demonstrated that individuals are able to understand words as a series of letters (as demonstrated in Sections 4.1 and 4.2) or phonemes [Zhao et al., 2018] as a primary task (without any other parallel task). It would be certainly very useful if users would be able to perceive such complex messages (words, sentences) composed of several tactons in the background as the application possibilities for multitasking would broaden drastically.

The insights gathered in this work will serve as foundation to further investigate the comprehension of textual information in background transmitted by wearable vibrotactile displays. Participants were not tested on comprehending words in background information as in the pre-training they were not trained on words but only on letters. Exposing them to words for longer periods would be necessary in order for them to be able to read words as units and thus get enough proficiency to be able to comprehend the words using automatic processing cognitive model. In other forms of reading such as visual, it is a well-established theory, that fast reading is attributed to words being read as units instead of letter by letter [Millar, 2004, Larson, 2004]. Such a word recognition as a unit is achieved through exposure to words (practice) [Whalen, 1991, Larson, 2004]. Analogously, for vibrotactile encoded information, I think that for testing in the background, users should be at a stage of proficiency where they would interpret words as units. Therefore, I plan to conduct such longitudinal user studies in the future work but currently it is out of the scope of this thesis.

4.4 Study 8: Passive Haptic Learning For Skin Reading

This section investigates the effects of using passive haptic learning to train the skill of reading text from vibrotactile patterns. The vibrotactile method of transmitting messages, skin-reading, is effective at conveying rich information but its active training method is quite demanding. Typically, learning to associate meanings (e.g letters or words) with such vibrotactile patterns involves active training, where users receive vibrotactile patterns accompanied by visual and audio cues representing the meaning [Luzhnica et al., 2016b]. Despite being effective, such a training requires full attention, is repetitive, extensive and tedious.

On the other hand, *passive haptic learning* [Seim et al., 2014a, Seim et al., 2015a] (*PHL*) can be used to train users passively without requiring their attention. This haptics-based teaching technique has been successfully used in numerous applications such as teaching people to play piano [Seim et al., 2015a] or type in braille [Seim et al., 2014a] without them being actively focused on training. During the training, they are exposed to audio and haptic stimuli that inform a skill, but they need not pay attention to it. Such a technique would be beneficial for training skin-reading as it might motivate potential users of skin reading that are interested but do not have the inclination to go through hours of active training.

All prior work on PHL for text system (Braille, Morse code, Stenography) learning taught the system incrementally; teaching letters in small groups based on words from a pangram [Seim et al., 2014a, Seim et al., 2015a]. It was assumed that the small groups and their semantic associations were necessary for learning of many letters; however, no prior work has contrasted this with a training method not requiring semantic grouping. On the other hand, a passive instruction method without having to develop word-based lessons may allow different learning durations, less rigid passive learning structures and less system development. Such non word-based training method without any semantic grouping has been successfully applied in active training for skin reading. Thus, this study contrasts these learning structures for PHL.

Additionally, As training is an extensive task, it would be useful if users could be trained with a default transmission speed but be able to understand messages with different transmission speed. This way users could increase the speed over time and they would not need re-training if the speed need to be changed. Perhaps PHL could enable this.

This section presents a user study investigating the following research questions:

- **RQ1:** *Can passive haptic learning can be used to train users for skin reading*?
- **RQ2:** *How does the duration of training stimuli affect recognition? Can users understand transmissions at a different speeds than the one used for training*?
- **RQ3:** *Is it necessary to semantically group letters when using PHL as a training method for vibrotactile skin reading? Or would a non word-based training be sufficient*?

4.4.1 Passive Haptic Leraning

Passive haptic learning (PHL) began with simple music sequence training for one hand and has since been explored for multi-limb skills, simultaneous actions, rhythm, other areas of the body and alphabetic codes for text entry [Seim et al., 2014a, Seim et al., 2015a, Seim et al., 2016, Seim et al., 2017]. The technique has been found in a limited number of cases and would benefit from further study. This work aims to replicate the technique of PHL and examine it for training users in vibrotactile skin reading. Furthermore, prior work [Seim et al., 2015a] contrasted two teaching structures for passive learning, but this work focused on teaching two-limb skills, and it has not been established whether a semantic chunking structure is beneficial to learning.

4.4.2 Procedure

The goals of this study were to investigate if PHL is useful to train for skin reading (RQ1), establish the effects of training stimuli speed on recognition results (RQ2), and compare a bottom up, letter by letter training (ABT) with a training based on words cues (WBT, RQ3). This study tests reception and knowledge before and after passive training and testing of letters and words.. PHL requires the attention of the user on a primary task while the training takes place in the background. The study intends to maintain the time and attention of participants within manageable margins so, the study uses only ten letters, enough to compose words, while limiting training time to around 30m. Note that, the method used to encode information is not limited to ten letters. In previous studies, it was used to encode the entire English alphabet, which participants learnt within three hours of active training (see sections 4.1 and 4.2). This study uses letters: 'A', 'C', 'E', 'G', 'H', 'I', 'M', 'N', 'S' and 'T', encoded with max. two vibromotors (see Figure 4.21). Given the native German language of the location where user study took place, German was used throughout the study for words and spelling. This study uses two training protocols (RQ3: ABT, WBT), three stimulation speeds during testing (RQ2: 100, 200, 300 ms) and measures of accuracy, repetition of stimuli and testing duration (RQ1).

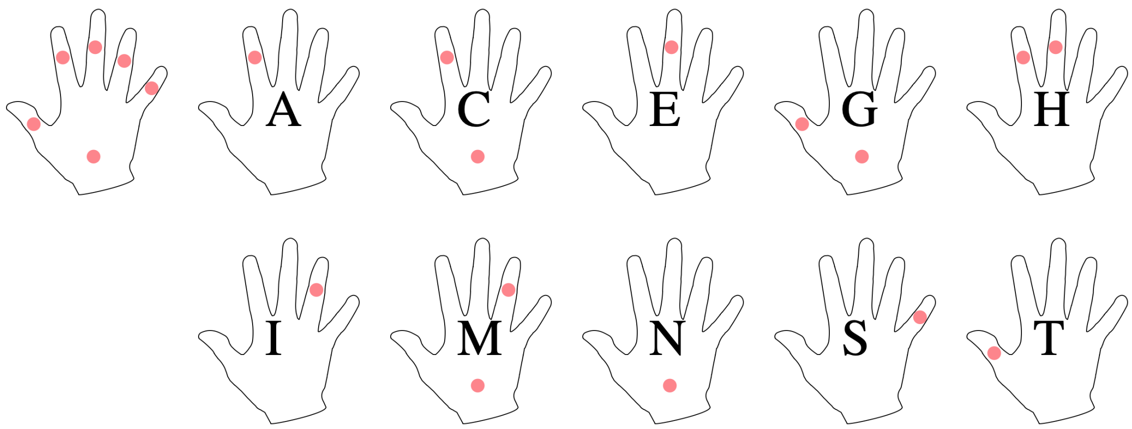

Figure 4.21: The wearable vibrotactile display layout and the encoding scheme of each letter used during the study 8.

Wearable Haptic Display Design

A layout design with six vibromotors on the back of the hand is used identical to the Study 5 presnted in Section 4.1 (see Figure 4.21). With it, the ten letters in the study can be encoded with combinations of one or two vibromotors. The rationale behind using only of six vibromotors is that only ten letters will be encoded for this user study. But, for encoding the entire alphabet, a layout with more vibromotors as proposed in Sections 4.2 and 3.4 would be a better choice. The vibromotors can be fitted in a fingerless glove, leaving the fingers free for interaction.

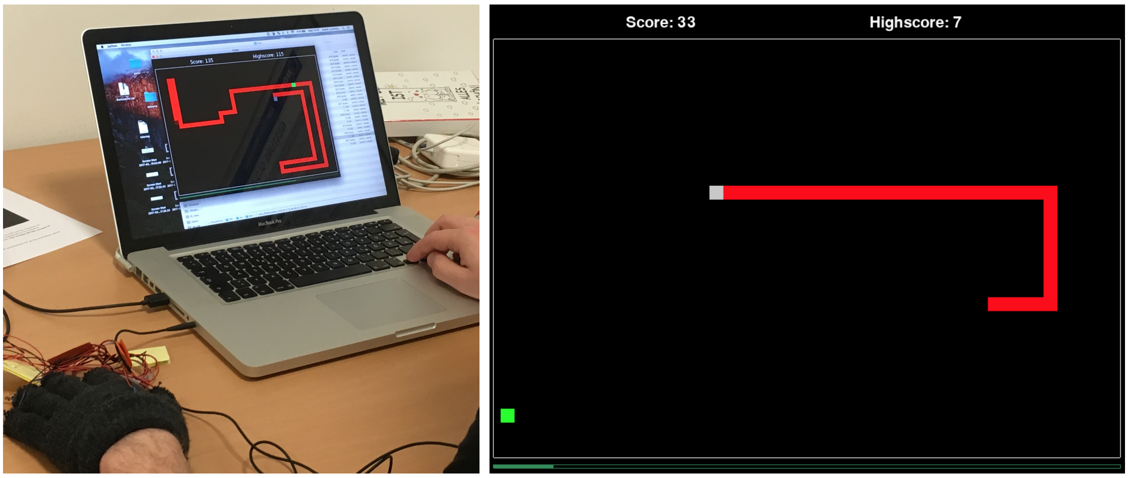

Figure 4.22: A participant (left) playing the game (right) while being trained to recognise letters using PHL.

Vibrotactile Patterns and Encoding

Each letter is encoded with one or two vibromotors using an OST (overlapped spatiotemporal) stimulation pattern described in Chapter 3. Moreover, the order of activation is prioritised by the sensitivity of the finger, since it yields a higher accuracy in identification of locus as revealed by Study 3 3.3. Sensitivity order is assumed according to studies suggesting that sensitivity decreases from the index finger towards the little finger: the index finger is more sensitive than the middle, ring, and pinky finger [Duncan and Boynton, 2007, Vega-Bermudez and Johnson, 2001, Hoggan et al., 2007]. The thumb is the lowest sensitive [Sterr et al., 2003]. For example, a letter encoded with index and pinky finger, activates the index vibromotor first, and then after a gap, the vibromotor on the pinky finger. Letter encoding uses a base duration (d) of 200 ms and a 10 ms gap (g) between the activation of vibromotors. So, the letter duration (ld) of a one vibromotor letter is 200 ms and 210 ms for two-vibromotor letters. When constructing words, a gap (bl) of 200 ms separates subsequent letters . Note that with longer training periods, users learn to recognise letters and words with shorter stimulation (see Sections 4.1 and 4.2). This study fixes training duration to 200ms and considers shorter durations during testing. The study aims at having training and testing in a single session and thus the decision for a longer duration.

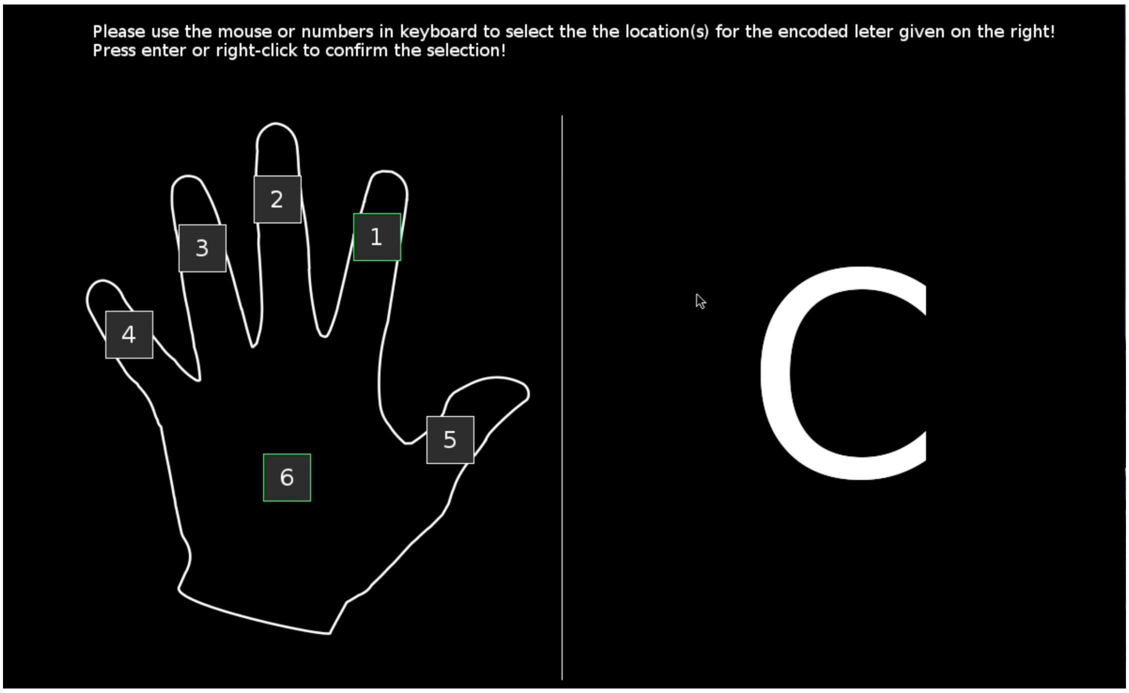

Figure 4.23: The user interface used for letter reconstruction test during Study 8.

Procedure

The entire study was organised in *rounds*, each serving the purpose of either training or testing. PHL takes place during training, where participants are engaged in playing a game as a primary task (internal implementation of the snake game[7]). Meanwhile, they are passively trained to recognise patterns (see Figure 4.22). Testing rounds use the active concentration of participant on the test. There are two training modes: WBT and ABT **Word Based Training (WBT)** uses word cues to passively train users to associate letters with vibrotactile patterns. WBT starts with an audio cue of a word (e.g. Ich) and continues with a series of audio cues of each letter of that word. the vibrotactile stimulation of the pattern representing the letter follows 50ms after its audio cue. The study used the words "ICH", "MAG", "ES", "NICHT", which together form a sentence ("Ich mag es nicht") from the children's book "Grunes Ei mit Schpeck" written by Dr Seuss. Each word is played in a loop 48 times before moving to the next one. WBT takes 32 minutes. Figure 4.24 illustrates the procedure.

[7] https://en.wikipedia.org/wiki/Snake_(video_game_genre)

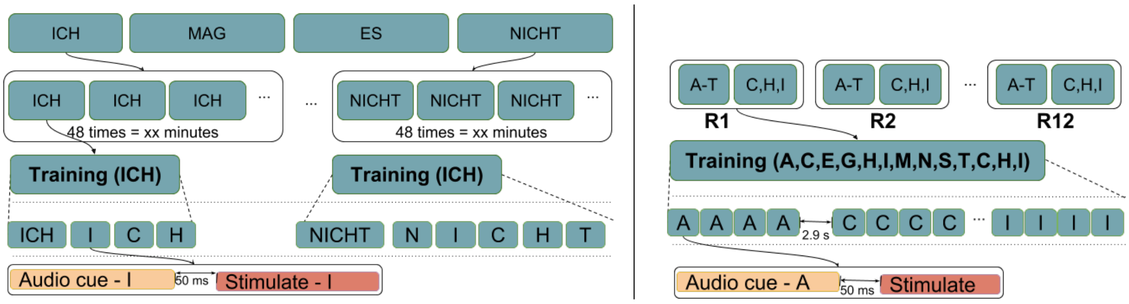

Figure 4.24: Training methods: word based training (left) where letters from a word are used to determine the order of the trained letters and sequential training (right) where letters are trained in alphabetical order.

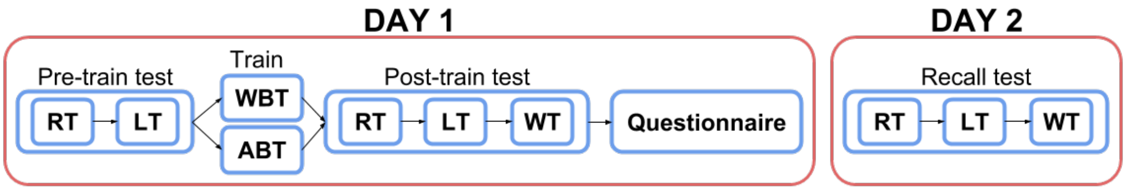

Figure 4.25: The entire procedure of PHL training and testing.

Alphabetical Based Training (ABT) uses letters in alphabetic order to passively train users to associate symbols with vibrotactile patterns. ABT starts with an audio cue which represents a letter of the German alphabet, followed by its vibrotactile cue after 50 ms. The process is repeated four times, moving to the next letter to compose one round. The entire procedure was repeated for 12 times (32 minutes) composing 12 rounds. In addition to the ten letters in alphabetical order, one round also contained the letters C, H and I at the end. Doing so, the number of letters stimulated in ABT is balanced with that of WBT where the letters C, H and I appear twice.

- **Reconstruction Testing (RT)**. Participants were asked to select (using the mouse or keyboard) which locations (vibromotors) are used to encode a given letter displayed on the screen. Figure 4.23 shows the user interface for RT.

- **Letter Testing (LT)**. Participants were stimulated with a pattern and asked to input the letter associated with it. They could repeat the stimuli before answering, and they were not notified whether their answer was correct.

- **Word Testing (WT)**. Like LT, users try to recognize stimulations. Participants were tested for words constructed only from letters that they trained. These include the four words used in WBT (ES, ICH, MAG, NICHT) and four additional words (IN, MIT, IST, GEHEN).

The first round of the study was a pre-test consisting of a round of RT and a round of LT. Pre-test served the purpose to familiarise participants with the testing procedure and to demonstrate their lack of knowledge about skin-reading. Thereafter, participants were exposed to the game and passive training. They were explicitly instructed to focus on the game. They were randomly assigned to two equal groups. The first group trained using WBT and the second one with ABT. After 32m of training, they were exposed to rounds of RT, LT and WT. Let us refer to this block as post-train testing. To study the effect of transmission speed, post train LT and WT were performed with a base duration of $d \in [100ms, 200ms, 300ms]$ coupled, in the case of words, with between letter duration of $bl \in [100ms, 200ms, 300ms]$. Finally, participants filled out a NASA TLX questionnaire, rating workload of the letter and word recognition. They were also asked to rate the three following sentences using five-level Likert scale (from strongly disagree to strongly agree):

- ***Effectiveness***: *The (voice) passive training of letters while playing is a good way of teaching to recognise vibrotactile encoded letters!*
- ***Annoyingness***: *The (voice) passive training of letters while playing the game was annoying!*
- ***Interruptedness***: *The (voice) passive training of letters while playing the game did prevent me from focusing on the game!*

On the very next day, participants were exposed to another testing, identical to the post-train testing. Let us refer to it as recall-testing, as its purpose was to evaluate how much users recall the next day. The entire procedure is depicted in Figure 4.25. In the pre-train testing one probe per letter was collected in both LT and RT. During post-train and recall, for each letter, one probe was collected in RT, six probes (two per each speed) in LT and three probes (one per speed) WT.

The audio cues used the German spelling of letters, and also german words are used during BWT given the native German language of the location where user study took place. The study did not use English as due to concerns that participants would get confused by the letters I and E. E in English spells the same as I in German, and thus as participants only hear the letters, they might form the wrong association.

Apparatus

The device consisted of an Arduino Due board which controls $3.4mm$ vibrotactile motors of type ROB-08449 (Voltage range: $2.3V \sim 3.6V$; Amplitude vibration: $0.8G$).

Participants

Twenty (20) individuals (13 male and 7 female) aged between 23 and 46 (M=32.7, STD=7.6) years old participated in this study. Half of participants used WBT. Only one of them was left handed. All of them used the left hand for stimulation and the right to interact with the computer as depicted in Figure 4.22.

4.4.3 Results

Let us define the following variables: accuracy, repetition and total duration. Repetition describes how many times a user repeated the stimulation (letter or word) in LT, WT rounds. Total duration represents the difference between the user response time-point and the first stimulation time-point including repetitions.

Accuracy will be defined differently for different test types. For the RT accuracy is defined to be 1 if the user provides the exact locations that encode the given letter, otherwise 0. Similarly, for LT the accuracy is a binary variable defined to be 1 if the user's response matches the stimulated letter. For WT, accuracy is defined in relation to the similarity of the stimulated word to the user's response. Word recognition accuracy for a pair of answer and stimulated word (a,s) is computed by the given expression:

$$\sigma(a,s) = 1 - \frac{d(a,s)}{\#s}, \tag{4.6}$$

where d is the Levenshtein distance [Levenshtein, 1966] between two words and $\#s$ represents the word length (number of letters). The Levenshtein distance is defined as the minimum single-letter edits (insertions, deletions or substitutions) required to change one word into the other [8].

[8] https://en.wikipedia.org/wiki/Levenshtein_distance

Test	Phase	Train	Speed	Accuracy	Duration	Repetition
Letter	Post	ABT	100ms	.74 (.44)	3.94 (3.81)	2.30 (3.26)
			200ms	.71 (.45)	4.26 (4.56)	1.91 (2.88)
			300ms	.74 (.44)	4.05 (3.66)	1.64 (2.43)
		WBT	100ms	.68 (.47)	3.95 (3.59)	1.10 (1.74)
			200ms	.64 (.48)	4.59 (5.57)	1.61 (3.74)
			300ms	.62 (.49)	4.22 (4.21)	1.24 (2.57)
	Recall	ABT	100ms	.70 (.46)	3.95 (3.56)	2.76 (2.98)
			200ms	.72 (.45)	3.30 (2.53)	2.44 (3.02)
			300ms	.72 (.45)	3.55 (3.09)	2.35 (2.72)
		WBT	100ms	.64 (.48)	3.74 (3.83)	1.46 (4.37)
			200ms	.68 (.47)	3.80 (4.15)	1.15 (2.29)
			300ms	.67 (.47)	4.44 (5.44)	1.52 (3.30)
Words	Post	ABT	100ms	.70 (.36)	6.52 (3.82)	6.35 (6.96)
			200ms	.69 (.34)	7.64 (5.89)	7.25 (10.3)
			300ms	.73 (.34)	6.17 (3.13)	4.94 (4.89)
		WBT	100ms	.64 (.41)	6.62 (3.94)	3.46 (3.01)
			200ms	.64 (.40)	6.40 (3.76)	3.65 (3.82)
			300ms	.71 (.36)	6.30 (3.32)	3.35 (3.78)
	Recall	ABT	100ms	.72 (.33)	6.05 (3.22)	5.22 (5.45)
			200ms	.74 (.33)	5.61 (3.25)	5.29 (5.35)
			300ms	.75 (.32)	5.88 (3.78)	4.29 (4.49)
		WBT	100ms	.69 (.40)	5.93 (3.40)	3.76 (3.48)
			200ms	.69 (.39)	5.53 (2.89)	3.34 (3.01)
			300ms	.76 (.37)	5.51 (2.20)	2.56 (2.88)

Table 4.8: Letter and word recognition results for the Study 8.

Let us define the testing phase (post-train, recall), speed (100 ms, 200 ms. 300 ms) and training method (ABT, WBT) as independent variables; the letter reconstruction accuracy, as well as recognition accuracy on word and letters as dependent variables. The repetition rate and total duration are considered to be dependent variables. Moreover, let us consider comparisons of performance (accuracy, duration, repetitions) in testing phases to answer RQ1, comparisons of performance as regards speed to answer RQ2, and effects of training method to answer RQ3.

Method	LT	RT
ABT	.04 (.20)	.07 (.26)
WBT	.09 (.29)	.09 (.29)
Both	.06 (.25)	.08 (.27)

Table 4.9: Pre-train letter recognition and reconstruction accuracy.

Phase	Train	Accuracy
Post	ABT	.70 (.46)
Post	WBT	.61 (.49)
Recall	ABT	.71 (.46)
Recall	WBT	.61 (.49)

Table 4.10: Post-train and recall letter reconstruction accuracy.

Letters

Table 4.9 lists letter reconstruction and recognition accuracies in the pre-train test. Participants managed to guess/identify letters with an accuracy of 6% and reconstruct them with an accuracy of 8% before the training. The results demonstrates their lack of knowledge about the encoding of letters. The letter recognition and reconstruction accuracies for the post-train and recall tests are presented in Table 4.8 and Table 4.10.

Considering that the recognition and reconstruction accuracy are binary values, a chi-squared analysis will be used to determine the significance of differences in accuracy.

As regards RQ1, a chi-squared analysis revealed no significant difference in recognition accuracy between the post-train phase ($M = 0.69, STD = 0.46$) and recall ($M = 0.69, STD = 0.46$); $\chi^2(1, N = 2400) = 0.0, p = 0.96$. There is also no significant difference in reconstruction accuracy between the post-train phase ($M = 0.66, STD = 0.48$) and recall ($M = 0.66, STD = 0.47$); $\chi^2(1, N = 400) = 0.0, p = 1.0$. Furthermore, a chi-squared analysis reveals that the differences in accuracy between the recognition ($M = 0.69, STD = 0.46$) and the reconstruction of letters ($M = 0.66, STD = 0.48$) are not significant ; $\chi^2(1, N = 2800) = 1.21, p = 0.27$.

Let us also explore the relationship between the letter recognition accuracy and the performance in the game while training which is presented in Figure 4.26. A Pearson correlation analysis reveals that there is no significant correlation between

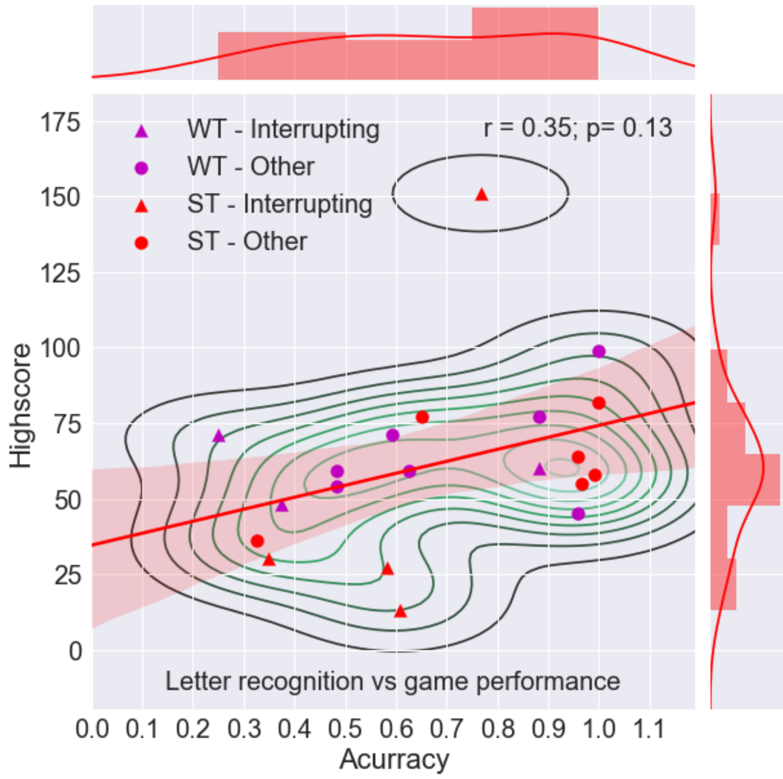

Figure 4.26: The relation between the average (per user) letter recognition accuracy and the user's high score. The bar plots on the top and on the side represent histograms and calculated the univariate distribution of the variable in the given axis. The contours represent the multivariate distribution of both variables. The straight line and the shades around it represent the fitted regression and its confidence. The Pearson correlation index and the confidence value are annotated as r and p. The colour represents the training method whereas the shape expresses the participant's rating on the training experience.

the average recognition accuracy and user's high-score in the game; $r = 0.35, p = 0.13$. Moreover, Figure 4.26 clearly shows that the recognition accuracy varies a lot among users. Four users do not even achieve 40% accuracy. On the other hand, there is a cluster of eight users that perform with accuracy over 88% and the rest lie in between.

Additionally, let us explore the total duration of users' response. Since the values are neither binary nor normally distributed (Shapiro-Wilk test, p ¡ 0.05), nonparametric tests will be used for determining the significance. The effects of phase on duration until response are analysed with Wilcoxon signed-rank test, considering that repeated measurements are being handled. The test reveals that indeed participants were **significantly** faster on the recall test ($MDN = 2.54, M = 3.8, STD = 3.88$) compared to the post-train ($MDN = 2.78, M = 4.17, STD = 4.29$); $V = 323582.5, p = 0.004$. Also repetition rate is non binary and not normally distributed and thus the same nonparametric test will be used. The test reveals that participant performed **significantly** less repetitions on post-train phase ($MDN = 1.0, M = 1.63, STD = 2.86$) than in recall ($MDN = 1.0, M = 1.95, STD = 3.23$); $V = 146372.0, p = 0.002$.

Regarding RQ2, there is no significant effect of speed in letter recognition accuracy; $\chi^2(2, N = 2400) = 0.0, p = 0.99$. The Kruskal-Wallis test will be used to analyse the effect of transmission speed on duration until response. The test reveals that the duration is not significantly affected by the transmission speed; $H(2400) = 0.85, p = 0.654$. A Kruskal-Wallis test reveals that transmission speed had also no effect on repetition rate; $H(2400) = 2.44, p = 0.295$.

As regards RQ3, there is a **significant** difference in recognition accuracy between participants trained with ABT ($M = 0.72, STD = 0.45$) and those trained with WBT ($M = 0.65, STD = 0.48$); $\chi^2(1, N = 2400) = 12.09, p < 0.001$. There is also a large, albeit non-significant difference in reconstruction accuracy between ABT ($M = 0.70, STD = 0.46$) and WBT ($M = 0.61, STD = 0.49$); $\chi^2(1, N = 400) = 3.6, p = 0.058$. To analyse how training method affects the response time and repetitions Wilcoxon rank-sum test will be used. The test reveals that the differences between WBT ($MDN = 2.72, M = 4.12, STD = 4.53$) and ABT ($MDN = 2.56, M = 3.84, STD = 3.6$) are not significant; $W = 1.39, p = 0.164$. Interestingly, participants trained using WBT did **significantly** fewer repetitions ($MDN = 0.0, M = 1.35, STD = 3.13$) than those trained with ABT ($MDN = 1.0, M = 2.23, STD = 2.91$); $W = -12.46, p = 0.0$.

Words

The average word recognition accuracy, duration and repetition rate are presented in the Table 4.8. Given that the word recognition accuracy is a real value and not normally distributed (Shapiro-Wilk test, p ¡ 0.05), nonparametric tests (Kruskal-Wallis, Wilcoxon rank-sum and Wilcoxon signed-rank) will be used for determining the significance.

As regards RQ1, a Wilcoxon signed-rank test reveals that indeed participants perform **significantly** better (accuracy) on the recall test ($MDN = 1.0, M = 0.72, STD = 0.36$) compared to the post-train ($MDN = 1.0, M = 0.68, STD = 0.37$); $V = 12990.5, p = 0.023$. Duration is also not normally distributed (Shapiro-Wilk, p ¡ 0.05). Comparing duration until response, a Wilcoxon signed-rank test reveals that participants were **significantly** faster on the recall test ($MDN = 5.05, M = 5.75, STD = 3.15$) compared to the post-train ($MDN = 5.39, M = 6.61, STD = 4.08$); $V = 45664.0, p = 0.0$. Participants also performed **significantly** more repetitions on post-train phase ($MDN = 3.0, M = 4.83, STD = 6.18$) than in recall ($MDN = 3.0, M = 4.08, STD = 4.33$); $V = 38467.0, p = 0.011$.

Regarding RQ2, A Kruskal-Wallis test reveals that the word recognition accuracy is not significantly affected by the transmission speed, $H(2) = 3.78, p = 0.15$. The duration is also not significantly affected by the transmission speed; $H(960) = 1.36, p = 0.507$. But, concerning repetition rate, the test reveals that the transmission speed had a **significant** effect on repetition rate; $H(960) = 14.09, p = 0.001$. A further post-hoc Wilcoxon signed-rank tests reveal that participants did **significantly** fewer repetition when the vibrations and the gap between letters was set to 300 ms ($MDN = 2.0, M = 3.78, STD = 4.16$) compared to 200 ms ($MDN = 3.0, M = 4.88, STD = 6.48$); $V = 13594.5, p = 0.0$ and 100 ms ($MDN = 3.0, M = 4.7, STD = 5.1$); $V = 13944.0, p = 0.0$. However the differences between 200 ms and 100 ms were not significant; $V = 18676.0, p = 0.592$.

As regards RQ3, a Wilcoxon rank-sum test reveals that the differences in accuracy between WBT ($MDN = 1.0, M = 0.69, STD = 0.39$) and ABT ($MDN = 1.0, M = 0.72, STD = 0.34$) are not significant; $W = -0.63, p = 0.531$. Similarly, the differences in duration between WBT ($MDN = 5.26, M = 6.05, STD = 3.31$) and ABT ($MDN = 5.16, M = 6.31, STD = 4.0$) are not significant; $W = 0.47, p = 0.638$. However, participants that were trained using WBT did **significantly** fewer repetitions ($MDN = 2.0, M = 3.35, STD = 3.36$) than the ones that used ABT ($MDN = 4.0, M = 5.56, STD = 6.6$); $W = -6.29, p = 0.0$.

Questionnaire

The users rating on the how effective the game based PHL is, how much it interrupts the game and whether it is annoying during the game, are visualised in Figure 4.27. The overall ratings are quite positive. However, there are a couple of participants that did provide some poor ratings. While the majority of the users thought that

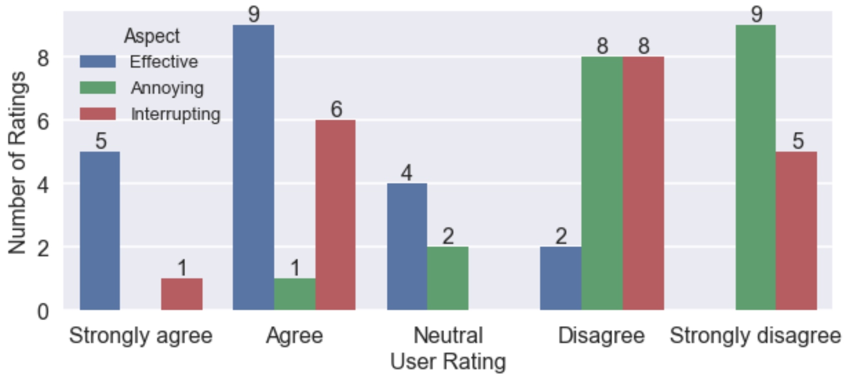

Figure 4.27: User ratings on how effective, interrupting and annoying the PHL is while playing the game.

it is effective, two users disagreed, and four others were neutral. On the matter of being annoying, one user did find it annoying, and two others were neutral on this. Additionally, seven users found it interrupting as they thought that the PHL did prevent them from focusing on the game.

The results of NASA TLX for letter and word recognition tasks depending on the training method are depicted in Figure 4.28. In addition to the six metrics contained in NASA TLX, the workload is calculated using the simplified R-TLX method (averaging all metrics where performance is inverted). So, let us compare the workload of letter and word recognition tasks between the training methods.

Given that the workload values are normally distributed (Shapiro-Wilk: p ¿ 0.05) and the variances of compared groups are homogenous (Levene: p ¿ 0.05), the independent t-test will be used. A t-test analysis reveals that the workload for letter recognition for participants that used ABT training method ($M = 4.22, STD = 1.47$) was lower that the workload of participants that trained used WBT ($M = 4.5, STD = 0.88$), but the differences are not significant; $t(20) = -0.52, p = 0.608$. When looking at the word recognition workload, on the contrary, participants that trained using ABT expressed a higher workload ($M = 5.68, STD = 1.4$) that participants that were trained using WBT ($M = 4.77, STD = 1.19$). However, again the differences are insignificant; $t(20) = 1.58, p = 0.13$.

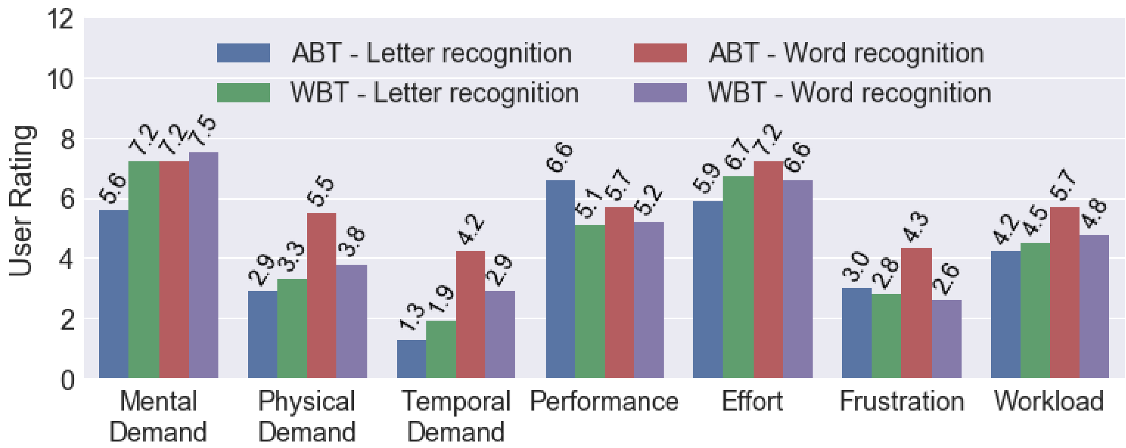

Figure 4.28: NASA TLX self evaluation metrics for letter and word and recognition tasks for the Study 8.

4.4.4 Discussion

Our user study was designed to investigate whether PHL could be used to train users for vibrotactile skin reading (RQ1) and explore different training methods (RQ3). Additionally, this study investigates whether the transmission speed compared to the one that was used to train participants affects their ability to perceive the encoded information (RQ2).

The results of this study show that overall, both phases, both training methods and all speeds combined, participants achieved a recognition accuracy of 69% on letters and 70% for words. To put this in perspective, Study 10 (see Section 6.2) reports that 20 participants achieved an accuracy of 95% on letter recognition after 5 iterations of active training rounds each followed by a reinforcement round (in total 10 rounds). The reported accuracy is higher with less training rounds compared to this study. Nevertheless, even though the learning rate is less than what participants could have potentially learned within the same time using active learning [Luzhnica and Veas, 2018a], recognising 69% of the letters using PHL means that in 32 minutes of training they were able to learn 7 letters in average. Thus in practice, for learning the entire Alphabet, one would need to reduce the number of letters within the 32 minutes of training and have more sessions (one per day) until users learn it entirely. Additionally, participants were able to not only recognise the letters but also reconstruct them which is consistent with the research on PHL of Morse code and Braille which showed successful reception along with reproduction [Seim et al., 2014a, Seim et al., 2016].

PHL has the benefit of letting users enjoy other activities while being trained and thus users would not need to stare at the screen and devote focus to training. Thus PHL could be used and presents an attractive alternative method for training vibrotactile skin-reading (RQ1). Whether, users would prefer spending more time in training but perform other activities during the same time (e.g. playing video games) or less time but focus actively on training, or even mix them, it would be up to individual preferences and should be considered as a trade-off. Moreover, six users did achieve an accuracy of or close to 100% (see Figure 4.26), meaning that for them, 32 minutes of PHL training was enough to learn 10 letters. Others demonstrated less learning from PHL. While this study was unable to determine the cause of their poor performance, such phenomena in the future could be explored in future work. Perhaps by tuning training parameters such as the time from the sound cue to vibrotactile stimulation, the volume, personalised training method (e.g. different number of letters for different users) etc... one could improve the learning effect for such users.

This study also investigates whether there is a trend that users who did well at learning also did poorly at the game, suggesting that they possibly attended to the stimuli actively. However, it found no such trend and those who did better at learning were also some of the ones who did best at the game.

The results of the word based (WBT) and alphabetically based training (ABT) methods (RQ3) demonstrated that both methods could be used for training. Nevertheless, findings on learning condition differences were surprising. Results suggest that the ABT condition enabled significantly better recognition and production of letters, and comparable performance on words; though this group required (significantly) more repetitions.

This would indicate that perhaps the ABT allowed comparable learning while also helping users think of letters individually rather than strongly tied to their word. The research team expected that learning from the disorganised ABT condition would be very challenging and that the cognitive benefit from semantic associations and small groups of letters in the WBT condition would allow those users to perform significantly better. Given the surprising results which suggest the promise in the ABT condition, this work clearly shows that further consideration of the ABT vs

WBT learning structure is needed and is relevant to all PHL work.

The analysis of the recognition in different phases (post-train and recall) show that participants were able to recall the learned information after one day. Retention and recall are well known for traditional learning methods; however, it is often asked, but still unknown, whether learning from PHL lasts. The breadth of passive tactile learning research - from piano to rehab - has yet to explore this important question [Seim et al., 2014a, Seim et al., 2015a, Seim et al., 2016]. Further research should investigate later recall tests in different PHL scenarios, but this initial result is encouraging that the effects of PHL are beyond short-term working memory. Moreover, the results show an improvement on the subsequent day.Perhaps performance improved after a night of sleep or a break as the literature suggests aids learning [Stickgold and Walker, 2013] and even motor learning [Walker et al., 2002].

Finally, the results show that participants could comprehend the transmitted information with the same accuracy regardless of the transmission speed (RQ3). This is consistent with related work which showed stimuli of different durations could be equally recognised on the fingers actively [Seim et al., 2015b], assuming that the minimum duration threshold has been considered. The results suggest that the duration of stimuli could be decoupled between training and the use of the device. Participants could train with one speed and once they learn the Alphabet and the use the device with faster speeds or even adjust the speed during usage without re-training.

4.5 Skin Reading for Sensory Substitution

Given the efficiency demonstrated in Section 4.2, vibrotactile skin reading using wearable devices presents an excellent opportunity for utilising in general purpose applications to facilitate multitasking or reduce demands on the predominant visual displays. For instance, users would be able to perceive their phone notifications, SMS, emails etc... while performing, e.g. driving, biking, working, etc... Skin reading could also find use in the application for users with specific impairments which is the focus of this section. This section proposes concepts and implementation of two mobile applications which capture the user's environment, describe it in the form of text and then convey its textual description to the user through a vibrotactile wearable display. The applications target users with hearing and vision impairments.

Note that the proposed applications are implemented to illustrate the capabilities of skin reading. However, they are not evaluated with the target users and such evaluation is out of the scope of this thesis.

4.5.1 Sensory Substitution

Sensory substitution has been a research subject for decades, and yet its applicability outside of the research is very limited. Thus creating scepticism among researchers that a full sensory substitution is not even possible [Spence, 2014]. Sensory substitution aims at re-channelling one sensory modality to another by changing the characteristics of the later. The attempts to re-channel vision [White et al., 1970] or speech [Gault, 1924] to tactile have been a long goal of the research community to enhance the lives of vision or auditory impaired individuals. Typically, the transformation vision to tactile stimuli is achieved by using a camera which captures the environment, transforms it to low-resolution image and then uses an array of actuators to *imprint* the image on the body [White et al., 1970]. Analogously, speech to tactile is converted by capturing the entire speech by a microphone, eventually extracting important signal features and then stimulating them to the skin [Gault, 1924, Novich, 2015]. While the methods of transforming the signal vary [Gault, 1924, Novich, 2015], they all tend to convey the entire richness and details of the captured environment to the skin.

Despite decades of research, yet there is no solution that can serve as complete hearing or seeing solution [Spence, 2014]. The reason for that might be connected with two fundamental problems to such an approach. First, the tactile sensation is not very high in resolution [Spence, 2014]. Second, there might be limiting cognitive constraints on processing the tactile information when attempting to convey the entire information contained within an image or sound [Spence, 2014]. On the other hand, recent works demonstrate that with few hours of training it is feasible to teach users to associate spatiotemporal patterns with symbols (e.g. letters of English Alphabet) and then combine such letters into more complex messages such as words and phrases (see Sections 4.1, 4.2, 4.4 and [Luzhnica et al., 2016b]. Thus, instead of trying to convey an entire image, one could express its essential aspects though text. Similarly, instead of letting users feel sound as the whole, one could provide only the text contained within the speech. Moreover, the active research and recent developments in machine learning applications have been very fruitful. Its applications such as speech recognition and object recognition have already been made available for consumer applications.

Now, of course, there are a considerable amount of details that will be lost when converting to text. An image is worth of 1000 words and yet in most of the cases, its essential representation can be described within a sentence. Similarly, the text will not express the rich emotions, sarcasm and other important aspects contained within the speech. However, the removal of such details lowers the bandwidth constraints and makes it possible to comprehend it as research shows that comprehension of text is possible. Thus, it has the potential to improve the life quality for millions of visually or auditory impaired individuals.

4.5.2 Conecept

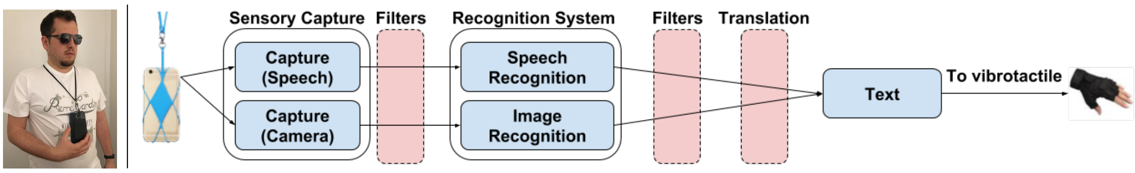

Figure 4.29: Application concept: (i) the environment is captured through camera or microphone, (ii) the content is transformed to text through image or speech recognition, (iii) alternatively the textual information is translated to the target language and finally (iv) transmitted to the wearable vibrotactile device.

This section proposes a wearable and mobile solution which transforms the captured environment to text and then conveys the text through a wearable vibrotactile display. The transformation is done in the following step:

1. Capture the environment using a camera (the area in front of the user) or microphone (the person talking to them).

2. Perform recognition of either speech or image and convert the captured signal to textual context.

3. If necessary, translate the text to a predefined language.

4. Transmit the text to wearable vibrotactile display.

The concept uses the same wearable display methods of conveying information as elaborated in Section 4.2. Note that as shown in Figure 4.29, to help the target users as much as possible, it would be useful some filtering of the information which could be done along the process. For instance, when capturing the speech of the user, it would be useful to identify the speech that originates from nearby sources and ignore the ones that come from far away. Similarly, when performing image recognition, there are many details to the image and the user might be interested only in some particular aspects of it. Such target user preferences and requirements should be considered when performing filtering of the information.

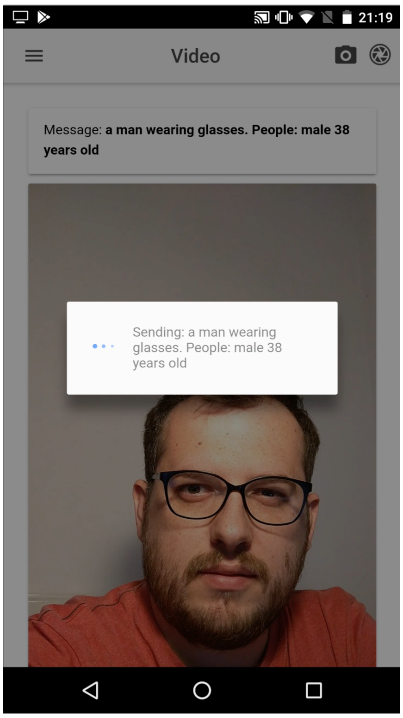

Figure 4.30: VTT recognising the objects and sending the recognised text to the wearable vibrotactile display.

4.5.3 Mobile Application Implementation

The concept was implemented in two mobile applications, each targeting a different user group: speech to tactile (**STT**) which targets users with hearing impairments and vision-to-tactile (**VST**) which targets the users with visual impairments. Both of them use the same concept as illustrated in Figure 4.29. The interaction concept

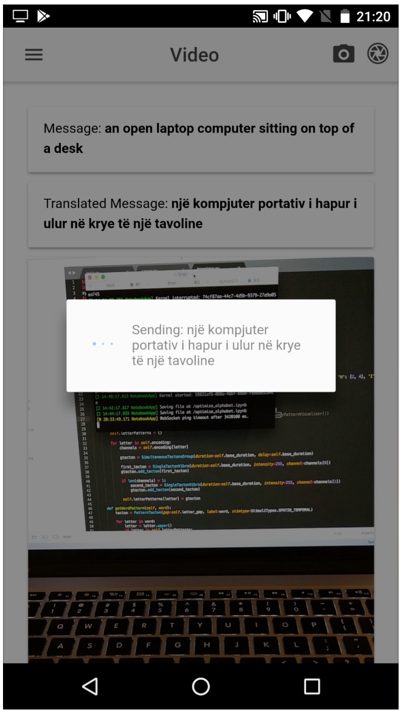

Figure 4.31: VTT with translation.

is the same. In both cases, the user would use a necklace phone holder which they would wear around their neck. They have two main interaction choices:
1. Press physical volume up button for a single snapshot of the environment. In the case of STT, the application records until it detects that the talking side is finished talking. For VST, this action takes a single image and then describes it.
2. Press volume down button for continuous capturing. For SST this would mean that the voice recording is always on. For VST this option would take periodical (e.g. every 30 seconds) snapshots of the environment where the frequency depends on a configuration parameter. In case the continues capturing is already on, this user action turns it off.

The main interactions intentionally rely on the physical buttons (see Figure 4.29) which are easy to reach on the mobile phone when positioned on the chest of the user (hanging on the necklace). Additionally, both applications allow users (or caregivers) to control different settings, mainly connected to the stimulation parameters and translation.

The proposed applications (see figures 4.30, 4.31 and 4.32) support Android and iOS operating systems, and they rely on existing models for both speech and object recognition. They rely on the speech recognition APIs offered by the mobile operating systems whereas object recognition is accomplished using the Microsoft's Cognitive Services API [9].

4.5.4 Discussion and Future Work

This section presented a concept and implementation of two mobile applications which capture the user's environment and then convey its textual description to the user through a vibrotactile wearable display. Such applications are fully mobile and wearable, and as such, they present a huge potential to provide assistance and enhance the lives of users with visual and hearing impairments. The representation of information in the textual form is chosen as recent works demonstrate that vibrotactile skin reading already is feasible (see Sections 4.1, 4.2, 4.4 and [Luzhnica et al., 2016b]).

The implemented applications run on smartphones due to their popularity. However, the same principles (sensor-text-tactile) could be used to provide the same functionality using other mobile or wearable devices equipped with the necessary sensors (microphone or camera) and processing units. For instance, vision to tactile application requires only a camera as a sensor which can be found in most of the smart glasses (e.g. Google glasses). Thus users could use such wearable devices instead of a smartphone to capture the environment.

In future work, it is planned to evaluate the presented prototypes with the target user groups, reiterate and improve the proposed and implemented prototypes by including their feedback. Additionally, in this version, no filter (Figure 4.29) was implemented as initially it should carefully be investigated and determined beforehand the content to be filtered. Thus, enhancing the applications with such filtering capabilities will be a goal for future work.

4.6 Summary

The main objectives of the work reported in this chapter are to provide an encoding of English Alphabet using overlapping spatiotemporal patterns and evaluate such

[9] `https://azure.microsoft.com/en-us/services/cognitive-services/directory/vision/`

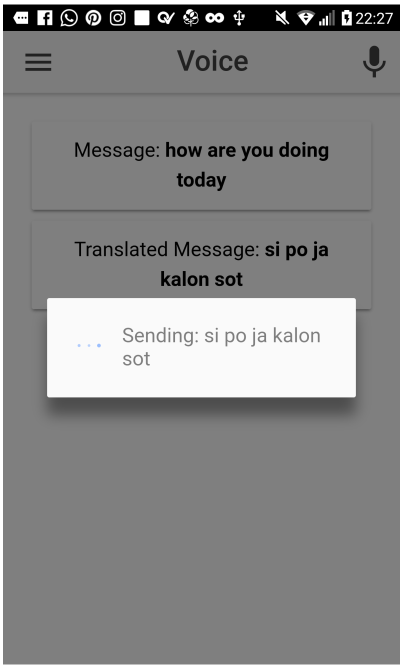

Figure 4.32: STT performing speech recognition, translating it to the user's language and then conveying the text through the wearable vibrotactile display.

patterns and the layouts proposed in previous chapter for skin reading. It evaluates the ability of participants to decode letters and words using wearable vibrotactile device. In addition this chapter also investigates other details of skin reading such as the background perception of vibrotactile encoded messages and passive ways of training. The conclusions in this chapter have been supported by four user studies.

The Study 5 in Section 4.1 validated empirically the feasibility of using such a wearable vibrotactile display with six vibromotors on the hand and forearms for skin reading. The study used overlapping spatiotemporal patterns and a letter frequency based encoding for letters of English Alphabet. The results showed that participants performed similarly on recognition of letters ($89\% - 92\%$) and words ($85\% - 90\%$) in both layouts (hand and forearms). Nevertheless, in both layouts, systematic errors occurred during both letter and word recognition. Further analysis showed that indeed there is room for improvement. Thus, a two step optimisation process was proposed to avoid such issues which resulted in a new layout of seven vibromotors, In addition to the layout a new encoding was proposed which was optimised in the basis minimising the probability that two subsequent letters in a word share a vibromotor to improve word recognition accuracy. This thesis also proposed a concrete algorithm for solving such a minimising problem which leverages the information of the structure of a language.

The Study 6 in Section 4.2 evaluated the proposed optimisations and it shows that such optimisation steps result in significant improvements in both letter and word recognition accuracies. Participants achieve an accuracy of 97% on letter recognition and 97% on word recognition within five hours of training. Such an outstanding accuracy of comprehension makes skin reading a good candidate for real-world applications.

The Study 7 in Section 4.3 investigated whether high speed vibrotactile encoded messages can be perceived in the background while performing other concurrent attention-demanding primary tasks. The results observed that users could very accurately comprehend vibrotactile encoded messages in the background and other parallel tasks did not affect users performance. Additionally, the comprehension of such messages did not affect the performance of the concurrent primary task. Such results promote the use of vibrotactile information transmission to facilitate multitasking.

Last but not least, the Study 8 in Section 4.4 investigated the potential of passive haptic learning (PHL) as a training tool for vibrotactile skin reading. The testing of the recognition of letters and words shows when trained (for 32 minutes), participants could recognise letters with an average accuracy of 69% and words with an accuracy 70%. Additionally, the study shows that PHL can be used regardless of whether the training is based on semantically grouped letters or alphabetically ordered ones. Moreover, the results show that participants recognitions accuracy was not affected by transmission speed indicating that they could be trained with a default speed and then proceed to use the system in different levels of speed without requiring a re-training. Overall such results demonstrate that PHL presents an alternative to active learning for training vibrotactile skin reading, while acknowledging that it is considerably less effective that active training.

Moreover, Section 4.5 proposes a novel technique of sensory substitution using vibrotactile wearable displays and the proposed conveying techniques by combining with speech and object recognition. Such application target users with visual or auditory impairments. In addition, it implements the concepts in two mobile applications.

Overall this chapter demonstrates that indeed the overlapping spatiotemporal patterns could be used for skin reading on the hand and forearms with relatively good accuracy. Moreover, optimising the encoding brings the accuracy to a near perfect (97% in both letters and words). However, for such results a layout of at least seven vibromotors is needed, which is provided in this thesis (see Sections 4.2 and 3.4).

Chapter 5

Conveying Continuous Numbers through Vibrotactile Wearable Displays

Vibrotactile technology has been used to communicate various types of information, supplementing or complementing other senses, supporting people with perceptual impairments, and augmenting those with normal perception. When encoding information, vibrotactile displays are commonly limited to transmitting a discrete set of tactile motives. The general approach is to encode symbols of generative symbols (e.g. letters) [Bliss et al., 1970, Geldard, 1957, Luzhnica et al., 2016b, Nicolau et al., 2013, Xu et al., 2011], that can be combined to form text [Geldard, 1957, Luzhnica et al., 2016b] as already discussed in Chapter 4.

However, many processes provide quantitative information that is not discrete but rather of continuous (real-valued) nature. For instance, the progress in a percentage of daily steps compared to the user's defined goal is a continuous value. One could still use discrete symbols to encode such values [Cauchard et al., 2016], by using having discrete tactons that represent numerical digits and then combine such numerical digits (in series) to form the desired value similarly to the process of forming words from letters (as applied in Chapter 4). That would be a good approach if high precision of comprehension is required. However, in several situations, users might benefit greatly from approximate information and some degree of inaccuracy is permitted as it is not crucial to the action taken by the user. For instance, when monitoring the running progress, the user might allow some degree of inaccuracy as it is not crucial to the activity of continuing or stopping to run.

The goal of the work in this chapter is to communicate quantitative values corresponding to continuous magnitudes for active feedback by providing a spatially continues feedback across the used wearable vibrotactile display. Such wearable vibrotactile displays for quantitative data could convey feedback on effort spent, the force applied by a tool, duration of a task or progress. The challenge of creating spatially continuous vibrotactile feedback is to imbue a continuous sensation with a limited, small number of vibromotors. For such a spatially continuous perception the phantom sensation is used.

The phantom effect has been used to create the illusion of spatial continuity [Cha et al., 2008, Israr and Poupyrev, 2011, Schneider et al., 2015, Seo and Choi, 2010]. It is an *interpolation* effect that occurs when two vibromotors are active, and the user attributes perception to a location in-between the active vibromotors due to an inherently low resolution of haptic perception [Alles, 1970]. Evaluation of spatially continuous tactile displays has so far mainly focused on the quality of perceived continuity, the consistency of perceived intensity across space, and the perceived smoothness of time-varying continuous *movements*.

For achieving the aforementioned goal, initially four vibrotactile layout displays are designed and an encoding concept is created. The vibrotactile layout displays and the encoding concept are evaluated in a comprehensive user study. The user study investigates user decoding precision of spatially modulated real-valued data, which is decoded by localising spatially continuous, temporally stationary (static) vibrotactile stimuli. It quantifies decoding precision across two vibrotactile display layouts (circular, worn around the wrist and the upper arm, and straight, worn along the forearm) each positioned at two different locations on the body. In each testing condition, it evaluates three perceptual models of phantom sensation from the literature with respect to localisation precision: the linear model, the log model, and the power model. Furthermore, this work introduces a data-driven method for vibrotactile display personalisation. Such a method adjusts perceptual models to idiosyncratic and spatial variations in perceptual sensitivity learned from user-specific data, and quantitatively compare localisation precision using all perceptual models with and without sensitivity adjustment.

Given the primary goal of this chapter is to convey inaccuracy tolerant continuous numerical values through wearable vibrotactile displays, this chapter targets the following research question:

For scenarios where high precision is not required, can we encode continuous values using a discrete number of actuators using phantom sensation? More precisely, how well (with what accuracy) can users decode such encoded values and does sensitivity adjustment increase such encoding/decoding accuracy?

The work described in this Chapter has already been published in a peer reviewed scientific paper [Luzhnica et al., 2017] (**P6**) and its findings resulted in contributions **C8** and **C9** listed in Section 1.2.

5.1 Progress-bar Inspired Vibrotactile Wearable Displays

The wearable vibrotactile displays designed in this chapter are inspired by progress-bars and mimic the same way of representing information. As shown in Figures 5.1 and 5.2, such progress-bars encode information visually by filling one part of the bar. The encoded value is the ratio between the filled part compared to the total length of the progress bar. Note that, from the visual representation it is rather difficult to decode precisely the encoded value and yet such information is sufficient and could be very useful in many scenarios (see Section 5.4.1 for potential use cases).

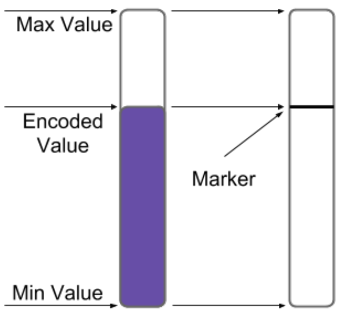

Figure 5.1: A straight progress-bar (left) and its simplified representation (right) where a marker indicated the value instead of the filled area.

A more simple representation of such progress-bar would be simply to mark the location of the value as shown on the Figures 5.1 and 5.2 (on the right). Such a

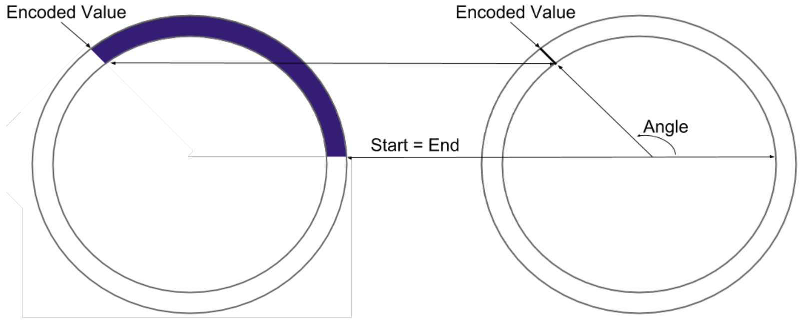

Figure 5.2: A circular progress-bar (left) and its simplified representation (right). The marker indicates the encoded value instead of a filled area. Alternatively, such a progress-bar could encode the angular value as illustrated on the right.

simplified representation will be mimicked by the vibrotactile displays where the marker will be stimulated using vibromotors of the wearable display.

Note that in addition to the straight progress-bars depicted in Figure 5.1, one could use also circular progress-bars as shown in Figure 5.2. Circular progress-bars are very useful in representing information for repetitive processes. Additionally, they could be used to represent angular information or orientation where the encoded angle is drawn from the centre of the circle to the location os marker.

5.2 Method

Human haptic perception has, depending on body location, a relatively low spatial resolution. Simultaneous stimulation of two or more locations in close proximity may only be perceived as a single stimulation somewhere *in between* vibromotors. This haptic illusion is typically referred to as phantom sensation [Alles, 1970]. The exact location of the perceived stimulus depends, among other factors, on the stimulation amplitudes [Alles, 1970, Schneider et al., 2015, Park et al., 2016]. This section starts by briefly describing three perceptual models of phantom sensation from the literature. Then it proposes a method for personalisation, extending these generic models by explicitly modelling and accounting for idiosyncratic and spatial variation in perceptual sensitivity. Furthermore, it outlines extending spatially continuous haptic displays to more than two tactons per dimension, and describes the estimation of local sensitivity from user data.

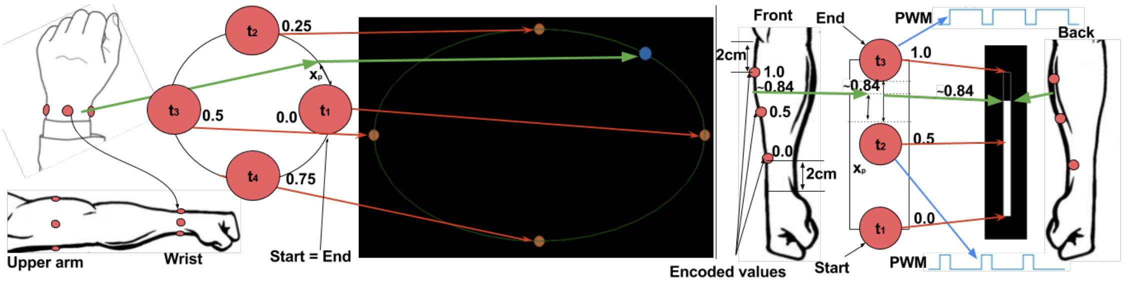

Figure 5.3: Circular layout, straight layout and their corresponding user interfaces used in the study (with black background). On the circular layout t_1 and t_2 are activated whereas on the straight one, t_2 and t_3 are activated (see also PWM signal) to encode the value.

5.2.1 Perceptual Models of Phantom Sensation

Perceptual models of phantom sensation for vibrotactile feedback can be described as a generative function $f : \mathcal{R}^T \rightarrow [0, 1]$, mapping stimulation amplitudes of a fixed set of T vibromotors onto a continuous location along one spatial axis at which stimulation is perceived. The inverse model $f^{-1} : [0, 1] \rightarrow \mathcal{R}^T$ determines the set of stimulation amplitudes required to invoke a phantom sensation at a given location.

The amplitude with which vibromotors are stimulated in practice is bounded within an interval $[A_{min}, A_{max}]$ by minimum perception thresholds, hardware constraints, and user comfort. Consequently, let us describe models with respect to normalised stimulation intensities $I \in [0, 1]$, derived using Eq. (5.1)

$$I = \frac{A - A_{min}}{A_{max} - A_{min}}, \quad A = I(A_{max} - A_{min}) + A_{min}. \tag{5.1}$$

Given a pair of vibromotors $i \in \{0, 1\}$, actuated with intensities I_i at locations $x_i = i$ along some spatial dimension, the location of the phantom sensation can be described as $x_p \in [0, 1]$.

Let us introduce the *linear* model, the *log* model and the *power* model [Alles, 1970, Israr and Poupyrev, 2011]. In the linear model, shown in Eqns. (5.2), x_p is proportional to the relative stimulation intensity of I_1 and, inversely, stimulation intensities are proportional to their proximity to x_p. In the log model, described by Eqns. (5.3) and (5.4), the stimulation intensities have a logarithmic relationship with perceived location (x_p). The recently proposed power model [Israr and Poupyrev, 2011, Schneider et al., 2015] is based on the energy summation of Pacinian channels [Makous et al., 1995]. Here, the perceived location is dependent on the square of stimulation intensities as in Eq. (5.5),

$$x_p^l = \frac{I_1}{I_0 + I_1}, \quad I_0^l = (1 - x_p), \quad I_1^l = x_p \tag{5.2}$$

$$I_0^g = \frac{\log(2 - x_p)}{\log(2)}, \quad I_1^g = \frac{\log(1 + x_p)}{\log(2)} \tag{5.3}$$

$$(1 + x_p^g)^{I_0} = (2 - x_p^g)^{I_1} \tag{5.4}$$

$$x_p^p = \frac{I_1^2}{I_0^2 + I_1^2}, \quad I_0^p = \sqrt{1 - x_p}, \quad I_1^p = \sqrt{x_p}. \tag{5.5}$$

Between, log and linear models, usually the log model is preferred [Alles, 1970, Seo and Choi, 2010, Schneider et al., 2015] mainly as the perceived intensity of phantom effect decays towards the middle [Alles, 1970, Seo and Choi, 2010]. Nevertheless, the results of [Seo and Choi, 2010] suggest that linear model might be better than log model for location accuracy between two vibromotor. As for the power model, the authors of [Schneider et al., 2015] noted participants' preference of the power and log model over the linear model whereas between the power and log, there was no significant preference of one or the other.

Constructing $f(A_0, A_1)$ and $f^{-1}(x_p)$ can be achieved by combining each model with Eq. (5.1). The phantom location $f(A_0, A_1)$ is estimated by mapping stimulation amplitudes to intensities using Eq. (5.1)(left) followed by estimating x_p^l, x_p^p or x_p^g. Determining the stimulation amplitudes that create a phantom sensation $f^{-1}(x_p)$ involves estimating intensities (I_0^l, I_1^l), (I_0^p, I_1^p) or (I_0^g, I_1^g) and corresponding amplitudes using Eq. (5.1)(right).

5.2.2 Sensitivity-adjusted Perceptual Models

All models in the literature implicitly assume a constant stimulation sensitivity across locations on the body at which vibromotors are placed. They can, therefore, be considered as *generic* models. Cholewiak [Cholewiak and Collins, 2003] provided evidence suggesting that there are fine-grained spatial differences in perceptual sensitivity on the forearm: the area around the wrist and upper part towards the elbow are more sensitive than the middle part of the forearm. As there are spatial variations in perceptual sensitivity, it is reasonable to expect that perceptual sensitivity also varies across users. One indication for this is provided by [Rahal et al., 2009], showing strong gender-based differences in preference between the linear model and the log model. Therefore, this work proposes to create *personalised* perceptual models by explicitly incorporating user-specific measures of local sensitivity at vibromotor locations.

Intuitively, the higher the stimulation sensitivity at some location the lower the stimulation intensity needs to be to create a fixed intensity sensation. Thus, to account for spatial variations in perceptual differences of stimulation sensitivity, this work proposes to scale intensities I_i with a user and location-specific scale factor $s_i \geq 1$. This local rescaling is independent of the perceptual model of phantom sensations and can, therefore, be applied to all three models as shown in Eqns. (5.6)-(5.8),

$$x_p^{ls} = \frac{s_1 I_1}{s_0 I_0 + s_1 I_1}, \quad I_0^{ls} = \frac{1 - x_p}{s_0}, \quad I_1^{ls} = \frac{x_p}{s_1} \tag{5.6}$$

$$I_0^{gs} = \frac{\log(2 - x_p)}{s_0 \log(2)}, \quad I_1^{g} = \frac{\log(1 + x_p)}{s_1 \log(2)} \tag{5.7}$$

$$(1 + x_p^{gs})^{s_0 I_0} = (2 - x_p^{gs})^{s_1 I_1} \tag{5.8}$$

$$x_p^{ps} = \frac{s_1^2 I_1^2}{s_0^2 I_0^2 + s_1^2 I_1^2}, \quad I_0^{ps} = \frac{\sqrt{1 - x_p}}{s_0}, \quad I_1^{ps} = \frac{\sqrt{x_p}}{s_1}. \tag{5.9}$$

For the generic and personalised log models, Eqns. (5.4) and (5.8)), $f(A_0, A_1)$ can not be computed in closed form. Thus, optimization techniques should be used for finding x_p^{gs}.

After formalising how models are extended to chains of $N > 2$ vibromotors, it will be described how sensitivity values s_i can be estimated from data.

5.2.3 Extension to Chains of $N > 2$ Vibromotors

When two vibromotors are further apart than a few inches (depending on their location on the body), the user senses two separate stimulations instead of one: one at each vibromotor location. With additional vibromotors, the area on the body can be increased without losing the phantom sensation. Thus. let us explore vibrotactile displays with chains of three and four tactons.

Let us consider the general case of representing values $v \in [0, 1]$ with N equidistant vibromotors $0 \leq i < N$ and M segments between vibromotors. In order to apply Eqns. (5.2)-(5.9), we need to determine the relevant pair (a, b) of vibromotors for stimulation and estimate the within-interval value x_p from v. This is accomplished using Eq. (5.10), where % is the modulo operator,

$$a = \lfloor vM \rfloor \,\%\, N, \quad b = (a+1)\,\%\, N, \quad x_p = vM - a. \tag{5.10}$$

Figure 5.3 illustrates a straight and a circular tactile display, the values v corresponding to vibromotor locations, a phantom sensation and its distance to the two closest vibromotors.

5.2.4 Data Driven Sensitivity Estimation

While sensitivity values s_i could naively be adjusted manually or pre-configured for different body positions, this work proposes to estimate them from calibration data. Let us consider a set of D datapoints $\{(y_j, I_1^j, ..., I_N^j)\}$, where $y_j \in [0,1]$ is the stimulus location perceived by the user when vibromotors are set to stimulation intensities $I_1^j, ..., I_N^j$.

The problem of finding optimal sensitivities $S : s_1, ..., s_N$ can be expressed as a minimisation problem of the mean squared error between encoded values and user responses:

$$S_o = \arg\min_S \frac{1}{D}\sum_{j=1}^{D}(y_j - v(S, I_1^j, ..., I_N^j))^2, \quad s_i \geq 1. \tag{5.11}$$

Note that this optimisation requires the evaluation of $v(S, I_1^j, ..., I_N^j) \in [0,1]$ in every step, which involves estimating x_p. While x_p can be calculated in closed form for the linear model and the power model using Eqns. (5.6) and (5.9), finding x_p for the log model involves solving one optimisation problem for every data point:

$$x_{po} = \arg\min_{x_p}((1 + x_p^{gs})^{s_a^j I_a^j} - (2 - x_p^{gs})^{s_b^j I_b^j})^2, \; 0 < x_p < 1. \tag{5.12}$$

For solving the optimisation problem, the L-BFGS-B algorithm [Byrd et al., 1995] is used, which is a quasi-Newton method that can handle simple bounding box constraints.

5.3 Study 9: Continuous Wearable Vibrotactile Displays

A user study was conducted to quantify real-valued data decoding precision using spatial modulation of a stationary stimulus. Decoding precision was measured under twenty-four conditions: two display layouts, two body positions per display layout, and three perceptual models with and without personalisation through sensitivity adjustment.

5.3.1 Participants

Participants were recruited among students of a local technical university. Sixteen participants (all right-handed, ten male, aged between 26 and 41, and six female, aged between 21 and 32) volunteered to take part in the study. Only one female participant had previously taken part in a study on haptics. Detailed participant characteristics, including wrist and upper arm circumference, and forearm length, are presented in Table 5.1.

Gender	Age	Wrist	Upper Arm	Forearm
Male	25.60 (2.89)	17.47 (1.23)	28.56 (2.64)	27.13 (1.67)
Female	29.67 (5.45)	15.33 (1.21)	26.33 (4.59)	23.37 (2.65)
All	27.12 (4.48)	16.67 (1.60)	27.72 (3.64)	25.72 (2.77)

Table 5.1: Participant characteristics (mean and standard deviation): Age, wrist circumference (cm), upper arm circumference (cm), and wrist to elbow forearm length (cm).

5.3.2 Apparatus

An Arduino Due board (see Fig. 5.4) controlled a set of $3.4mm$ vibrotactile motors of type ROB-08449 (Voltage range: $2.3V - 3.6V$; Amplitude vibration: $0.8G$). Instead of changing the vibration amplitude directly, different intensities of vibration are generated by varying PWM duty cycles.

Int his work, two vibrotactile display layouts are designed to encode real-valued data through spatial modulation. The circular layout (Fig. 5.4, left) with four equidistant vibromotors was worn around the wrist and the upper arm. This layout

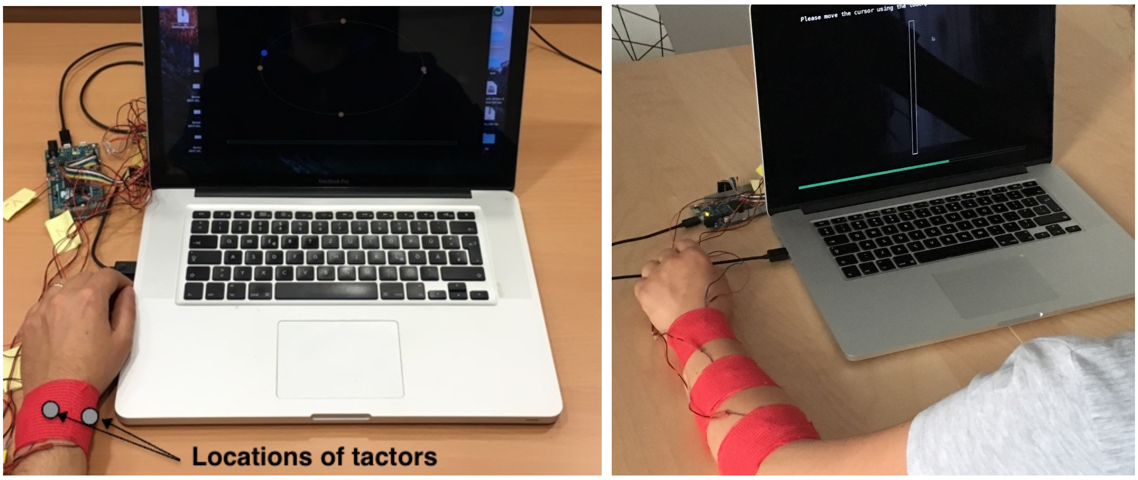

Figure 5.4: Participants performing the study on the wrist (left) and forearm (right). Note that the hand (left) is supinated relative to the rest position, which may result in user inputs being shifted.

is envisioned to be useful for encoding values that represent the state of repetitive or circular processes, orientation or angular information. The straight layout (Fig. 5.4, right) was fitted with three equidistant vibromotors and was worn at the front and back of the forearm. The straight layout applies to the more general case of encoding real-valued scalars within a bounded range.

5.3.3 Procedure

Each participant tested both display layouts (counterbalanced start with either circular or straight layout) in both positions (counterbalanced start in either position) on their non-dominant (left) arm, with four trial conditions in total. Note that the distance between vibromotors varies with wrist circumference, upper arm circumference, and forearm length of each user.

Introduction Phase Each experiment (i.e. each pair of display layout and position) was started with an introduction phase ($3 - 4$ min) involving a verbal explanation and a trial of the vibrotactile display. These varied depending on the display layout.

Straight Layout: It was explained that stimulus locations represented percentage values as follows: the vibromotor position at the wrist corresponded to 0%, the position of the vibromotor closest to the elbow corresponding to 100%, and values in between increased linearly along the arm. Then, the graphical user interface (GUI) for testing was shown on screen, consisting of a single vertical progress bar (see Fig. 5.3, right). It was explained that the empty bar corresponded to 0% the full bar corresponded to 100%. In preparation for the trial run, 80 values were sampled uniformly. Every value was then stimulated with the corresponding GUI representation shown on the screen. This was to familiarise participants with the vibrotactile display and to test the phantom effect: as a condition to proceed to the test phase, each participant was asked whether they felt one or two stimuli at a time, whether stimuli were also perceived in between vibromotors, and whether they found a correspondence between the location of the stimuli and the values shown on the screen.

Circular Layout: It was explained that the stimulus location represented a direction as on a compass, where the top vibromotor corresponded to North, the right vibromotor corresponded to East, and so forth. Then, the GUI for testing was shown on screen, consisting of a circle with four red dots (see Figure 5.3, left). It was explained that the top dot corresponded to North, the right dot corresponded to East, and so forth. In preparation for the trial run, 72 values were sampled uniformly. Every value was then stimulated with the corresponding GUI representation shown on the screen, where a marker in the shape of a blue dot appeared at the location corresponding to the stimulus. Participants proceeded to the test phase under the same conditions as explained for the straight layout.

Test Phase In preparation for the test phase, the sequence of values was shuffled such that stimuli were applied in a new random order. The test phase uses the exact samples of stimuli as the introduction phase. One stimulus was applied, and participants were asked to mark the value corresponding to the perceived stimulus on the GUI. For straight layouts, users were asked to use the mouse to click on the empty progress bar at the location that corresponded to the stimulus. For circular layouts, users were asked to adjust a blue visual marker initially set at North by moving the mouse and to confirm the location with a click. Participants received the stimulus until their response was confirmed and the next stimulus was applied. For each stimulus, the participant's response and the set of vibromotor intensities are logged. Upon completion of the test phase, participants continued with the introduction phase for the next combination of layout and position.

5.3.4 Data Preprocessing

During the experiment, it was noticed that the circular display was shifted when participants' hands were supinated or pronated relative to the rest position (see Figure 5.4, left), introducing a systematic bias in user responses. In order to avoid systematic over-estimation of errors due to this misalignment, all user inputs by the negative mean error were shifted. The mean errors were estimated independently for each user and body position.

5.3.5 Results

For statistical analysis, let us considered two body positions for each layout, circular (C): 1) wrist and 2) upper arm, and straight (S): 1) back of the forearm and 2) front of the forearm, and six perceptual models: power (P), linear (L), log (G), sensitivity-adjusted linear (LS), sensitivity-adjusted power (PS) and sensitivity-adjusted log (GS). The independent variables were layout $\in$ {C,S}, body position $\in$ {1,2} and model $\in$ {P, L, G, PS, LS and GS}.

From each participant's data for a given layout and position, six sets of encoded values v were generated, one for each model, by transforming logged vibromotor intensities to stimulus locations (see Eqns. (5.2)- (5.9)). This avoids repeating measurements with participants and justifies paired t-test comparisons of random samples. The data within each combination of user, layout and position was randomly split into equally large training and test sets. The training set was used to infer user and location-specific sensitivities for sensitivity-adjusted models. The test set was used to evaluate all models. Thus, all the error rates reported below are based on the test set only.

The absolute decoding error was estimated as $\epsilon = |v - y| \in [0, 1]$, where $y \in [0, 1]$ is the user response and $v \in [0, 1]$ is the encoded value. Within each combination of layout, position, and model, each participant's central tendency of absolute error was estimated. As absolute the errors were not normally distributed, the median is used as the central tendency. This aggregate error was the dependent variable in this analysis. In total, statistical analysis was based on 16 (users) $\times$ 2 (layouts) $\times$ 2 (positions) $\times$ 6 (models) = 384 measurements.

Layout	**Position**	**Generic Models**			**Sensitivity-adjusted Models**		
		ϵ_l	ϵ_p	ϵ_g	ϵ_{ls}	ϵ_{ps}	ϵ_{gs}
Circular	**Upper Arm**	.059 (.014)	**.054 (.014)**	.056 (.014)	.053 (.016)	**.051 (.014)**	.054 (.013)
	Wrist	.052 (.009)	**.046 (.006)**	.049 (.008)	.049 (.010)	**.044 (.008)**	.048 (.010)
	Both	.056 (.012)	**.050 (.012)**	.053 (.011)	.051 (.013)	**.048 (.012)**	.051 (.012)
Straight	**Front**	.082 (.031)	.082 (.026)	**.080 (.029)**	.070 (.018)	.074 (.022)	**.068 (.020)**
	Back	.090 (.022)	**.085 (.027)**	.086 (.022)	.067 (.018)	.073 (.027)	**.064 (.020)**
	Both	.086 (.027)	.084 (.026)	**.083 (.026)**	.069 (.018)	.074 (.024)	**.066 (.020)**

Table 5.2: Decoding error for each layout, position and model. Notation: ϵ_l - linear model, ϵ_p - power model, ϵ_g - log model, ϵ_{ls} - sensitivity-adjusted linear model, ϵ_{ps} - sensitivity-adjusted power model and ϵ_{gs} - sensitivity-adjusted log model.

Layout	ϵ_l **vs** ϵ_p	ϵ_l **vs** ϵ_g	ϵ_p **vs** ϵ_g	ϵ_l **vs** ϵ_{ls}	ϵ_p **vs** ϵ_{ps}	ϵ_g **vs** ϵ_{gs}	ϵ_{ls} **vs** ϵ_{ps}	ϵ_{ls} vs ϵ_{gs}	ϵ_{ps} vs ϵ_{gs}
Circular	**0.001**	**0**	**0.008**	**0.006**	0.205	0.175	0.069	0.925	**0.032**
Straight	0.603	**0.008**	0.697	**0.001**	**0.01**	**0.001**	0.116	0.116	**0.006**

Table 5.3: Statistical significance in p-values of paired t-tests.

The decoding errors under all tested conditions are given in Table 5.2 and depicted in Fig. 5.7. For circular layouts, the power model (P) performed best on average among all generic models, and the sensitivity-adjusted power model (PS) showed lowest mean error overall. For straight layouts, the log model (G) outperformed other generic models on average, and the sensitivity-adjusted log model (GS) best-explained user responses overall. Under every single condition, the sensitivity adjusted model produced a lower mean decoding error compared to the corresponding generic model.

The effects of independent variables on decoding error were analysed using factorial ANOVA. Significant effects were found for layout; $F(1, 360) = 175.56, p = 0.0$, and model; $F(1, 360) = 4.64, p = 0.0$. A significant interaction effect was found between layout and model; $F(1, 360) = 2.7, p = 0.02$. There was no significant interaction effect between layout and gender; $F(1, 360) = 0.017, p = 0.98$, or model and gender; $F(1, 360) = 1.7, p = 0.13$.

A separate factorial ANOVA test for each layout is also performed. With the circular layout, it was found a significant effect of body position; $F(1, 180) = 14.15, p = 0.0$, but none of the model; $F(1, 180) = 1.76, p = 0.12$. For the straight layout, body position had no significant effect; $F(1, 180) = 0.20, p = 0.65$, but the model did; $F(1, 180) = 4.12, p = 0.001$.

The results of post-hoc paired t-tests are presented in Table 5.3. Comparing generic models on the circular layout, all models differed significantly from each other, whereas on the straight layout only a significant difference between the linear model and the log model was found. Comparing generic and sensitivity-adjusted models on the circular layout, only the linear model showed a significant difference. On the straight layout, all sensitivity-adjusted models showed significantly lower error than corresponding generic models. Among sensitivity-adjusted models, the power model and the log model differed significantly with both layouts.

Inspecting the decoding error observed with circular layouts and the sensitivity-adjusted power model, there was no significant difference between body positions; $t(32) = 2.04, p = 0.058$. Equally, there was no significant difference between body positions; $t(32) = 0.65, p = 0.52$ among straight layouts with sensitivity-adjusted log model.

Pearson correlation analysis confirmed that there was no significant correlation between the error on the straight layout (using GS) and the forearm size; $r =$

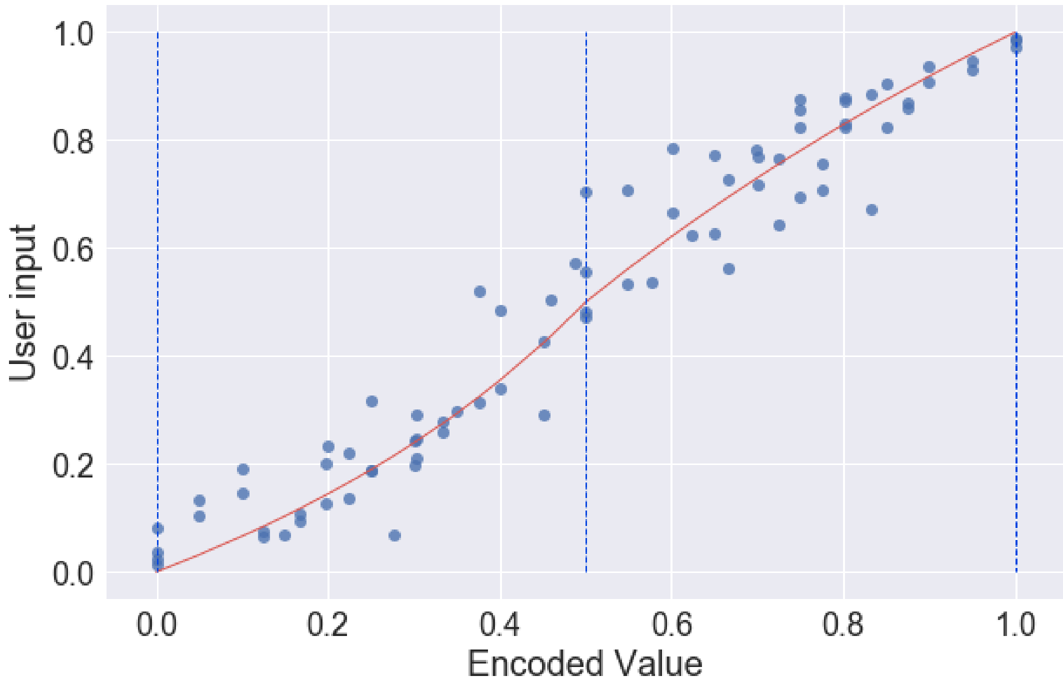

Figure 5.5: The sensitivity-adjust linear model (continuous red line) approximates user's perception of the encoded value on a straight display. The x-axis represents the encoded value using a linear model and the dotted lines represent positions of vibromotors.

$0.21, p = 0.22$. Similarly, there was neither a significant correlation between the error on the upper arm display (with PS) and the circumference of the upper arm $r = -0.07, p = 0.77$ nor the wrist display the circumference of the wrist $r = -0.35, p = 0.17$.

5.4 Discussion

This study demonstrated that paricipants can decode a real-valued values with a mean error of only 4.4% for circular displays (on the wrist) and 6.4% for straight displays (on the back of the forearm). This low error was achieved with the proposed method for personalisation of perceptual models for tactile displays using sensitivity adjustment. The proposed method consistently outperformed the corresponding generic model with regards to decoding error of real-valued data. This improvement was significant for the best model (the log model) for straight layouts, validating our

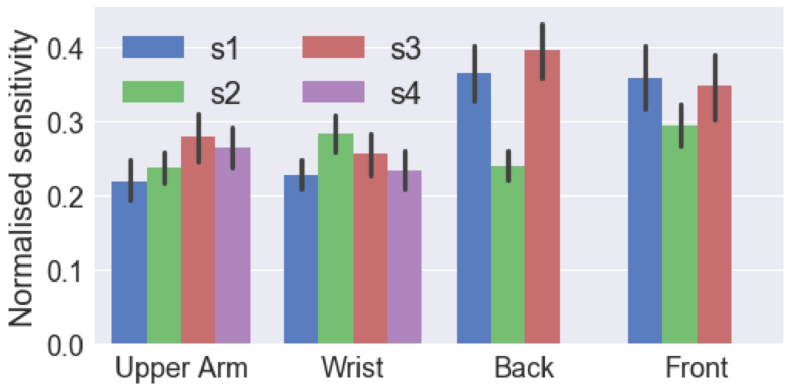

Figure 5.6: Normalised optimal user sensitivities.

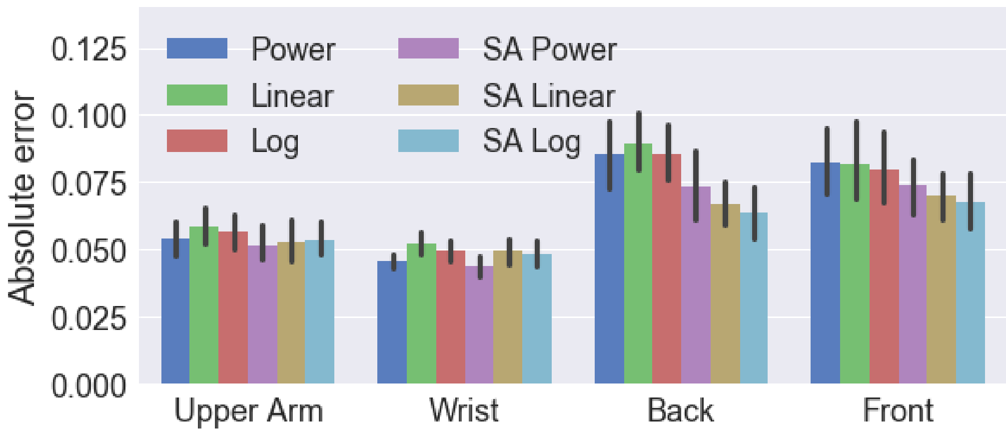

Figure 5.7: Absolute errors for each layout, position and model.

approach. The user study showed that error-reducing local perceptual sensitivities can be inferred from user data. An example of how a linear sensitivity-adjusted model approximates user's perception is presented in Figure 5.5.

Wrist-based circular displays showed lower decoding error than displays worn around the upper arm. However, there are practical aspects that differentiate those displays further. Wrist-based displays may be easier to offer as mainstream products (e.g. alongside smartwatches). The potential for movement around the wrist can, however, pose problems. The user might interpret values differently when decoding values from a rotated display, perhaps relative to the body position. This could be avoided by using motion sensors on the wrist to detect and compensate for the orientation of the wrist when encoding a value. Such sensors are typically available on wristbands, but determining the orientation is prone to error. This problem does not occur on the upper arm as pronation and supination are not physically possible.

Similarly, a higher accuracy was observed on the back of the forearm than at the front, but it is often in contact and quite often attached to the body. This could affect the perception of vibration if the motors are pushed towards the body.

The circular layout performed significantly better than the straight layout. To some degree that is not a fair comparison, as the former had four and the later had three vibromotors. Depending on the application and the nature of encoded values, one layout may be better suited than the other (and vice-versa). For the applications representing a state of a circular or repetitive processes (e.g. direction guide, representing an angle, the progress of a lap in a racing game) the circular display would be a great choice. For other applications, the proximity of two extreme values in a circular layout and the risk of them being mistaken for each other would make it an impractical choice.

Figure 5.6 presents the sensitivities of locations for each display averaged among all users. Note that as the ratio between s_i values within S is important, the sensitivities are normalised within S ($\sum_i s_i = 1$). Focusing on the front and back displays, we see that areas near the wrist and the upper part of the forearm are more sensitive than the middle part. This is in line with findings on vibrotactile localisation on the forearm by Cholewiak [Cholewiak and Collins, 2003]. There is considerable variation across participants (visualised by black lines) at each location. This could be attributed to idiosyncratic variations of sensitivity levels. In addition to skin sensitivity, factors influencing how strong a vibration is perceived include how tight a vibromotor is attached to the skin, vibromotor orientation and manufacturing inaccuracies, especially for cheaper devices. Thus, the test equipment can directly affect the model.

As presented in the Tables 5.2 and 5.3 the sensitivity adjusted models outperform generic models in predicting the user perceived value by capturing different levels of sensitivity of the locations, personal sensitivity variations and the device (vibromotor) characteristics within a small set of parameters (s_i). It is interesting that, on circular layouts, the sensitivity-adjusted power model is most accurate, whereas, on straight layouts, the sensitivity-adjusted log model performed best. As there is no closed form solution to predicting locations with the log model, the process of computing optimal sensitivities is computationally expensive. However, once the sensitivity parameters S_o are estimated, encoding values can be done in closed form for all models (Eqns. 5.6, 5.9 and 5.7). Thus, this is not expected to be an important issue in most of the cases, but designers of wearable devices might consider this in very specific cases (e.g. this process needs to occur often and the calculations need to happen in device).

There is room for further improvement on the decoding error rate. In this study participants used such a display for the first time. Long term use and feedback on user performance is expected to result in better perception and recognition of encoded values. Other haptic related studies provide evidence of learning effects when using vibrotactile devices for longer duration [Kaul and Rohs, 2017, Luzhnica et al., 2016b].

5.4.1 Potential Usage Scenarios

The proposed tactile displays can be used in scenarios where angular, repetitive or quantitive information can be encoded, and high precision is not required. Some such potential usage scenarios are listed in the following:

- **Gaming**. Both displays can be used in gaming as a supportive modality of interaction. The circular display could be used to encode states such as current position in lap when playing a car racing game (e.g. at 45% of the lap), orientation such as the enemy is coming from left-behind (225°) in a combat game, navigation such as turn right or take the 60° turn in any game that requires navigation (car racing, combat, etc...), multiplayer interaction such as someone is asking for a ball to your top-right (75°) in a football game. The straight display could be used to indicate the progress of current level of the game, the level of ammunition in a combat game, level of damage or the force of hitting the opponent in a fighting/combat game, travelling speed/racing in a car racing game, etc...

- **Healthcare**. Patients with hand amputee prosthesis or neuroprosthesis lack the tactile and kinesthetic sensation on the hand. While such prosthesis can help regain grasp function and some basic movements, the user still needs feedback on simple things such as: how much force is applied, how far is the wrist rotating. In such scenarios, a straight display can encode the pressure applied to an object while grasping. The circular display can be used to encode the orientation of the hand while pronating and supinating (where the thumb could be used as a reference vector). Such information is crucial for proving closed loop control. Although, as for the wearable vibrotactile displays used in this work, the chosen locations are upper arm and forearm, they can only be directly applied in scenarios where one of the arms is sensible (perhaps only one arm is with prosthesis). Nevertheless, the same principle used in this work can be utilised to provide a haptic display in other body locations.

- **Industry/Factory**. In industrial and production line settings, often employees need to interact with the machine through a touch screen or other conventional interfaces for adopting different settings, observe the results of the interaction (the product being produced) and maybe obtain instructions during this process. Some information can be transmitted through tactile displays to reduce the visual overloading. Employees can feel some of the current settings (e.g. intensity, temperature, the amount of fuel, etc..) and concentrate on the process of the product without needing to look at the display of machine.

- **Fitness**. Awareness for everyday activity could be effectively conveyed through such a haptic display. For instance, the level of progress while running, walking, stair climbing goal achievements can be reminded to the user (80% of the goal is achieved). A straight display would be more appropriate for such encoding but a circular display would be more practical as it could fit within a smartwatch or wristband. Such devices are very often used to track the activities and adding vibromotors would make them bidirectional (sensing and feedback).

5.5 Summary

This work investigated user decoding precision of spatially modulated real-valued data, which was decoded by localising spatially continuous, temporally stationary vibrotactile stimuli. A user study quantified decoding precision using six perceptual models across two vibrotactile display layouts.

With the goal of improving the decoding accuracy, this work developed personalised sensitivity-adjusted perceptual models for tactile displays based on the linear, log and power models. The user study evidenced that the proposed models consistently outperform corresponding generic models, increasing the accuracy of decoding information by participants. The best sensitivity-adjusted models approximated users' perception with an average error of only 6.4% for straight display layouts and only 4.4% for circular display layouts.

Increasing decoding accuracy has the potential to improve the sense of realism and acceptance by users, and facilitate new applications of tactile displays in a wide range of real-world application areas.

Chapter 6

Exploring Interactions for Skin Reading

As portrayed in the Chapter 4 vibrotactile skin-reading uses wearable vibrotactile displays to convey dynamically generated textual information. Such wearable displays have potential to be used in a broad range of applications. While vibrotactile skin reading is very different from visual reading, there are common patterns and practices from readers that might be shared across all kind of readings. A good example of such common practices are the evidences produced by studying reading patterns of visual reading and Braille reading. Despite of being very different in nature, interaction with the text and the navigation through it, remains fairly similar at its core.

A common belief that reading is a sequential task, where eyes *glide smoothly across the page*, is merely an illusion [Rayner, 1998]. At the word level, well-established research postulated that words are recognised as units [Larson, 2004, Fisher, 1975, Reicher, 1969, Cattel, 1886] and they are even recognised before individual letters [Cattel, 1886]. Reading depends on the mechanics of the visual system to stop at fixed spots in the text (fixations) and jump quickly to other spots (saccades, covering about 8 letter spaces) [Rayner, 1998]. Skilled readers fixate on about 2/3 of the words in a text. Beside forward movements to advance in reading, they reread nearby material backwards in the text about 10 to 15% of the time, occasionally driven by breakdowns on comprehension. Conversely, beginning readers fixate every word (often more than once), perform shorter saccades, and up to 50% of their eye movements are regressions, as they rely more on context to identify words [Rayner, 1998]. Obtaining meaning from printed words is not sequential; it depends on processing words as units and uses backward jumps at word level to aid understanding. In Braille, it is not possible to form a global shape recognition of the entire word, so the text has to be processed character by character [Daneman, 1988, Millar, 2004, Millar, 2003]. The perception and flow of information in Braille are controlled by moving the hand forward and occasionally backwards to revisit information [Millar, 2003, Hughes et al., 2011]. Thereby, Braille readers control reading speed, focus on particular letters or re-scan entire words.

On the other hand, vibrotactile skin reading is passive: a pattern of vibrations is stimulated by the device from start to end while users have no control over the transmission. Yet, vibrotactile displays can evoke the perception of words as units, by means of tactile animations [Kirman, 1974b, Tan et al., 1997]. The question that drives this research is what interactions are needed for efficient skin-reading? Users may not understand parts of the text due to lack of concentration or training. They need ways to pause, resume and jump to previous units of meaning or change the speed of transmission to account for progress in their reading skills.

To compensate for the drawback of being passive, this chapter investigates what kind of interactions are necessary for vibrotactile skin reading and the modalities of such interactions in order to equip users with means of controlling the reading process. Thus this chapter targets the following research question:

What interactions are necessary for skin reading? What is the preferred modality for such interactions?

To investigate such skin reading interaction, initially, an interaction concept is created to enable reading interactions for vibrotactile displays. A formative study trained novice users to recognise letters and words, and tested their behaviour while skin-reading sentences. The study analyses participants' interaction behaviours and a questionnaire to determine what interactions are useful and what are the preferred means of interactions. Finally, this work maps the interactions to gestures and discuss the wearable design choices that could allow the used wearable vibrotactile display to be extended for supporting such gesture-based interaction concept. The work described in this Chapter has already been published in a peer reviewed scientific paper [Luzhnica and Veas, 2018a] (**P7**) and its findings resulted in contributions **C10** and **C11** listed in Section 1.2.

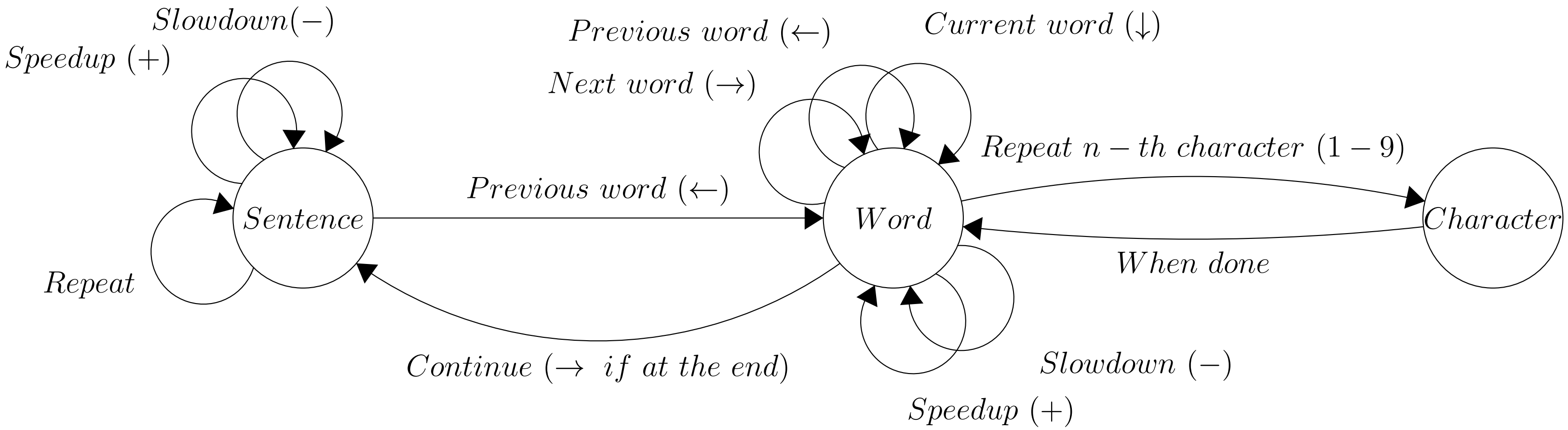

Figure 6.1: Interaction concept during sentence transmission. States (Sentence, Word and Characters) represent what the system is transmitting to user.

6.1 Interaction Concept

This section describes the design of an interaction concept for textual vibrotactile skin reading, illustrated in Figure 6.1. The concept is based on a virtual fixation point metaphor. The fixation point represents the word that is currently being transmitted to the user which will be referred to as the current word. While perceiving text/sentences, at whatever point in time, users have the possibility to request retransmission of the current word. In this case, the system transmits the current word and transfers itself into the pausing state (corresponds to Word state in Figure 6.1) where no further text is transmitted until resumed.

For elaboration purposes, let us assume that the user paused on the n-th word of a text. While on the pausing state, the user can repeat the current (n) word or navigate to the previous ($n-1$) word in the text. In this case, the fixation point shifts to the left in the text and the ($n-1$) word is transmitted and becomes the current word. At this point, the user can repeat the current word ($n-1$), regress to the previous one ($n-2$), or go to the next one (n). Hereby, the user navigates back an forth and scans the text. If the fixation point is at position n and the user navigates to the next word (beyond the point where it was paused), the system resumes and starts transmitting the rest of the words. Additionally, when the system is in the pause mode, the user can repeat particular characters of the fixated word. Furthermore, users can also change the speed of transmission which would proportionally change gaps between characters and words; and the activation time of each vibration motor (see Figure 6.4).

6.2 Study 10: Investigating Interactions and their Modalities for Skin Reading

To investigate the proposed interaction concept and determine which of the interactions are useful for the user, a user study was conducted. The user study combines participant training and testing of characters, words and sentences. An additional goal of the study was to investigate the word recognition process. However, the topic is out of the scope of this chapter, and thus the results and findings concerning this investigation are deferred to future work.

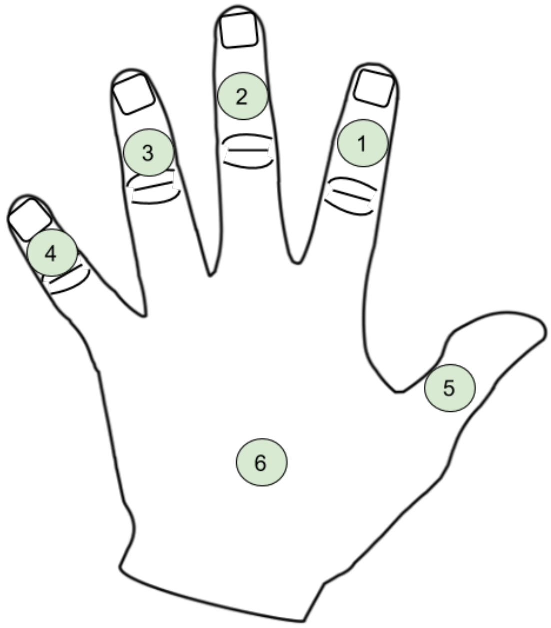

Figure 6.2: The used vibrotactile display containing six vibromotors. The design has been borrowed from the Section Section 4.1. The number of the vibromotor indicates the priority of activation.

.

6.2.1 Wearable Haptic Display Design

A layout design with six vibromotors on the back of the hand is used identical to the User Study 5 described in the Section 4.1 (see Figure 6.3). With it, the ten letters in the study can be encoded with combinations of one or two vibromotors. The rationale behind using only of six vibromotors is that only ten letters will be encoded for this user study. But, for encoding the entire alphabet, a layout with more vibromotors as proposed in Sections 4.2 and 3.4 would be a better choice.

6.2.2 Vibrotactile Patterns and Encoding

Each letter is encoded with one or two vibromotors using an OST (overlapped spatiotemporal) stimulation pattern described in Chapter 3. Moreover, the order of activation is prioritised by the sensitivity of the finger, since it yields a higher accuracy in identification of locus as revealed by Study 3 in Section 3.3. Sensitivity order is assumed according to studies suggesting that sensitivity decreases from the index finger towards the little finger: the index finger is more sensitive than the middle, ring, and pinky finger [Duncan and Boynton, 2007, Vega-Bermudez and Johnson, 2001, Hoggan et al., 2007]. The thumb is the lowest sensitive [Sterr et al., 2003]. For instance, for a letter encoded in index and pinky finger, the vibromotor on index finger would be activated first, and then after a gap, the vibromotor placed on the pinky finger.

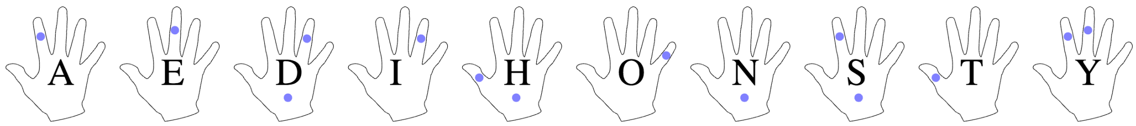

Figure 6.3: The encoding scheme of each character used during the study.

Figure 6.3 illustrates which vibromotors are used to encode each of the characters used in the study. In addition, Figure 6.4 illustrates the technical details of the stimulation process of characters, words and sentences. Character encoding uses a base duration (d) of 200 ms and a 10 ms gap (g) between the activation of vibromotors. This means that the duration of a character (ld) is 200 ms for one vibromotor and 210 ms for two-vibromotor letters. When constructing words, a between letter gap (bl) of 200 ms is used to separate sequential letters. With such encoding, a word containing four characters can be transmitted within 1400-1440 ms. Note that, users can be trained to recognise letters and words with shorter duration when exposed to longer training periods (see Section 4.2). However, this study aimed at having training and testing in a single session. Hence, a longer durations was used. Additionally, sentence encoding uses a between word gap of 600 ms.

6.2.3 Characters

To keep the study in a manageable time, this study use only ten characters: A, E, I, O, T, N, S, H, D and Y. A small alphabet ensures a shorter period of training for characters and still enables us to create words and sentences for investigating interactions.

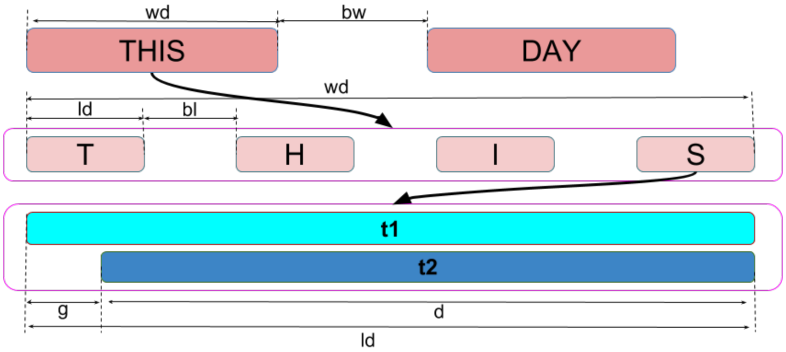

Figure 6.4: Stimulation process of characters, words and sentences. Base duration (d) represents the activation time of a vibromotor. Words in a sentence are transmitted in series separated by a gap (bw = 600 ms). Within words, characters are transmitted in series with a gap in between (bl = 200 ms). The characters are encoded using OST patterns where vibromotors are activated in sequence with a gap in between (g = 10ms). Note that this figure is included for sake of completeness as the same concept has already been introduced in the Chapter 4.

Length	2-char	3-char	4-char	5-char
Words	is he it	tea say hot	easy does this	shiny stand notes

Table 6.1: The list of words that are used during the user study.

6.2.4 Words

With the chosen characters, a list of 12 words words is composed, containing two to five characters. The words have been selected from the list of basic English words[1] and they will be used to train the user with words.

6.2.5 Sentences

In addition a list of 29 sentences (see Table 6.2) is composed. But only 15 of them were stimulated during testing, whereas the rest are there just to create more choices. The goal of the sentences testing was to observe how participants interact with the system while reading a sentence and how well they perform.

[1] https://simple.wikipedia.org/wiki/Wikipedia:List_of_1000_basic_words

He is the one, **She is the one**, **The tea is hot**, The sea is hot, **It is hot**, **He is hot**, She is hot, **Today is hot**, He says no, **She says no**, I say yes, **I say no**, **He says yes**, She says yes, I say yes, **This is easy**, It was easy, It is easy, **I hate this**, He hates this, **Hide this idea**, **It does stand**, This does stand, **It is done**, This is done, **This is shiny**, It is shiny, **It is noisy**, This is noisy.

Table 6.2: Sentences used during the user study. All sentences (29) were presented to the user in a list to select from. Only the bold marked ones (15) are used to test the user. The rest of the sentences (14) are used as decoys to make the process more challenging.

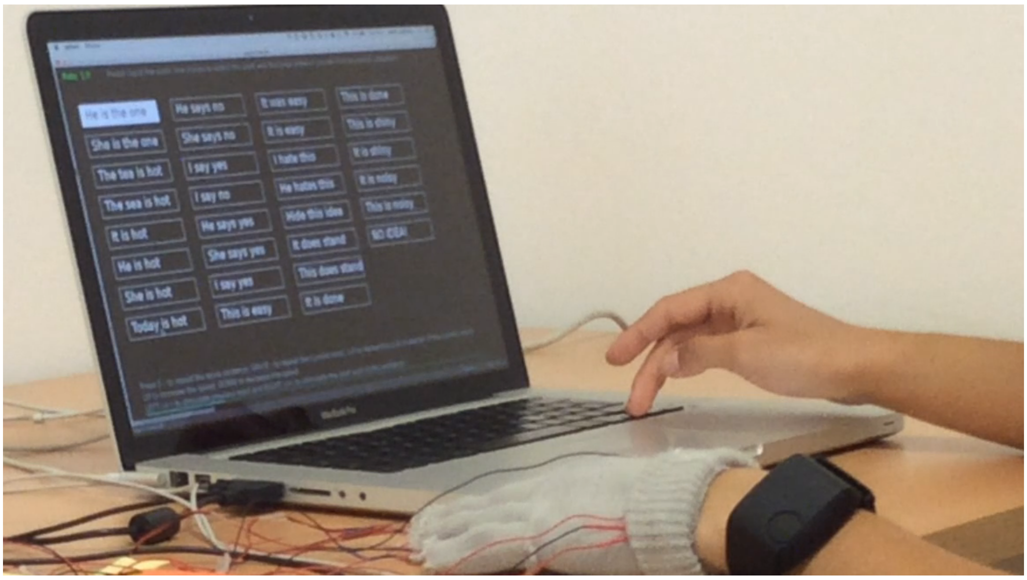

Figure 6.5: A participant performing a round of sentence reinforcement.

6.2.6 Procedure

The entire study was organised in several blocks which will be referred as rounds, each serving different purposes:

- **Character Training** trained users to associate a symbol with a vibrotactile pattern. During this process, participants were stimulated with patterns representing a character, the character was displayed on the screen, and an audio spelling of the character is simultaneously played as shown in Figure 6.6. Such a simultaneous technique of tactile, auditory and visual stimulation has been demonstrated to be efficient [Luzhnica et al., 2016b].

- **Character Reinforcement.** Participants were stimulated with a pattern and asked to input the character associated with it. After entering the answer, they were notified whether their input was correct and saw the correct answer (see Figure 6.6). This way they would learn from their mistakes. Participants were allowed to repeat the stimuli before answering.

- **Word Training** exposed users to simultaneous stimulation of vibrotactile, auditory and visual stimuli all representing a word. The process was similar to character training, but instead of characters words were used. Words were transmitted as a series of characters (see Figure 6.4).

- **Word Reinforcement**. This is similar to character reinforcement, but words are used instead. Participants are allowed repeat the stimuli. After stimulation, they were asked to select the answer from a list constructed with all the words shown in Table 6.1, plus 21 other words words including "No idea!". Participants were instructed to choose "No idea!" if they do not know what is stimulated. It was also pointed out to them that for every stimulated word there are other similar words in the list and thus they should avoid guessing based on few characters. Upon entering the answer, participants were informed on the display what would have been the correct answer.

- **Words Testing:** was similar to word reinforcement, but participants were not allowed to repeat the stimuli, and they were not notified the correct answer.

- **Sentence Reinforcement.** The process was similar to the word reinforcement but used sentences instead of words.

Participants were stimulated with a sentence and asked to select an answer from a list of sentences. As shown in Table 6.2, the list of choices contained 29 sentences plus the "No idea" option. However, users were tested only in 15 sentences and the rest of them were used to make the process more challenging. Each sentence was composed of three to four words (see Table 6.2). The main purpose of sentence reinforcement was to study interactions during skin reading.

Participants went through five rounds of character training and reinforcement as shown in Figure 6.6. The first round was split into two short rounds, where only half of the characters are used in each (A-T and N-Y). This way participants were

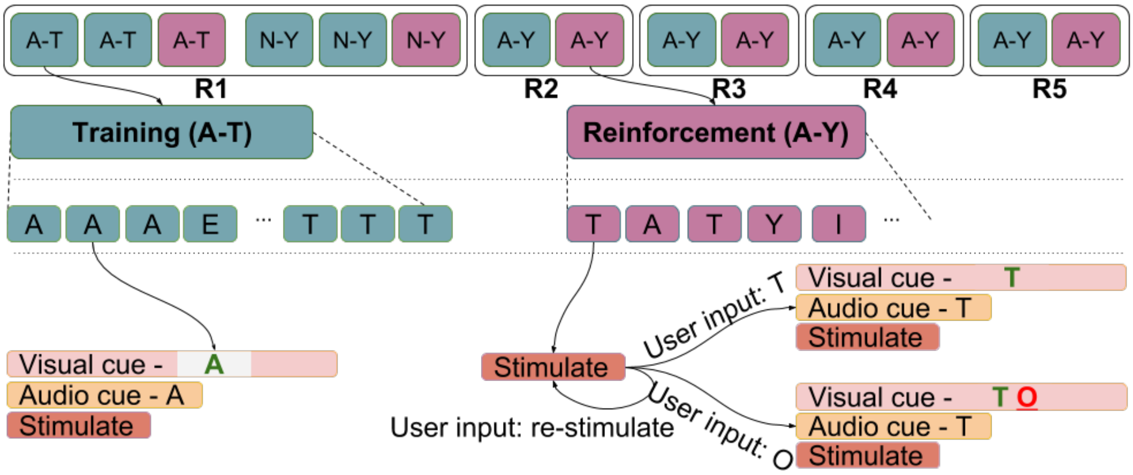

Figure 6.6: Character training and reinforcement process. Initially only five characters were introduced (A-T) then the next five letters (N-Y). For the the rest of letter training, all ten characters were used (A-Y). Colour coding: ● - train round, ● - reinforcement round.

introduced first to five characters and then to the next five. The following four rounds used all ten characters. In every training round, a character was displayed three times whereas in the reinforcement rounds each character was tested twice.

Thereafter, participants went through a round of word testing followed by a round of reinforcement using words from Table 6.1. Then, users went through four rounds of word training each followed by a round of word reinforcement. Finally, users were subject to one round of word testing followed by one round of word reinforcement where they were exposed to a combination of words they trained on (from Table 6.1) and an equal number (balanced by word length) of words they did not trained on. To finish the study, participants were subject of one round of sentence reinforcement where each of 15 sentences (see Table 6.2) was tested once.

6.2.7 Interaction

During the sentence reinforcement rounds, participants could use the interaction concept presented in Figure 6.1. They, could repeat the current word by using the keyboard space key, navigate words using left and right arrow keys, change transmission speed by using up and down arrow keys. They could choose to re-stimulate only a particular letter by pressing a number from 1-9, which would re-stimulate the n-th letter of the current word. Additionally, participants were able to repeat the entire sentence by pressing S. During word reinforcement rounds, participants were also able to repeat the entire word or particular characters of it. Similarly, during the character reinforcement, participants could repeat the character.

6.2.8 Data Collection

First, all user responses and interactions during testing and reinforcement rounds were logged. Additionally, at the end of the session, users filled a questionnaire asking questions about how they would use such a wearable device on a daily basis. Initially, they were asked to rate whether they find the interactions for skin reading as: (i) Necessary, (ii) Optional or (iii) Not useful. Also, for each available interaction, users were asked to rate how often they think they would use it by selecting one of the available choices:

- Continuously - every couple of seconds or minutes,
- Often - every couple of hours,
- Not very often - every couple of days or weeks,
- Rarely - few times only, and
- Never.

Furthermore, three modes of interaction were proposed:

- Gesture-based: the user uses hand gestures to interact,
- Smartphone-based: an application in the user's smartphone is used to interact,
- Physical buttons based: physical buttons would be added to the vibrotactile glove for interaction.

Participants were asked to rate (0-10) how suitable each of the proposed modality would be for interactions with the wearable display. Furthermore, they were asked to choose one modality they would use for interactions/commands that they would use more often and one for the interactions they would use rarely.

6.2.9 Apparatus

Our device consisted of an Arduino Due board which controls $3.4mm$ vibrotactile motors of type ROB-08449 (Voltage range: $2.3V \sim 3.6V$; Amplitude vibration: $0.8G$).

6.2.10 Participants

Twenty-two (22) individuals (12 male and 10 female) aged between 17 and 38 (M=26.7, STD=5.5) years old participated in this study. The overall study took approximately 90 minutes. Only one of them was left-handed. All of them used the left hand for stimulation and the right to interact with the computer as depicted in Figure 6.5. One participant, for personal time constraint reasons, completed the character and word rounds but did not continue with sentence reinforcement round.

6.2.11 Results

Let us define four common variables: accuracy, repetition, total duration. Accuracy will be defined as a binary variable set to be one if the user's answer is correct. Repetition describes how many times a user repeated the stimulation (character, word or sentence) within a reinforcement round. The total duration represents the entire duration from the time stimulation was first initiated by the system until the user responded, including repetitions. Additionally, let us define an interaction to be a repetition of any kind. E.g. during sentence reinforcement round, each repetition such as: current word, previous word, next word, a particular letter of the current word or the entire sentence is considered to be an interaction. Although, this section will provide a brief overview of performance to give an impression of users training level prior to sentence reading, it will mostly focus on interactions during sentence reinforcement round as the reading performance is out of the scope of this section.

Performance

Each character reinforcement round collected 20 probes (2×10 characters) for each user. Table 6.3 presents the results of character recognition, including the average accuracy, repetition, duration and total duration. By the third round, participants could already recall characters with a high accuracy (M=95%). While on the next

Round	Accuracy	Total Duration (s)	Repetitions
1a	0.98 (0.15)	3.29 (4.03)	0.22 (0.60)
1b	0.82 (0.38)	5.06 (8.41)	0.57 (1.28)
2	0.86 (0.25)	5.57 (9.58)	0.72 (1.52)
3	0.95 (0.23)	5.58 (23.8)	0.72 (1.81)
4	0.95 (0.22)	4.39 (6.56)	0.63 (1.37)
5	0.95 (0.22)	3.33 (5.58)	0.39 (0.90)

Table 6.3: [
Results of character reinforcement rounds. Note that we consider the first two rounds (1a and 1b) as two parts of round one as each of them contained only half the characters. This table shows the correct recognition rate (accuracy), average total duration, average duration and average repetition rate.

Round Type	Accuracy	Total Duration (s)	Repetitions
With Repetition	0.81 (0.39)	12.41 (12.27)	2.32 (4.28)
No Repetition	0.55 (0.50)	9.34 (8.45)	0.0 (0.0)

Table 6.4: Results of word recognition in the last reinforcement (with repetitions) and testing (no repetitions) rounds.

two rounds the accuracy does not improve, there is an improvement in repetition and duration which could be interpreted as them being more confident.

For the word recognition, let us focus on the last round of reinforcement and testing as they are considered to be the end result of the word training process. Note that, in the reinforcement round users are allowed to repeat word or letters whereas in the testing repetitions are not allowed. In each of the two rounds, 24 probes were collected for each user. The recognition accuracy, total duration and repetition rate are presented in Table 6.4. Additionally, the user recognition accuracy (averaged per user) is shown in the Figure 6.7. Both Table 6.4 and Figure 6.7 reveal that when repetitions are allowed, participants achieve a higher accuracy. Furthermore, a chi squared test reveals that participants achieve a significant higher accuracy ($M = 0.81, STD = 0.39$) in the round where they can perform repetitions compared to the round where they are not allowed to repeat ($M = 0.55, STD = 0.5$); $\chi^2(2, 1056) = 81.67, p = 0.0$.

The sentence recognition round collected 15 probes for each user. The average accuracy, number of interactions, duration and total duration for sentence recogni-

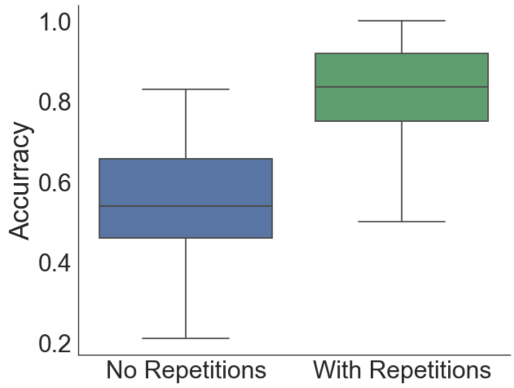

Figure 6.7: Averaged user accuracy for the last reinforcement (with repetitions) and testing rounds (without repetitions).

tion are presented in Table 6.5. On average, users needed 37.29 seconds to recognise sentences with an accuracy of 82%. Figure 6.8 shows histograms of recognition accuracy, the number of interactions and total duration averaged per user. While total duration is somehow evenly distributed, that is not the case for the recognition accuracy. The vast majority of users achieved a good accuracy (see histograms in Figure 6.8).

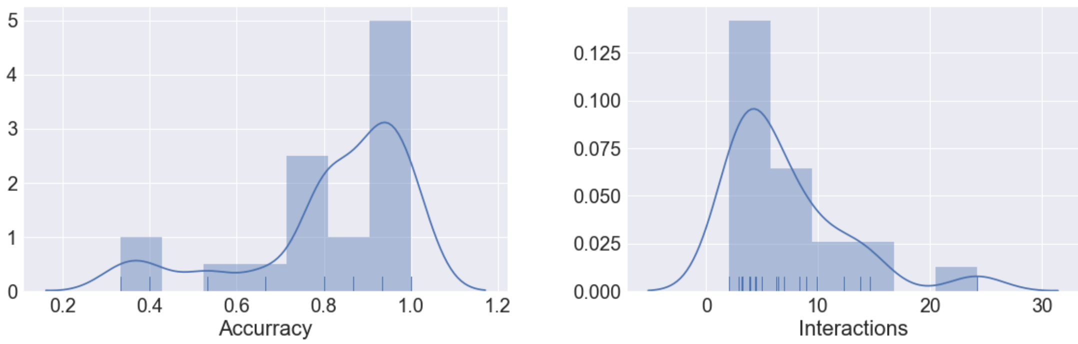

Figure 6.8: Histograms showing the distribution of the accuracy and the number of interaction during sentence recognition.

Interactions

First, let us analyse the interactions in the last round of word reinforcement. On average participants performed 2.32 (SD=4.28) interactions for each word. From the

Accuracy	Interactions	Total Duration (s)
0.82 (0.38)	7.14 (7.59)	37.29 (27.04)

Table 6.5: Averaged sentence recognition results (M, STD).

interactions, participants repeated 67% of words completely (at least once) whereas they repeated one or more single characters only in 6.8% of the words. Table 6.5 shows that participants needed on average 7.14 interactions to recognise sentences. Moreover, the histogram in Figure 6.10 shows that the vast majority of users needed a relatively low number of interactions. Eleven users needed on average five or fewer interactions per sentence but some users performed even over 20 interactions per sentence. Generally, word repetition was used (at least once) in 82.8% of the sentences, character repetition in 6.3% of them and sentences were repeated entirely in 21.2% cases. Within word interactions, 62.8% of them were repetition of current word, 25.6% of them were repetition of the previous word and only 11.6% were a repetition of next word in the sentence.

Figure 6.9 shows the overall state transitions probabilities between interactions during sentence reading. The chart is constructed from the interaction data of the sentence reinforcement rounds. The transition plot in Figure 6.11 shows the state transitions probabilities for the first ten iterations of interactions. The start state represents the time point when the sentence is fully transmitted the first time. The finish state represents the user providing the answer for sentence recognition.

The most likely interaction at the start is repeating the current word (probability = 0.57). Participants were also likely to start with previous word interaction (0.23) which could be interpreted as they already understood the current word. Precipitants were also likely to start with repeating the entire sentence (0.17). After a current word repetition, participants were most likely to continue with another current word repetition (0.6), meaning that they did not understand the word from the last repetition. They were also fairly likely to continue with previous word (0.16) or next word repetitions (0.1). They were quite likely to provide the answer (0.10). After a previous word repetition, participants were most likely to continue with another previous word repetition (0.38), meaning that they understood the word that was repeated and they were scanning the sentence backwards. They were also highly likely to continue with current word repetition (0.36), in cases where they did not

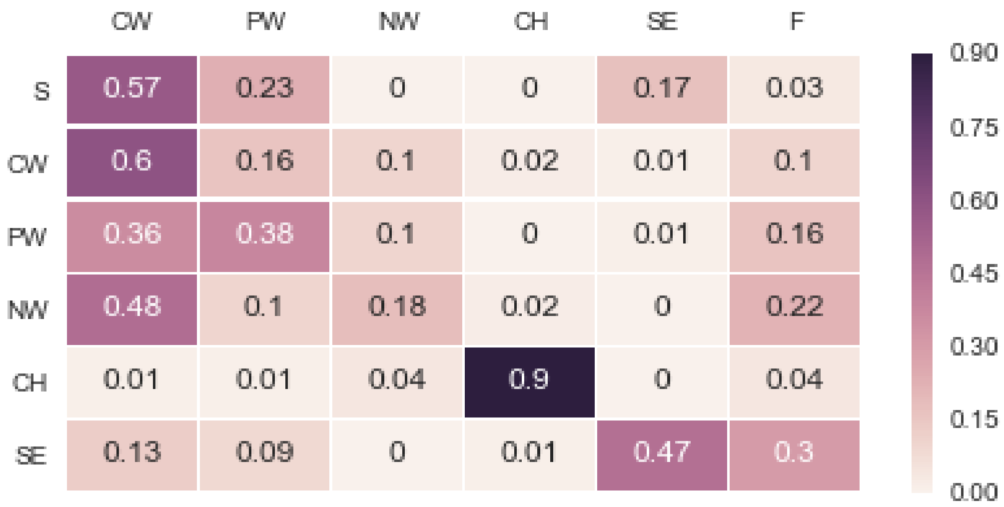

Figure 6.9: State transitions probabilities between interactions for sentence reading constructed from the sentence reinforcement round. Interaction states: S -start, CW - current word, PW -previous word, NW - next word, C - character, SE - sentence, F- finish.

understand the repeated word. On the other hand, they were relatively less likely to continue with a next word interaction (0.1). They were quite likely to provide the answer (0.16), meaning that they finished backwards scanning. After the next word repetition, participants were most likely to use a current word repetition (0.48) in the case where they did not understand the repeated word. They were also likely to use the next word interaction (0.18) again; scanning forward the sentence, or use the previous word interaction (0.1). They were quite likely to provide the answer (0.22), meaning that they finished forward scanning.

After the entire sentence repetition, users were most likely to continue with another sentence repetition (0.47), which could be interpreted as some users were simply repeating the sentence over and over until they were able to understand it completely. Such an interaction was relatively less but still likely followed by current word repetition (0.13) or previous word repetition (0.09). Users also were likely (0.3) to provide an answer. A similar behaviour pattern occurs after character repetition. Users were most likely (0.9) to repeat character again as users who used this interaction were repeating different letters of the current word.

Only five users adjusted the transmission speed during sentence recognition. Two of them set it to a higher than the default speed whereas three did slow it down.

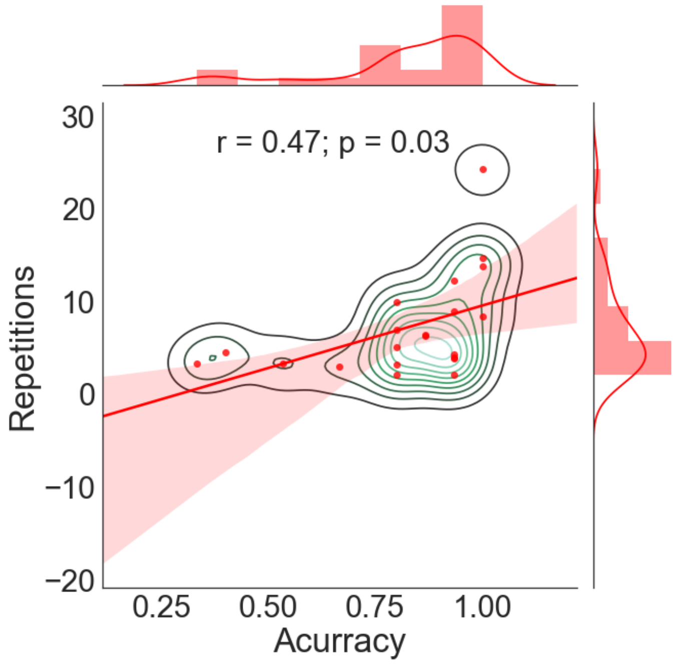

Figure 6.10: The relation between average sentence recognition accuracy and average number of interactions. The bar plots on the top and on the side represent histograms of the variable in the given axis. The contours represent the multivariate distribution of both variables. The straight line and the shades around it represent the fitted regression and its confidence. The Pearson correlation index and the confidence value are annotated as r and p.

Additionally, let us explore the relationship between sentence recognition accuracy and the number of interactions. Figure 6.10 shows that there is a positive Pearson correlation between the average recognition accuracy and the average number of interactions, meaning that users that interacted more, also recognised sentences more accurately; $r = 0.47, p = 0.03$.

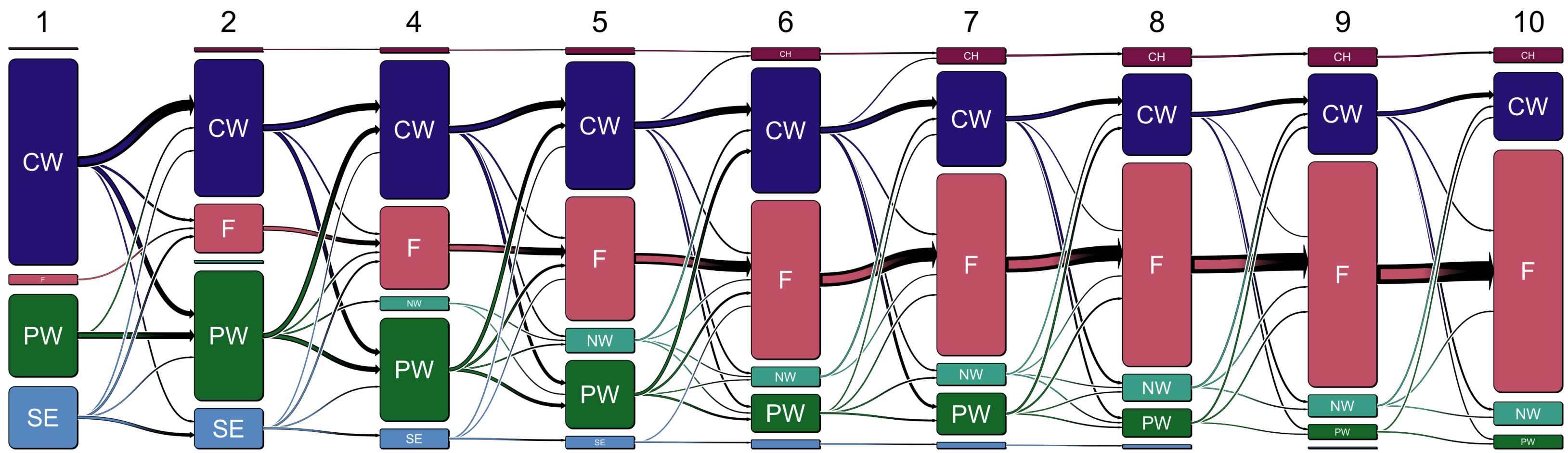

Figure 6.11: State transitions diagram between interactions for the first ten interaction durin g the sentence reading. Interaction states: S -start, CW - current word, PW -previous word, NW - next word, C - character, SE - sentence, F- finish. The size of the bar represents the probability of being in that state for the given interaction whereas the width of the arrow represents the probability of the state transition.

Questionnaire

On the question of how useful they found the interaction concept when performing sentence reading, 76% (16) of users rated it as "Necessary", 24% (5) of them rated it as "Optional" and no user rated it as "Not useful". When asked why users thought such an interaction was optional, they all argued that with a proper amount of training they would get proficient and there would not be a need for such interaction.

User ratings on how frequent they think each of the commands/interactions they would use are presented in Figure 6.12. For all word repetition interactions: repeat the current word, previous word, and next word, the vast majority of participants thought that they would use them continuously (every couple of seconds or minutes). Quite contrary, for adjusting the speed the majority if users think that they would rarely use. For repeating the n-th character of the current word, there is some divergence. While most of the users think that they would never use it, two users think they would use it continuously, and another four think they would use it often.

User ratings on how suitable the proposed modalities of interaction would be for skin reading application are presented in Figure 6.13. Gesture interaction received the highest rating ($M = 7.9, STD = 1.7$). But, a paired t-test reveals that the difference with physical buttons interaction ($M = 6.57, STD = 2.99$) is not significant ; $t(42) = 1.72, p = 0.101$. On the other hand users rated gesture interaction ($M = 7.9, STD = 1.7$) significantly higher than interaction using a smart phone ($M = 4.76, STD = 2.64$); $t(42) = 4.1, p = 0.001$[2]. Additionally, when users were asked to choose one preferred modality of interaction for commands they would regularly and for ones they would rarely use, users mainly prefer gesture-based interaction for regular interactions and smartphone-based interaction for rarely used interactions (see Figure 6.13).

6.2.12 Discussion

This study was designed to investigate and identify useful interactions for skin reading with a wearable vibrotactile display. The evidence from all sources such as user performance, interaction behaviour and questionnaire point out that when performing skin reading, users benefit from means of interactions with the vibrotactile

[2]The significance level is considered $\alpha = 0.025$ according to Bonferroni correction for two comparisons

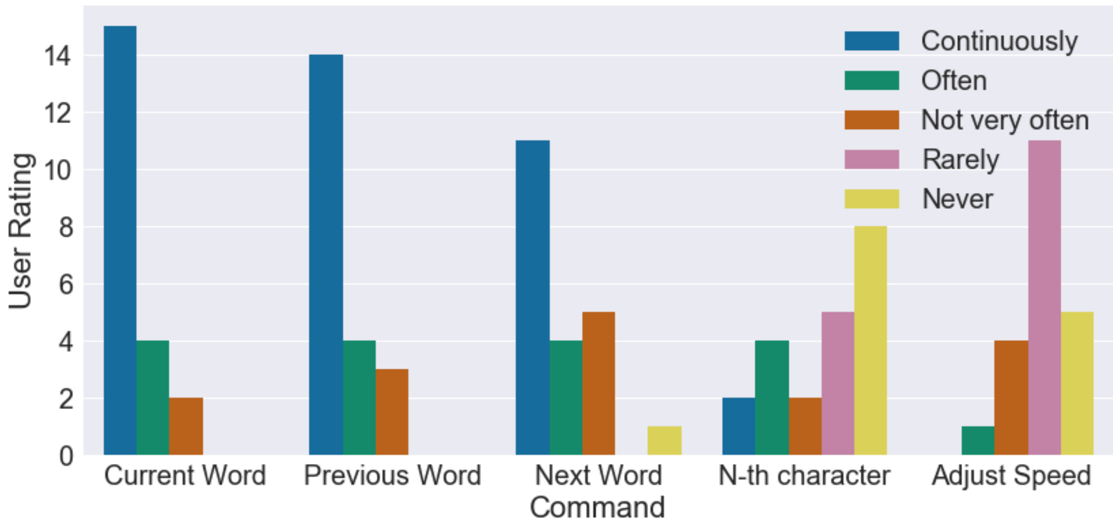

Figure 6.12: User ratings on how frequent they think they would use each of the interactions for text reading through a vibrotactile display.

display. First, the majority (76%) of the users explicitly expressed in the questionnaire that they think having interactions similar to the aforementioned experiment is necessary for skin reading. While some users (24%) expressed that when proficient, they would not need interactions, none of the believed that such interactions were not useful at all.

The necessity for interactions is also expressed in users' performance during sentence and words recognition. Interactions had a positive effect on the recognition accuracy. Participants performed significantly better (see Figure 6.7) in the word recognition rounds where repetition was allowed. Additionally, participants who on average performed more interactions in sentence recognition, achieved a better accuracy (see Figure 6.10).

Interaction usage also demonstrates that the interactions were necessary. Participants, on the last round of word recognition with repetition, on average needed 2.32 (SD=4.28) interactions for each word. Furthermore, they used on average 7.14 interactions for each sentence during sentence recognition.

As participants had no prior experience in skin reading, this study shows that at least for novice users, interactions are crucial. Users with minimal training can start perceiving words and phrases and would be able to understand them if the navigation interactions are at their disposal. Thereby, users do not need to become proficient

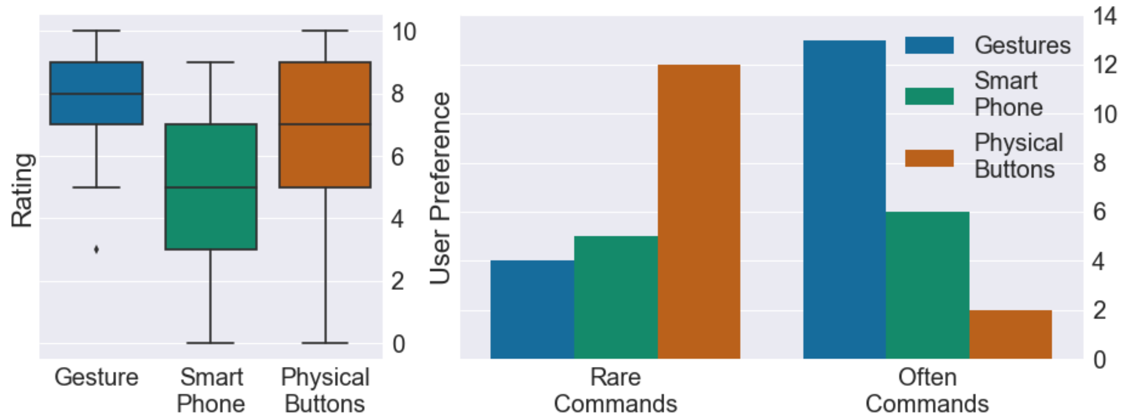

Figure 6.13: The box plot on the left visualises user ratings (0-10) on how suitable different modalities would be for interactions with a vibrotactile display during skin reading. The bar plot on the right visualises user preferences choices on which modality would be more useful for interactions/commands that they would use very often (at least every couple of hours) versus the interactions they would use rarely.

before they can start using such a wearable device. On the other hand, as users' skin reading skills increase, they would presumably need fewer interactions. They might learn to recognise entire words as units similar to visual reading [Woodworth, 1937, Smith, 1969, Fisher, 1975, Reicher, 1969, Cattel, 1886]. Nevertheless, that does not invalidate the need for interactions. First, users might need to repeat certain words from time to time as result of attention breaks or simply due to misperception. In both visual and Braille reading such interactions occur very often even if readers are not aware of it as it occurs unconsciously and naturally. Readers jump backwards to revisit already visited letters and words [Larson, 2004, Rayner, 1998, Rayner et al., 2001, Rayner et al., 2010]. This phenomenon is known as back regression, and skilled readers make regressions back in $10 - 15\%$ of the reading time [Rayner, 1998, Rayner et al., 2001, Rayner et al., 2010]. Such regression is common practice also in Braille reading [Millar, 2003, Hughes et al., 2011].

Besides emphasising the importance of interactions in skin reading, this study can be used to derive details of which interactions are most important. Both behaviour analysis and questionnaire analysis confirm that word repetition, and navigation interactions are critical. Participants used word interaction in 82% of sentences and the vast majority of users were convinced that such interactions would constantly be used. Repeating the n-th character of a word was rated as unnecessary by users and was mostly irrelevant to sentence recognition. A closer look at the usage of this interaction reveals that it was only used by seven participants. One participant used it in every sentence; one user used it only in two sentences whereas five users used it only in one sentence. Thus, only one user used character repetitions regularly in sentence recognition, whereas the rest did try to use them in one or two sentences to explore how well that would work but did not continue afterwards with the rest of the sentences. Adjusting the speed seems not to be frequently used. It was used only by five participants. Participants set the speed they were comfortable with and used it for the rest of the sentence recognition. This was also mirrored in the questionnaire where most of the participants expressed that they would only use it a few times during the lifetime of such device. Most of the participants pointed out that they would use it in accordance with the progress of their skin reading skill.

Interestingly and contrary to expectations, the repetition of the entire sentence was used by some users. Although, they were a small number, yet it was not expected for any user to rely on it. The prior expectation was that at first, users might be tempted to try it, but they would realise that it is challenging with their level of training to perceive the entire sentence at once. Such a scenario did occur with five users, where they used it only in four sentences or less. As shown in the Figure 6.11, at the first, second or third interactions some users switch to word-based interactions. Initially, they were curious and tried it for one or two iterations (see Figure 6.11), but then realised it was difficult and switched to other interactions. However, there were users who persisted using sentence repeats. Three users relied on this interaction almost entirely (used in more than 12 sentences), and two others used it moderately (in 7-9 sentences). Figure 6.11 shows that some participants repeatedly used this interaction until they provided an answer. Although it was interesting to provide such interaction within the frame of this study and explore user behaviour, only a small number of users used them. Moreover, in a real-world scenario, where text contains multiple sentences, and they are much longer, repeating the entire text might not very be useful as it does not scale.

The questionnaire reveals that the preferred modality for constantly used interaction would be gesture-based, whereas participants would prefer a settings smartphone application for rarely used interactions (adjusting the speed). For the constantly used interaction, one would need to provide a gesture-based system that would support the basic interactions: repeat the current word, go to previous word (and transmit it) and go to next word. Such a system would be a simplified version of the interaction system initially designed (see Figure 6.1). Stopping and resuming to transmit the rest of the text could be automatically achieved using the current word and next word interaction as explained in Section Interaction Concept.

6.3 Gesture Based Interaction

. One conclusion of the above described user study is that word navigation and repetition interaction are necessary for vibrotactile skin reading and that the preferred modality of such interactions is gesture based. Besides user preferences, from the engineering, designing and manufacturing perspective, a gesture-based interaction would be a perfect fit for a glove-based vibrotactile wearable display as the same glove could be equipped with sensing capabilities to enable gesture recognition. In this section, the interactions that resulted from the previous user study are mapped to hand gesture interactions and it is explored what sensors would be required to recognise them.

6.3.1 Gesture Mapping

The previous study concluded that only the interactions related to word navigation are essential for skin reading. In this section each of them is mapped to a hand-based gesture. The used gestures should be easy to remember, fast to perform, and contain simple movements so that they can be recognised with a minimal set of sensors. Thus, let us map the previous word interaction to swiping left gesture and next word interaction to swiping right. The mapping is natural, as it corresponds to movement of focus point within the sentence. Additionally, the interactions are easy to perform and considered in the research community as natural hand gestures [Romaszewski et al., 2014, Glomb et al., 2012, Luzhnica et al., 2016a] meaning that most users would be familiar with them. As for current word interaction, let us map it to swipe up gesture as it shares the simplicity and popularity with the other selected gestures.

6.3.2 Gesture Recognition

Gesture recognition has been a subject of many researchers. At a core abstract level, the main approaches for gesture recognition has been based on either environmental sensors such as: camera [Dardas and Georganas, 2011, Chen et al., 2007, Garg et al., 2009, Dardas and Georganas, 2011, Birk et al., 1997, Hasan and Mishra, 2012], radar signals [Fan et al., 2016, Lien et al., 2016] wifi [Li et al., 2016, Pu et al., 2013], etc... or hand/body-worn sensors such as : motion sensors, flexion and pressure sensors [Luzhnica et al., 2016a, Murakami and Taguchi, 1991, Xu, 2006, Neto et al., 2013, Zhang et al., 2009]. Even though each of them has its advantages and disadvantages, for the skin reading, wearable based sensors approach is a more suitable solution. First, it is not bound to the location (where sensors are placed) and second, adding the sensors to the same wearable glove, makes the setup much more convenient.

Many gesture recognition systems using wearable sensor have been proposed over the years, most of them rely on either hand worn sensors [Luzhnica et al., 2016a, Murakami and Taguchi, 1991, Xu, 2006, Neto et al., 2013, Zhang et al., 2009] or wrist-based sensors [Han et al., 2015, Van Vlaenderen et al., 2015, Xu et al., 2015, Zhao et al., 2015]. For recognizing gestures, typically statistical and machine learning approaches such as neural networks [Murakami and Taguchi, 1991, Xu, 2006], support vector machines [Luzhnica et al., 2016a], linear discriminant analysis (LDA) [Luzhnica et al., 2016a, Glomb et al., 2012], logistic regression [Luzhnica et al., 2016a, Zhao et al., 2015], decision trees [Zhao et al., 2015], etc.. have been employed.

The gesture recognition system used in this work utilises an existing framework and dataset from a previous work (Luzhnica et al. [Luzhnica et al., 2016a]). There the authors, collected data from 18 participants performing 31 gestures using a custom made data glove, where each participant performed each gesture 5-10 times. The data were annotated manually. Their data glove was equipped with seven inertial measurement units (IMU), one on each finger, one on the back of the palm and one on the wrist. Additionally, the glove was equipped with 13 bend sensors to cover main finger joint and wrist movement; and also five pressure sensors on each fingertip. The recording frequency of data glove was 85Hz (85 frames per second). For their recognition system, the authors used a sliding window approach upon which they extracted time domain features such as minimum, maximum, range, average, standard deviation and signal energy and time domain features such as Fast Fourier Transform for every sliding window. The authors evaluated different parameters for sliding window and different machine learning algorithms. They concluded that using LDA for dimensional reduction and then logistic regression for classification yielded the best results, where they achieved a f_1 score accuracy of 98.5%.

The apraoch of data processing, algorithm, training and evaluation procedure will be compleatly borrowed from the previous work [Luzhnica et al., 2016a]) and thus some of the extensive details (c.f., [Luzhnica et al., 2016a]) will be skiped in this section. From the dataset provided by the authors [Luzhnica et al., 2016a], only three (swipe left, right and up) gestures are utilised. Thus this work also explores what sensors are required to correctly identify the given gestures. More precisely, two particular sensors from the orginial set are explored: IMU on the back of the and the IMU on the wrist. Such motion sensors should be sufficient to capture motion characteristics of the intended gestures.

Considering that the number of gesture classes is reduced to three, the number of windows with rest class is very imbalanced. The rest class represents the data where the user is not performing any gesture such as: not moving or performing arbitrary movements. Thus he number of windows with rest class is reduced by randomly sampling a portion (5%) of them. In total, the entire resulting dataset (both training and test) contains 2959 windows. Similarly to LDA orginal work on building a gesture recognition [Luzhnica et al., 2016a] system, for dimensional reduction and logistic regression for classification trained on the training set which represents the 80% the data. The test set (the rest 20% the data) will be used to report on performance.

With only the IMU on back of the hand, the resulting classifier achieves a f_1 score accuracy of 98.4% on the test set. Using only the IMU on the wrist results in an accuracy of 96.5%.

6.3.3 Lessons Learnt

Overall using a single motion sensor (IMU) one would be able to recognise the necessary gesture-based interactions with a very high accuracy. The gestures can be better recognised by placing the sensor on the back of the hand (98.4%) as opposed

	Wrist				Palm			
	I	**L**	**R**	**U**	**I**	**L**	**R**	**U**
I	**385**	4	2	2	**386**	3	1	2
L	5	**61**	2	0	1	**68**	0	0
R	2	3	**54**	0	2	0	**57**	0
U	0	0	0	**60**	0	0	0	**60**

Table 6.6: Confusion matrix for classification in the test set using IMU on the wrist (left) and IMU on back of the hand (right). Classes: I - rest, L - swipe left, R - swipe right, U - swipe up.

to placing on the wrist (96.5%). The confusion matrix presented on Table 6.6 reveals that when using only the IMU on the wrist, there are some more misclassifications for classes left, right and the rest. Such misclassifications are less evident when using the IMU on the back of the hand. This could be explained by the fact that such gestures involve physical flexion and extension of the wrist which can easily be captured by the sensor on the hand but not on the wrist.

However, besides the accuracy, there are design and practical implications that might influence the decision of sensor location. First, there is a vibromotor located on the back of the hand of the used vibrotactile wearable glove (see Table 6.2), which is approximately located nearby the IMU on the data glove used to record the data [Luzhnica et al., 2016a]. Having both IMU and a vibromotor nearby might introduce noise in measurements when the vibromotor is active. A possible overcome could be achieved by shifting in opposite direction (left and right) to maximise the space in between. Alternatively, one could move the IMU on the palm side of the hand. On the other hand, a wrist-worn device makes it impossible to wear it along watches or wristbands. However, considering that a lot of users might already possess smart watches or wristbands equipped with a motion sensor, the motion data from their existing watch or wristband could be used to classify the required hand gestures. This would reduce both costs and power consumption of the wearable vibrotactile display.

6.4 Limitations and Future Work

One limitation of the proposed interaction system is that it deals only with text reading and not text comprehension, and thus it does not offer means of navigation beyond neighbouring words. For instance, while reading, users might want to revisit text 3-4 sentences backwards to better comprehend the text. Although, jumping larger portions of the text such as sentences could be provided analogously to the current proposed interactions. Such interaction could be mapped to rotation-based gestures like hand pronation and supination which could be easily be recognised using the motion sensors that proposed in this work. For that, we would need to train users for longer periods so that they would be able to perceive and understand larger messages in the first place. Furthermore, this work does not evaluate whether hand motion while performing the gesture could affect the ability of the user to perceive information during skin reading. Such effects need to be studied. If that were the case, a less convenient solution would be to use one hand for skin reading and the other one for interaction. Both limitations mentioned above will be considered for conducting additional studies in the future which is outside of the scope of this thesis.

6.5 Summary

This chapter investigates interactions for skin reading using a wearable vibrotactile display. Initially, an interaction concept for skin reading is proposed. In addition, a formative study with 22 users is conducted to evaluate the proposed concept during word and sentence reading with a six-channel wearable vibrotactile display. Participants were trained to recognise ten characters, trained on words and then tested on word and sentence recognition, during which they used the designed interactions. Furthermore, participants filled a questionnaire expressing their opinion about the interaction concept in general, different interactions within it and their preferred modality of interaction.

The results of the conducted user study and analysis of questionnaire indicated that interactions are beneficial for skin reading. Furthermore, this study shaped the proposed interaction concept by characterising interactions like character repetition as not necessary and transmission speed adjustment as less important. As a result, the end concept contains three main interactions, all of them providing word repetition and navigation within the sentence. For interaction modality, participants preferred gesture-based interactions. Following such a preference, the interaction concept was mapped to swipe-based hand gestures. Furthermore, motion sensors on the back of the hand and wrist are used to examine how well such gesture-based interactions would be recognisable using machine learning algorithms. The results of the study showed that a single motion sensor either on the wrist or hand is sufficient to recognise the gesture-based interactions with a high accuracy. Also, hand (98.4%) is a better choice for locating the sensor compared to wrist (96.5%) in terms of gesture recognition accuracy. Such a sensor could be incorporated into the same wearable glove providing one single solution for both skin reading and interaction. Thus, this work could serve as a guideline for designers and manufactures of such wearable vibrotactile displays.

Chapter 7

Summary and Conclusion

Conveying rich content through wearable vibrotactile displays presents opportunities for applications in various use cases involving general purpose usage or sensory substitution for users with visual or auditory impairment. When conveying such information, systems typically suffer from slow speed of transmission, comprehension inaccuracy (from users) or extensive training. The relationship between those three dimensions is inverse as improving one aspect degrades other aspects. This thesis investigates and proposes methods for conveying textual information as well as continuous numbers through wearable vibrotactile displays. In both cases, it starts from the ground up, by constructing stimulation methods, information encoding and wearable layouts to convey the information. When encoding information, it leverages the sensitivity of the stimulated locations and uses data-driven approaches to optimise the encoding in order to increase the information comprehension accuracy.

In the case of textual information, the process should optimise for both transmission speed and comprehension as for the vast majority of use cases; users should be able to perceive the text fast enough and also understand in precisely. Thus initially, in Chapter 3, this thesis proposes vibrotactile patterns that are discriminable when stimulated within relatively short durations ($\sim$100ms). Results of three user studies (in sections 3.1, 3.2 and 3.3) make a convincing case that the proposed sensitivity prioritised overlapping spatiotemporal (**OST**) patterns indeed provide a suitable mechanism for accurate perception when stimulated in short durations. In addition, Chapter 3 also takes the task of designing wearable display layouts that are suitable for stimulating such OST patterns. Initially, it uses hand based (on the back of the hand) and forearms based layouts with six vibromotors (see Section 3.1).

After evaluating the applicability of such layouts for skin reading in Chapter 4, it proposes a hand based layout with 8 vibromotors (see Section 3.4) which provides enough OST patterns to encode up to 36 discrete symbols (e.g. numbers, letters, phonemes, etc..).

Chapter 4 utilises the proposed layouts and OST patterns for skin reading. First, a letter encoding is introduced based on the frequency of letters, which maps more frequent letters to OST patterns that are composed of fewer vibromotors. Second, the methods to combine a series of patterns into words and sentences are introducted. Additionally, a training program was developed to teach users to associate OST patterns with the represented letters. A study (in Section 4.1) puts participants through the training program and evaluates their comprehension level of vibrotactile encoded letters and also words represented as a series of letters. Despite demonstrating that when using the proposed methods (patterns, stimulation, layouts and encoding) users can comprehend the encoded information with relatively high accuracy, the user study also reveals some limitations of wearable designs, OST patterns and encoding which are responsible for misinterpretation of encoded information by users. To account for such limitations, this thesis proposes a two-step optimisation process which extends the wearable layout and also optimises the encoding of symbols using a data-driven approach that leverages the bigram probability distribution for a given language. A second user study (in Section 4.2) evaluates such optimisations and reveals that they drastically improve the comprehension accuracy of both letters and words where participants were able to recognise them (both letters and words) with an accuracy of 97%.

Chapter 4 also presents two other user studies investigating the interactions of skin reading with other activities. The first of one (user study 7 in Section 4.3) evidences that vibrotactile encoded symbols can be perceived very accurately also while performing other tasks without affecting the performance of the other concurrent tasks. Such results emphasise the potential of skin reading in multimodal interactions and multitasking scenarios. Whereas the later user study (user study 8 in Section 4.4) evidences that passive haptic learning (**PHL**) could be an alternative for training users for skin reading. Although the learning effect seems to be not at the same level as the active learning where users focus actively on the training, PHL allows users to be trained while performing other activities such as playing video games which might make the training process more appealing for some users.

Chapter 5 deals with encoding of continuous numerical values for inaccuracy tolerant applications. It provides progressive bar inspired layouts and intuitive information encoding which require no training and have a small number of actuators as a tradeoff for a level of inaccuracy in the comprehension of encoded information. It proposes four vibrotactile wearable layouts in a straight form or a circular one and also a phantom based encoding of information to encode continuous numerical values. Furthermore, it proposes to extend the existing phantom based perceptual models by incorporating the sensitivity of the stimulated location into the model. To infer the relative sensitivity of locations, it proposes a data-driven optimisation method which requires a short calibration. The entire process of calibration and optimisation results in personalised user sensitivity adjusted models able to predict users' perception more accurately. Thus such sensitivity adjusted perceptual models are able to encode information more precisely than the other generic existing models. Overall, Chapter 5 shows that the designed layouts, the proposed stimulation method and the sensitivity adjusted models could be used to encode continuous numerical values that can be decoded with a relatively tolerable inaccuracy ($5\% - 7\%$) for many applications.

Lastly, Chapter 6 addresses the interaction aspect with a vibrotactile wearable display. It proposes to incorporate interactions into the skin reading process to allow the user to control the flow of information. Initially, an interaction concept is introduced and evaluated using a user study. The results of the study show that only higher level unit interactions such as word base interactions are necessary whereas the interactions with letters are neither utilised nor desired by users. In addition, the user study reveals, that users prefer hand gesture interactions as an interaction modality. As a result, Chapter 6 also proposes hand gestures for such interactions and a wearable sensor-based solution for recognising such gestures.

Overall this thesis uses two main approaches to increase the performance of comprehension when conveying information and such approaches are leading contributors to the success of conveying the textual information as well as in the conveying of continuous numerical information. The first approach is to leverage the sensitivity of the location when stimulating patterns. In the case of conveying discrete symbols (for textual information in skin reading), the OST patterns are used where the stimulation is prioritised based on the sensitivity of the location. In the case of continuous numerical values, again, the sensitivity of the location is leveraged to build sensitivity adjusted perceptual models which can be used to predict user's perception better and therefore more accurately encode the information. The second approach is the use of data-driven optimisation methods to optimise and refine the encoding of information. For encoding discrete symbols, the encoding is optimised by using the bigram probability distribution for a given language in order to minimise the occurrences of two subsequent letters (in a word) that share a vibromotor. This is achieved by defining a cost function which measures such sharing of vibromotors for any given encoding in a language and then constructing an algorithm which minimises such a cost function. Similarly, when encoding continuous numerical values, a data-driven optimisation is proposed, that minimises a defined cost function to drive the relative sensitivities of the stimulated locations. In both cases, such data-driven optimisations significantly improve the comprehension accuracy of the encoded information.

7.1 Research Questions

At the beginning of this thesis (see Section 1.1), four main research questions were defined, which then were addressed using ten user studies and described in four chapters of this thesis. Note that as for the first research question, four user studies were conducted (see Chapter 3), there were several research sub-questions that those studies answered. However, they were all related and centred around the first research question of this thesis as they all investigated different details and granularities of vibrotactile patterns and layouts of wearable vibrotactile displays. Similarly, for the second research question of this thesis, four user studies were conducted which answered other sub-questions centred around the second research question of this thesis. This section will discuss those four main research questions and elaborate on how they were addressed and answered but without focusing on the sub-questions of each study. Note, for the sake of completeness and readability, some of the clarifications included in this section overlap with a description of the research questions in the Introduction (in Section 1.1) of this thesis.

7.1.1 RQ1: Constructing vibrotactile Patterns Optimised for Throughput and Perception

Do the overlapping spatiotemporal patterns result in better identification accuracy than the baseline spatial patterns on the hands and forearms?

This research question takes the challenge of finding vibrotactile patterns which are optimised to be short so that they can allow high throughput when combined in complex messages and at the same time are highly distinguishable so that they can be identified correctly. This problem is addressed in Chapter 3 where it proposes the overlapping spatiotemporal (OST) patterns as a good option for maximising throughput and accuracy. Four user studies investigate the accuracy of such vibrotactile patterns compared to spatial patterns and also investigate suitable body positions for such vibrotactile patterns.

The first user study (in Section 3.1) proposes and evaluates overlapping spatiotemporal (OST) in comparison with spatial patterns. In addition, it proposes and investigates three wearable vibrotactile layouts for perceiving such patterns. The study is able to identify two of the proposed wearable layouts as being suitable for the perception of such patterns and also revealed that OST patterns are perceived more accurretely than spatial patterns. The third user study (in Section 3.3) reveals that prioritising the activation of vibromotors based on the sensitivity of locus increases the identification accuracy of OST patterns significantly. Moreover, the fourth study (in Section 3.3) extends the hand based layout by adding more actuators and then investigates the perception of OST patterns in the proposed new layout. The end result of this user study is a layout with eight suitable locations of actuators which are suitable for perceiving OST patterns accurately.

To directly approach the research question, overall the studies presented in Chapter 3 provide evidence that indeed overlapping spatiotemporal patterns result in a better identification accuracy than the baseline spatial patterns on the hands and forearms and the identification is further improved by prioritising the activation of vibromotors based on the sensitivity of locus.

7.1.2 RQ2: SkinReading - Conveying Natural Messages

Are overlapping spatiotemporal patterns suitable for vibrotactile skin reading on the hands and forearms? More specifically, what performance on the recognition of let-

ters and words can participants achieve with few hours of training?

The second research question investigates the feasibility of using such OST vibrotactile pattern first to encode discrete symbols representing the letters of English Alphabet and then to investigate whether the combination of such vibrotactile symbols can be used to form more complex messages such as words and phrases. This thesis first proposes an encoding of the English Alphabet to OST vibrotactile patterns. In four user studies presented in Chapter 4, users are trained to recognise the vibrotactile alphabet. Then, the studies investigate how well users can recognise symbols encoded by OST patterns, and also how well users can read words when such symbols are combined into words. Moreover, they identify problems related to the encoding of the symbols and construction of patterns and reiterate the entire process in order to maximise the recognition accuracy.

The first user study (in Section 4.1) proposes an encoding for all the letters of English Alphabet which maps every letter to an OST pattern. The encoding uses the frequency letters in the English language to provide an efficient encoding scheme. In addition, the study proposes a training program which is used to teach participants the proposed encoding. The study evaluates the performance of participants on the recognition of letters and words after training using wearable vibrotactile displays on the hands and forearms. Results show that participants are able to comprehend the information with relatively high accuracy but also it reveals that there are potential improvements related to layout and encoding. Thus, the second user study (in Section 4.2) proposes a two-step optimisation process which optimises the layout and encoding. Furthermore, an evaluation of the impact of such optimisations shows a drastic improvement in the comprehension of letters and words. Moreover, this study investigates the knowledge decay of encoding over time as well as the transferability of encoding knowledge on the untrained body location.

The aforementioned user studies clearly answer the research question. The first user study (in Section 4.1) shows that indeed OST patterns are suitable for vibrotactile skin reading on the hands and forearms using wearable vibrotactile displays with six vibromotors. Although, the results are very promising and the achieved comprehension accuracies (of letters and words) stand out compared to the state of the art (see Figure 2.5 and Table 2.2), this study showed that the layout needs to be extended with at least one more vibromotor to encode the entire English Alphabet as the OST patterns should not use more the two vibromotors. It also pointed out that the encoding needs to be optimised when conveying words as a series of letters. Using the extended layout and the optimised encoding in the next user study (in Section 4.2), participants achieved an astonishing 97% accuracy in both letter and word recognition.

7.1.3 RQ3: Conveying Inaccuracy Tolerant Quantitative Values through Wearable Vibrotactile Displays

For scenarios where high precision is not required, can we encode continuous values using a discrete number of actuators using phantom sensation? More precisely, how well (with what accuracy) can users decode such encoded values and does sensitivity adjustment increase such encoding/decoding accuracy?

The third research question moves away from the discrete symbols and focuses on use cases where continuous numerical values need to be presented to the user. Unlike in discrete tactons where mistaking of two neighbouring values (e.g. A for B) is considered to be a high error when dealing with numerical values the magnitude of error is very important. For instance, mistaking the value of 67% for 65% or 70% might not be a drastic problem, but mistaking it for 15%, might be. Thus this research question targets use cases which are tolerant to a degree of inaccuracy.

In Chapter 5 a concept is proposed where phantom sensation is used to convey such continuous numerical values using progress-bar inspired wearable vibrotactile displays organised in chains of vibromotors. In addition, novel sensitivity adjusted perceptual models are proposed to better predict users' perception. To provide an answer to our research question, a user study (see Section 5.3) investigates whether such method can be used to convey continuous numerical values and the impact of the sensitivity adjusted perceptual models. The results show that not only the phantom sensation could be used for encoding continuous numerical values, but the proposed sensitivity adjusted models can significantly increase the accuracy of comprehension. Using circular displays with four vibromotors, participants were able to decode values with an average of error of 4.4% using wrist layout and 5.1% using upper arm layout. When using straight displays with three vibromotors, participants decoded values with an average error of 6.4% on the dorsal (back) part and 6.8% on the ventral (front) part of the forearm.

7.1.4 RQ4: Interactions for Skin Reading

What interactions are necessary for skin reading? What is the preferred modality for such interactions?

When perceiving information, there is a need for repeating or navigating parts of it by controlling its flow. Users might need to repeat certain words from time to time as a result of attention breaks or simply due to misperception. In both visual and Braille reading such interactions occur very often even if readers are not aware of them as they occur unconsciously and naturally. However, in vibrotactile skin reading, the user does not control what information is currently being conveyed, and thus interactions need to be provided to enable such control.

Chapter 6 investigates what interactions are necessary and what is the preferred modality of such interactions. A study reveals that only higher unit interactions such as word interactions (previous, next and repeating the current word) are necessary and users would prefer hand-based gesture interaction for such interactions.

7.2 Contributions

Overall, this thesis provides several contributions relevant to several communities including human-computer interaction, wearable computing, and psychophysics. The contributions lie in the creation of wearable haptic prototypes, stimulation methods, information encoding, user training and methods for interacting with such wearable devices, which have been described in details in chapters 3 to 6. This section highlights the contributions and elaborates their relevance.

7.2.1 Vibrotactile Patterns

This thesis (see Chapter 3) proposes overlapping spatiotemporal patterns which are shorter than sequential temporal patterns and can be identified more accurately than spatial patterns. Moreover, it proposes to prioritise the activation of vibromotors based on spatial acuity to maximise the perception and identification accuracy. Such patterns are crucial building blocks of the skin reading and one of the major contributors to its success using the proposed methods in this thesis. Having patterns the are discriminable in a very short duration, makes it possible to have complex messages conveyed in a short duration, which is very often the major precondition for potential applications of skin reading as the information needs to be delivered precisely and as fast as possible to be applicable.

Although OST patterns are created with skin reading in mind, they do not necessarily have to be combined to form more complex messages. They could be used as stand-alone where they would represent complete messages such as states, warnings, errors, numbers, etc...

7.2.2 Wearable Vibrotactile Display Designs

While throughout the thesis, several layouts have been investigated for skin reading on the back of the hands and forearms, it was revealed that at least seven vibromotors are necessary to encode the entire English Alphabet. So a wearable hand based layout with eight vibromotors on the back of the hand was proposed in Section 3.4, where its locations have been tested to be suitable in terms of perceiving OST patterns. The wearable layout has been designed to be packed as a wearable glove where the fingertips could be left uncovered not to hinder the interaction with everyday objects. A subset with seven vibromotors is tested in Section 4.2 and it evidenced that it is very adequate for skin reading. Although in Section 4.2 only seven out of eight vibromotors have been used as they were enough to encode the English Alphabet, the layout of eight vibromotors could be used to encode up to 36 symbols, which could be used either to encode the alphabet and other symbols (e.g. digits) or to encode larger alphabets of other languages.

Additionally, four other layouts on the forearm and upper arm are proposed for using them to encode continuous numerical values using phantom sensation (in Section 5.3). Such layouts are able to encode continuous values of directional or circular nature. The circular layouts have been designed to be packed in the form of vibrotactile bracelet on the wrist or upper arms whereas the directional ones can be packed inside a wearable forearms sleeve.

7.2.3 Methods for Conveying and Encoding Letters of English Alphabet and Textual Information

When conveying information through means of vibrotactile displays, besides having perceivable patterns, it is required to have an encoding which maps the vibrotactile patterns to the encoded information. Initially, Section 3.1 proposes a frequency based encoding of the English Alphabet which follows a simple principle: the more frequent a letter appears in a language, the fewer vibromotors are used to encode it. In principle, this is a very appropriate encoding when considering only symbols/letters as the OST patterns with more vibromotors take longer to stimulate and also the more vibromotors are used (specifically if more than 2), the higher is the chance of masking one location. Thus such an encoding contributes to faster transmission of information in general and also fewer comprehension errors on average.

However, as the Section 4.1 reveals, such an encoding is not ideal for combining symbols in series as when subsequent letters share a vibromotor their patterns are more prone to masking. Thus Section 4.1.6 proposes an optimised encoding which delivers outstanding accuracy in terms of its information (encoded letters and words) being comprehended by users. In addition, this thesis proposes methods for combining letters into words to form textual information.

7.2.4 Methods for Optimising the Encoding for a given Language

In order to avoid systematic errors when conveying words as a series of letters, Section 4.1.6 proposes a method and an algorithm to optimise the encoding of a language based on the bigram frequency in the given language. Moreover, it provides clear steps on how to apply this encoding where it defines a cost function and then it provides an algorithm for minimising it, which in turn results in an optimised encoding. The user study conducted in Section 4.1.6 reveals that such an optimised encoding delivers outstanding accuracy in terms of being comprehended by users.

Even though in Section 4.1.6, the optimisation is applied to encoding of letters of English Alphabet, it has broader applicability. The same methodology can be applied to other forms of encodings where the basic unit is not a letter. For instance, the same methodology can be applied when the basic encoding unit is a phoneme.

Equations 4.3 and 4.4 could be used almost one to one where only one modification is required. The bigram probability distribution ($BF(l_1, l_2)$ in the original equations) would need to be replaced with the probability distribution of bi-phonemes which would represent the probability of two subsequent phonemes appearing after each other in a text. Apart from such minor change, the rest remains the same, including the definition of minimisation function provided by Equation 4.5 and the Algorithm 1 that solves it.

7.2.5 Training Methods for Learning Skin Reading

This thesis investigates and proposes several methods for training skin reading which are a subject of Chapter 4. It first (in Section 4.1) proposes an active training which combines simultaneous visual, auditory and vibrotactile stimulations in order to train participants to associate the vibrotactile patterns with the encoded information. As sections 4.1 and 4.2 show, such a method is very efficient. Then (in Section 4.4) it also investigates using passive haptic learning as a training method for skin reading. It demonstrates that indeed passive haptic training could be used to train users while they perform other activities (e.g. playing games) which might make the whole process of training less tedious. However, the results of the user study presented in Section 4.4 also show that such passive training is less effective compared to active training, especially for some users.

Additionally, Section 4.1 proposes a training program for teaching participants the skin reading. The program is composed of five sessions and uses the first three sessions to teach participants the alphabet where letters are introduced gradually. Whereas the last two sessions serve the purpose of familiarising the participants with faster transmission of information. The program uses repetitive rounds of training and reinforcement testing both of which teach users the encoding. Sections 4.1 and 4.2 demonstrate that such a program is very effective. Although it might be a bit extensive and then perhaps it could be shortened. It remains up to future work to investigate whether the same results could be achieved using a less extensive training program.

7.2.6 Background Perception of Vibrotactile Symbols

Section 4.3 demonstrates that the vibrotactile symbols (encoded by the proposed methods) can be comprehended in the background while performing other primary tasks. Moreover, their performance is not affected by the absence of presence a primary task and vice versa. The performance of the primary task is not affected by the presence or absence of stimuli of background vibrotactile encoded symbols. Such a revelation gives a tremendous perspective to conveying of vibrotactile information in multitasking scenarios.

One limitation of the study presented in Section 4.3 is that it does not evaluate the comprehension of more complex messages such as words and sentences in the background. The comprehension of complex messages would undoubtedly be of the most interest and will be considered in the future work.

7.2.7 Knowledge Transferability of the Encoding in Untrained Body Parts

Section 4.2 investigates and evidences that the knowledge of encoding acquired during the training can be transferred to an untrained body part without any training. In other words, participants can be trained on one hand, and the use the wearable vibrotactile device on the other hand and still be able to recognise the encoded letters without any additional training.

7.2.8 Methods for Conveying Continuous Numerical Values

Chapter 5 proposes a method of using phantom sensation to encode continuous numerical values via wearable vibrotactile displays on the forearms, wrist and upper arm. It proposes to combine a chain of vibromotors and also the phantom sensation to stimulate the perception between two vibromotors to provide an encoding for continuous numerical values.

Such wearable vibrotactile displays target use cases where high precision is not required as clearly the phantom sensation can be controlled only up to a degree. Use cases might include progress or states in games, progress during sports activities (e.g. running, jogging, etc..), grasp force on an object for individuals who lack the sensation (with prosthesis or hand amputee), etc...

7.2.9 Extending Perceptual Models for Phantom Sensation

Chapter 5 extends the existing state of the art perceptual models by proposing sensitivity adjusted perceptual models which are better at estimating the perceived stimuli when applying the phantom effect. The contribution goes beyond proposing models that include sensation as Chapter 5 also provides a data-driven approach for inferring the sensitivities of locations. In turn, this increases the accuracy of comprehension for the encoded value using the phantom sensation.

This approach uses a short calibration process to collect user data which then are used to minimise a defined cost function that in turn reveals the sensitivities of locations for which the models can predict the user perception at best. Such method makes also the encoding personalised to the user and thus promotes the personalisations of vibrotactile wearable devices.

While such models are used to improve the encoding/decoding accuracy, given that they can more accurately (compared to existing models) predict user's perception, they can be utilised for other purposes such as vibrotactile animations which are a typical application of phantom sensation [Israr et al., 2012].

7.2.10 Interaction Techniques for Skin Reading and Gesture Recognition

Wearable vibrotactile displays are unidirectional in the sense that they convey information, and usually, there is no way to control the flow of information. By studying the reading process in visual and Braille reading this thesis realises the need for bidirectional interaction with vibrotactile wearable displays. Thus, Chapter 6 proposes to incorporate interactions for skin reading which would allow the user to control the flow of information and navigate through the text. In addition to proposing and evaluating an interaction model, it proposes a mapping of such interaction to intuitive gestures which then can be recognised using wearable sensors.

One limitation of this interaction technique is that it does not go beyond word based interactions in the sense that it does not provide any means of navigating through more extensive portions of text (e.g. jumping one or more sentences forward or backwards). Although Chapter 6 hints on its benefits and it suggest technical ways to implement such functionality using gesture recognition which is very analogous with the already implemented functionality, the user study does not evaluate its usability aspect. Such an evaluation should be considered in future work.

7.2.11 Wearable Sensory Substitution using Mobile Devices and Skin Reading

This thesis proposes a novel technique for using mobile phones to provide a solution for sensory substitution in combination with wearable vibrotactile devices (see Section 4.5). The solution uses mobile devices to capture the environment which then is processed, and its content is recognised using machine learning models. The recognise content is represented in a textual form and then transmitted to the user through a wearable vibrotactile display. Besides the conceptual work, such concepts are also implemented in the form of mobile applications. Future work needs to address the evaluation of usability with the targeted user groups, which is omitted in this work.

7.2.12 Hand Gesture Recognition System

Given that hand gestures are used for interacting with wearable vibrotactile displays this thesis, the author developed a mechanism for hang gesture recognition based on hand worn sensors. However, the contribution goes beyond recognising a set of small gestures. Throughout the course of this thesis, a generic mechanism and methodology was developed for recognising hand-based gestures using wearable sensors. The methodology is not an essential part of this thesis, and thus its details are not included. However, such work was already published in a peer-reviewed conference paper, and for the details, the reader is referred to the published paper [Luzhnica et al., 2016a].

In summary, the method uses a sliding window to extract features in time and frequency domain which then are used as input to linear discriminative analysis and logistic regression for classifying the gestures. The work used that methodology to successfully classify 31 gestures with very high accuracy (98.5%). The contribution here lies on methodology and mainly on data processing and the pipeline of used algorithms to build the gesture recognition system.

7.3 Reflection and Future Opportunities

The work in this thesis was driven by several user studies from which it was able to design and evaluate methods for skin reading as well as various phenomena related to it. Its findings and revelations present major contributions in vibrotactile communication. However, reflections upon the findings of the presented work and its limitations lead to discussions as well as open up promising directions for future work, which are explored in this section.

7.3.1 Longitudinal Studies

Chapters 3 and 4 propose patterns that deliver high recognition accuracy when stimulated in relatively short duration which then are very suitable for encoding symbols and combine them in words. The user study in Section 4.1 shows that participants achieve an outstanding accuracy of letter and word recognition. Nevertheless, the user studies that evaluate skin reading are limited to a few hours (≈ 5) of training and testing. It would be certainly of most interest to conduct longitudinal studies that observe user's progress through a more extended period of time as several interesting observations would be made.

First, it would be interesting to observe how long users need to be trained to be able to become proficient on reading larger portions of text such as a full story. The ability to read long stories would be very applicable in some particular applications such as in sensory substitution for users with visual or adulatory impairments. Additionally, with more training and more experience, perhaps one could further increase the transmission speed and the messages would still be comprehensible. Perhaps one could find the limit to the shortest duration that OST patterns could encode symbols.

Having proficient participants in reading textual information, would allow us to investigate cognitive processes being reading or more specific word recognition models for skin reading. For instance, in visual reading, users recognise words as units instead of reading them serially letter per letter [Larson, 2004], which is assumed to be a major factor for reading speed that can be achieved [Millar, 2004]. On the other hand, in Braille reading words seem not to be read as units as several investigations suggest [Millar, 2004]. Interestingly, words in vibrotactile Morse Code reading seem to be read as units as indicates [Tan et al., 1997]. It is surprising that such reading process in Morse Code is more similar to visual reading than Braille reading in terms of unit recognition. As for vibrotactile skin reading, one would expect to be similar to vibrotactile Morse Code reading, given the shared similarities. However, a longitudinal user study would make it possible to investigate such a phenomena.

Moreover, having proficient users in text reading who presumably recognise words as units, it would make it possible to investigate the background perception of more complex messages such as words and sentences while performing another primary task as the user study in Section 4.3 was limited to symbols only. Additionally, it would enable to extend and evaluate the interaction model proposed in Chapter 6 by including interaction that navigates larger portions of text as already hinted in Chapter 6.

In addition, longer studies would provide enough time with participants to evaluate the comprehension of information outdoors and in different conditions such as running, talking, driving. For such scenarios, there would be many potential use cases, especially considering the wearable nature of proposed layouts in this thesis.

7.3.2 Sensory Substitution

Section 4.5 proposes a novel technique of sensory substitution using vibrotactile wearable displays and the proposed conveying techniques by combining with speech and object recognition. Such application target users with visual or auditory impairments.

One limitation of this work is that it does not evaluate the usability aspect of such applications with target users and such an evaluation should be considered in the future work. Moreover, future work should also elicit requirements of what information are necessary to convey. For instance, a picture may contain several objects but the target visual impaired users would be perhaps interested only in some particular subset of them (e.g. street names, persons, vehicles, shops, etc..). Similarly, it would not make sense to convey the speech to vibrotactile when the speaking person is far away as it might not be of interest to the target users. Thus filtering uninteresting information would improve the usability of such application. What exactly is to be filtered should be carefully elicited using target users groups.

Moreover, especially in the case of visually impaired users, it might be useful to convey data not form images only but also other materials such as books, text documents, web pages etc... Such documents might contain more complex information such charts, math formulas, reference links etc.. Text also contains punctuations and numbers. All of the aforementioned information should be carefully considered to extend the methods of conveying that is proposed in this thesis and especially the vocabulary of encoded symbols.

7.3.3 Extending the Dictionary

As the previous section hinted, in order to extend the application opportunities of skin reading, it might be necessary to extend the vocabulary of encoded symbols. The hand-based layout proposed in Section 3.4 contains eight vibromotors which are sufficient to encode 36 symbols with OST patterns composed one or two vibromotors. That is enough to encode all letters of English Alphabet and also digits (0-9). However, several alphabets of other languages contain more than 26 letters. The text also contains punctuations. Additionally, for sensory substitution other meta-symbols might be necessary to encode which would indicate the nature of information (e.g. e hyperlink, image, etc...). Thus for more symbols, additional vibromotors need to be added to existing proposed layouts.

One possibility would be to combine the back of the hand with the forearm which would provide enough space to add additional vibromotors. Another possibility would be to use two hands, each with 8 vibromotors. Such two hands layout would be able to encode up to 132 symbols, which would be more sufficient for encoding letters, numerical digits, punctuations, meta-symbols and perhaps some of the most common words, for most of the alphabets.

7.4 Summary

This chapter summarised the contributions of this thesis and emphasised the findings revealed through its chapters. It also discussed further research ideas that have been crystallised while working on the contributions of this thesis with the goal of providing research directions for future work. I humbly thank the reader for reaching the end of this thesis.

Appendix A

Gesture Recognition System

In the following the paper [Luzhnica et al., 2016a] on methodology of gesture recognition based on wearable sensors will be included. Note that as this was considered out of the scope of this thesis, its material was not included in the thesis. However for references, it will be included in here. Note that the paper is included in the original version that is published and thus the font, formatting will be different of that used in this thesis. Additionally, the paper also contains the original page numbers from the original proceedings [Luzhnica et al., 2016a].

A Sliding Window Approach to Natural Hand Gesture Recognition using a Custom Data Glove

Granit Luzhnica* Know-Center GmbH

Jörg Simon† Know-Center GmbH

Elisabeth Lex‡ Knowledge Technologies Institute, Graz Univ. of Technology & Know-Center GmbH

Viktoria Pammer§ Knowledge Technologies Institute, Graz Univ. of Technology & Know-Center GmbH

Abstract

This paper explores the recognition of hand gestures based on a data glove equipped with motion, bending and pressure sensors. We selected 31 natural and interaction-oriented hand gestures that can be adopted for general-purpose control of and communication with computing systems. The data glove is custom-built, and contains 13 bend sensors, 7 motion sensors, 5 pressure sensors and a magnetometer. We present the data collection experiment, as well as the design, selection and evaluation of a classification algorithm. As we use a sliding window approach to data processing, our algorithm is suitable for stream data processing. Algorithm selection and feature engineering resulted in a combination of linear discriminant analysis and logistic regression with which we achieve an accuracy of over 98.5% on a continuous data stream scenario. When removing the computationally expensive FFT-based features, we still achieve an accuracy of 98.2%.

Index Terms: C.3 [Special-Purpose and Application-Based Systems]: Signal processing systems I.5.2 [Design Methodology]: Classifier design and evaluation I.5.2 [Design Methodology]: Feature evaluation and selection I.5.2 [Design Methodology]: Pattern analysis I.5.4 [Applications]: Signal processing H.5.2 [User Interfaces]: Input devices and strategies H.5.2 [User Interfaces]: Interaction styles

1 Introduction

Gesture recognition has been an active field of research for more than two decades in human computer interaction. Initially, the motivation was to detect and recognise sign language [1, 14, 33, 41]. The goal mostly was to develop computing systems that could understand and translate sign language. More recently, gesture recognition has gained interest as basis for gesture based interaction in a wide range of use cases, such as crisis management [45], TV remote controlling [34], interacting with computer [18, 24], gaming interfaces [23, 26, 45, 52], augmented reality applications [17, 43, 48, 50], hands-free interaction in car driving [27], providing virtual training for car driving [50] or detecting a driver's fatigue [25]. In the medical area, robot nurses are envisioned to detect surgeon's hand gestures and to assist with necessary surgical instrument [45]. In another type of use case, computer systems detect gestures in order to understand user activities. For instance, robots have been envisioned to analyse gestures in order to track which tasks are already completed in order to be able to seamlessly take over with the next steps [6, 35]. Sometimes, it is useful to only observe and document the gestures, as in the case of assembly lines to document the work for quality assurance [40]. More in general, the goal to detect assembly line tasks is an area of active research [20, 32, 46, 51]. Gesture recognition has also been explored in the context of logging activities of daily life: In [38], the authors explore the possibility to detect eating habits via recognising the gestures for eating and drinking (bringing the hand to the mouth). In [39], activity logging based on both smartwatch and smartphone sensing is used to detect drinking too much coffee or not eating.

With this work we contribute to the field of gesture recognition by exploring the recognition of natural and interaction-oriented hand gestures based on sensors worn on users' hands. To that end, we designed a custom data glove equipped with sensors that capture both motion and state of the hand and fingers. We concentrated on gestures that are widely known and that can reasonably be adopted to control and communicate with computing systems. Our envisioned use case is that of mapping out a general-purpose gesture alphabet. It should be easy to learn for users, and should be able to replace some of the interactions with computing systems (selecting, browsing, etc.) that are currently performed via mouse or smartphone gestures.

We approached this goal by conducting a data collection experiment in which multiple users performed such gestures. In parallel to sensing, the gestures were manually annotated with gesture names. This resulted in a labelled set of hand gestures, which we used to extract representative features and to train a supervised learning algorithm. Then, we evaluated the performance of our algorithm "online", i.e. on a continuous sensor data stream. The contributions of this work are three-fold:

- A data set of natural hand gestures, which were gathered in a data collection experiment with 18 adults, and are manually annotated with gesture names.
- Features selection - We identified characteristic features for gestures and investigated similarities between gestures.
- Algorithm selection - We identified a performant algorithm for classifying gestures in a continuous sensor data stream.

2 Related Work

We identified two strands of research that are relevant for our work: firstly, research that deals with vision based systems for gesture recognition and secondly research that deals with wearable sensors for gesture recognition. In the first case, the gesture recognition relies on an infrastructure built into the environment (e.g., using Kinect or webcam) whereas in the second case, the gesture recognition relies on wearable sensor technologies like data gloves, armbands or smartwatches.

Vision based systems for gesture recognition. Typically a camera that is mounted in the environment records human hands and the system extracts features from the individual frames of the recording. Sometimes there is a filtering process involved which removes unwanted objects like e.g heads from the image or video [7]. Typically, postures are predicted [5, 7] and then a grammar is con-

*e-mail:gluzhnica@know-center.at
†e-mail:jsimon@kow-center.at
‡e-mail: elex@know-center.at
§e-mail: viktoria.pammer@tugraz.at

IEEE Symposium on 3D User Interfaces 2016
19–20 March, Greenville, SC, USA
978-1-5090-0842-1/16/$31.00 ©2016 IEEE

structed to recognise gestures, where a gesture is defined as a sequence of postures [9]. For instance, in [7], the authors first detect and track hands and then recognise ten postures with an accuracy of 96% in camera images with a multi-class SVM. Similarly, in [5] a single web cam is used as source, from which the authors extract Haar-like features and use AdaBoost to discriminate between four postures: two finger, palm, fist, little finger. The authors achieve an accuracy of over 90%. In [1], the authors achieve an accuracy of 99% by using PCA and a Euclidian distance based classifier to recognise 25 international hand alphabet postures from images of the gestures. An alternative vision based approach is to use coloured gloves in which different parts of the hand are marked with different colours, making it much easier to track gestures [12].

Wearable sensor based systems for gesture recognition. The majority of wearable sensor systems for gesture detection are gloves equipped with sensors. In most research endeavours, gloves are custom-built. In [29], a recurrent neural network (RNN) is used to recognise the following Japanese sign language gestures: father, mother, brother, sister, memorise, forget, like, hate, skilled and unskilled. The gestures are constructed using 42 previously recognised postures representing Japanese letter alphabet with an accuracy of 96%. The data were generated using VPLDataGlove. In [50], the authors use a feed-forward neural network capable of distinguishing between 15 gestures with an accuracy of 98%. Data were recorded with a CyberGlove with 18 sensors. More recently, a feed-forward neural network was used to construct a hand gesture recognition system for interacting with robots [30] . Using data from CyberGlove II (providing 22 joint-angle measurements), the authors were able to recognise 10 different artificial gestures with an accuracy of 99.8% and 30 gestures with an accuracy of 96.3%. In this work, the authors initially performed segmentation by identifying whether each of the data readings belongs to a gesture. These segments were used as input to the machine learning algorithm. In [52], data from EMG and a wrist-worn accelerometer were used to build a system that recognises 18 gestures with an accuracy of 91.7%. The defined gestures were used to play a virtual rubic's cube game. In [36], authors used a list of 22 natural hand gestures. The gestures originally stem from [10]. While analysing the data, the authors first resampled and interpolated the data. They then used LDA to discriminate between the resampled segments with an accuracy of 92.8%. This shows that there is a clear separability between those gestures. However, in a live recognition system, such resampling and interpolation is not possible unless the start and end time of the gesture that needs to be recognised can be detected.

Recently, there has been a growing research interest in gesture recognition based on consumer good technology like smart watches (e.g., [11, 44, 49, 53]). In [49], the authors report the classification of 37 interaction oriented gestures, i.e. gestures that are intended to be used for controlling other devices (turning the arm, simulating a click, pinch to zoom, etc..). The gestures are detected based on smartwatch sensor data only, and with an accuracy of 98% by using Naive Bayes algorithm. The authors report different numbers in a subsequent demo paper, namely 27 gestures with 96% accuracy using Logistic Regression or Decision Trees [53]. However, the data in the latter paper were collected only by a single participant; and the participant performed gestures from a fixed arm position.

When comparing two approaches, vision based systems are more sensitive to the environment. Lighting conditions, scene and background details are issues that affect such systems [45]. In the case of cameras, there might be also privacy issues; and different countries have different regulations concerning video recordings in non-private environments. Wearable sensor based systems, especially glove based ones, can be uncommfortable [45] or even pose a hygienic problem [19]. On the other hand, wearable technologies provide in principle the possibility for higher privacy as the data are a priori more anonymous than pictures or videos. When it comes to accuracy, many authors report very high accuracies (in the higher 90s) for the selected set of gestures using either technique [1, 7, 30, 50].

In this paper, we take the wearable sensors approach. In contrast to some other works, we emphasise capturing the dynamics of the gestures, and present gesture recognition using a custom data glove. Our work also differs from previous work since we take a sliding window approach in combination with dual labelling in the test set. Sliding window is a technique for data preprocessing in which information is extracted (statistics, aggregates, features, etc...) over a "sliding window" that contains a fixed number of samples. This enables us to use representative features that aggregate sensor data over time. Interestingly, the sliding window approach, to the best of our knowledge, has not been used in gesture recognition system even though it very common approach in activity recognition [21, 22, 31].

3 GESTURE DETECTION SYSTEM

In this section, we describe the design of the overall gesture detection system: What kind of gestures should our system detect, and what kind of information/sensors are required for this purpose?

3.1 Interaction-Oriented and Natural Hand Gestures

As described in the introduction, we are interested in a general-purpose gesture alphabet with which to control computers and communicate with them. Essentially, it would be possible to develop a completely novel gesture language for such a purpose. A study looking at inventing custom gestures [15] showed however, that a user can only remember a very limited number (about two) of such artificial gestures. Therefore, we are looking at gestures that are widely known, even though there may be cultural differences regarding their popularity and meaning. Additionally, there should be a plausible relationship between the gesture and an interaction between human and computer.

These criteria resulted in the following 31 hand gestures (see Table 1). Our gesture set was initially based on the list of 22 natural gestures described in [10]. We added the following: The numbers one to five, as they would be useful to select items; popular touch-based swipe gestures such as swipe left, right and down (up was already on the original list), as these would be useful for navigation. Finally, we added lateral grasp (Grasp 2) and palmar grasp (Grasp 1) gestures, as we think that grasping objects would be useful in interaction with 3D virtual objects. After the first trial, the gesture walk was discarded as it was difficult to perform due the IMU chips on the fingers.

3.2 Custom Data Glove

The above gestures vary a lot in their dynamics: Some gestures contain a lot of complex motions (e.g continue) whereas some are very close to a posture (e.g. numbers one, two, ...).

We planned a data glove that emphasises motion detection of the fingers (which implies that we would have motion sensors on the fingers); as well as hand postures (which implies that we would use bend sensors). The glove is depicted in Figure 1.

We placed two bend sensors on each finger. The upper sensor measures the bending (which translates to angle) of the finger relative to the hand, whereas the lower sensor measures the bending between middle segment and base segment of the finger. Another bend sensor is placed between the thumb and index finger in order to measure the distance/angle between them. Two more bend sensors (in opposite direction) are placed on the wrist in order to be able to measure wrist flexion/extension. Overall, this gives 13 bend sensors. Additionally, each finger tip is equipped with a pressure sensor (5 pressure sensors). Furthermore, 7 IMUs[1] are placed, one

[1]IMU (inertial measurement unit) chip contains a gyroscope and an accelerometer

Gesture	Description
(1) One	Number one by extending index finger
(2) Two	Number two by extending index and middle finger
(3) Three	Number three by extending index, middle and ring finger
(4) Four	Number four by extending all fingers except thumb
(5) Five	Number five by extending all fingers
Thumbs up	Thump stretched pointing up, other fingers form fist
Thumbs down	Thump stretched pointing down, other fingers form fist
Point to self	Pointing at self with thumb
Shoot	Hand in form of a gun and then vibrate
Scissor	Stimulating scissors with two fingers
Cutthroat	Using index finger
Continue	Waving like circular motion with the flat hand
Knocking	Forming a fist and moving the fist up and down
Waving	Shaking the flat hand left and right
Come here	Flat hand with palm upwards: Simultaneous flexing the all fingers but the thumb
Go away	Hand with palm downwards, all fingers but thumb flexed. Simultaneous stretching them
Push away	Flat hand with palm pointing fore wards, then moving the whole hand forward
Never mind	Flat hand with palm pointing left above the head, then moving the whole hand left
Talking	Thumb and 4 fingers pointing forward. Then moving 4 fingers up and down
Calling	Hand is a fist, but thumb and small finger are extended
Walking	Hand is a fist, but making a walking motion with the index and middle finger
Shoulder pat	patting with the open hand on a virtual shoulder
Point	Pointing in front with index finger
Swipe left	Stretched hand with palm pointing left, flexing it completely to the right first, then flexing it to the left, in a circular motion
Swipe right	swiping with palm pointing right, and left to right motion
Swipe up	swiping with palm pointing up, and bottom to top motion
Down	swiping with palm pointing down, and top to bottom motion
Turn	Hand rotation
Zoom	Reverse pinch using index finger and thumb
Grasp 1	Palmar grasp (in the experiment we used a glass)
Grasp 2	Lateral grasp (in the experiment we used a pen)

Table 1: List of 31 interaction-oriented hand gestures.

at the top of each finger, one on the back of the hand and one on the wrist. The wrist IMU is placed exactly at the position where a watch would be. This allows the data recorded with the glove to also be treated as if it came from a smartwatch by simply ignoring the input from other sensors. Finally, a magnetometer is placed on the back of the hand.

At the beginning of our study, various data gloves had already been available commercially. All of them emphasise bend sensors in fingers, and thus focus on hand postures. In contrast, our glove contains both bend and motions sensors (gyroscope + accelerometer) on each finger, thus focussing more on hand motion, i.e. the dynamic aspects of gestures. The MiMu Glove[2] is used to produce music by some means of gesture detection. It employs one IMU at the wrist, 4 bend sensors at the fingers, and vibrators at the underarm to provide haptic feedback. Fifth Dimensional Technologies[3] offers two gloves that are equipped with bend sensors and abduction sensors between fingers. CyberGlove Systems [4] offers CyberGlove II equipped with two bend sensors on each finger, four abduction sensors, sensors measuring thumb crossover, palm arch, wrist flexion and wrist abduction. Virtual Labs[5] offers a range of data glove products (VMG Lite, VMG 10, VMG30, VMG 30 Plus), all of them equipped with bend sensors on the fingers, 9-DOF orientation sensors for hand and wrist, as well as tactile feedback vibrators.

[2]http://mimugloves.com
[3]http://www.5dt.com/
[4]http://www.cyberglovesystems.com
[5]http://www.virtualmotionlabs.com/

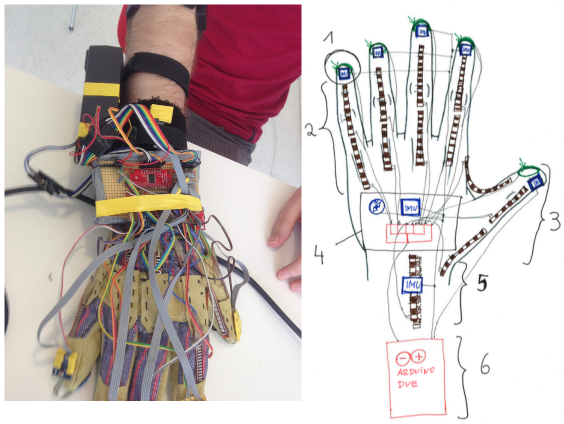

Figure 1: Custom data glove, and scheme of sensor positions. 1: An IMU and a pressure sensor are placed on each finger tip. 2: Two bend sensors cover the two main joints of each finger. 3: The thumb is special as it has 3 bend sensors. 4: An IMU and a Magnetometer is on the top of the hand. Here also an analog multiplexer is mounted to combine all the bend sensors. 5: An IMU is placed on the wrist. Additionally, one bend sensor at the top of the wrist and one at the bottom of the wrist give the angle of the hand to the forearm. 6: All sensors are connected to an Arduino board to collect the data and send it to a computer.

Our custom data glove is a hardware prototype and as such it has some limitations, mainly regarding usability: For long-term wearing, the glove should for instance be made of more comfortable material, be made of smaller and not visible electronic components, should be available in different sizes, and be wirelessly connected to the computing unit.

4 Data Collection Experiment

We collected sensory data annotated with gesture names in the subsequently described data collection experiment.

4.1 Participants

We collected data from 18 healthy adults: 11 males and 7 females. Participants were aged between 24 and 40 years.

4.2 Procedure

Before starting with the data recording, the purpose and procedure of the experiment were explained. Participants were asked to remain seated during the experiment in an office chair. In front of them (on the desk), a monitor was placed. The monitor was used to display the instructions of the experiment. The overall setup is shown in Figure 2.

For each gesture, the following steps were performed in the given sequence:

1. Name of the gesture was shown on the screen (2s)
2. A video was displayed; it showed an actor performing the gesture (without glove; 6s-7s)
3. A counter was shown on the screen alarming the participants that the recording was about to start (3s)

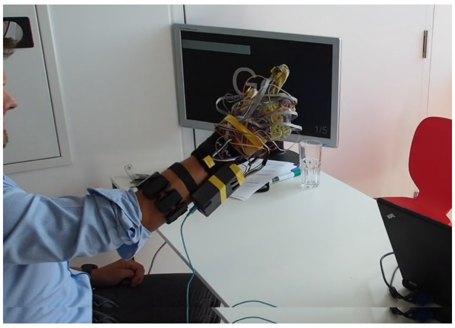

Figure 2: Participant performs "One (1)" gesture while the progress bar is on the screen

4. The participant was asked (audio and text on screen) to start performing the gesture. A progress bar was displayed on screen, indicating the time the participant had to finish the gesture (3*s*). The appearance of the progress bar started the time window called "automatic labelling" (4) in Figure 3.

5. When the participant actually started the gesture, the experiment observer pressed a button on the keyboard. This indicated the start point of the time window called "manual labelling" (5) in Figure 3.

6. When the participant ended the gesture, the experiment observer released the button. This was the end point of the time window called "manual labelling" (6) in Figure 3.

7. When the progress bar ended, the time window called "automatic labelling" (7) in Figure 3 was ended.

The timeline of one gesture is illustrated in Figure 3. Every gesture was performed several times by every participant (5 or 10 times depending on willingness of participant) in a row. The gesture name and the video of the actor performing the gesture (steps 1 and 2) were shown only for the first repetition of the gesture, whereas the counter, progress bar and labelling (steps 3-7) were the same in every repetition.

4.3 Data Annotation

Figure 3 illustrates how the data collection experiment procedure relates to the continuous sensor signals that we recorded. Here we comment on two things regarding data annotation: Firstly, we used *manual labels* as the ground truth, i.e. the labels we refer to in training and testing. *Automatic labels* are used to perform sanity check on *manual labels* e.g. sometimes it happened that the experimenter forgot to label a gesture. We discarded such data. Secondly, when sliding windows are moved over the continuous sensor signals, then there are windows with no gesture in it (window 1 in Figure 3), with partial gestures in it (windows 1 and 5 in Figure 3), and windows with full gestures in it (windows 2 and 3 in Figure 3). For algorithm design (Section 5) only windows with no or full gestures were used, while for evaluation of the selected algorithm in realistic settings, the algorithm was also evaluated on windows that contain a partial gesture (Section 6).

5 Algorithm Design

In this section, we describe the process of selecting the best performing supervised learning algorithm, and the optimum configuration. By configuration we mean selecting parameters for window slicing and parameters for spectral components of the window. Overall, we prepared a list of all possible configurations and

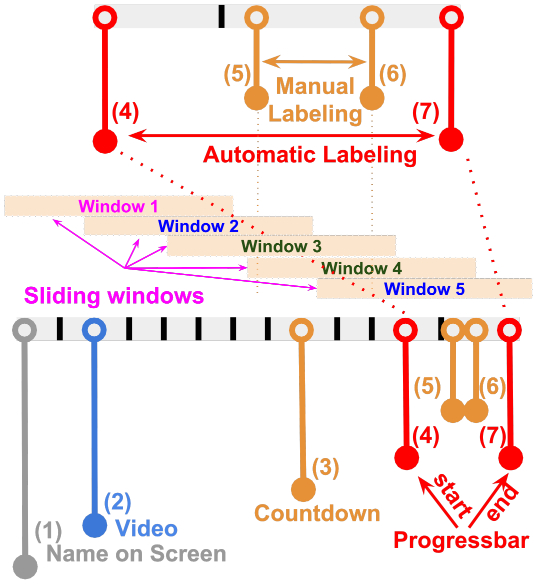

Figure 3: Experiment timeline for a single repetition and sliding windows construction

cross-validated each of them against the set of all chosen learning algorithms. Finally, in order to choose the best configuration, we averaged the cross validation results over all algorithms and considered configurations with the highest average scores. The details of each step are given below.

5.1 Data Pre-processing

Since in the accelerometer readings the gravity component is present and we only need to know the real acceleration value, we first removed the gravity component from the data readings. The gravity component was removed using a complementary filter [13], which typically gives satisfying results and is computationally less expensive than a Kalman Filter [16]. In addition, in order to get independent axis accelerator and gyroscope values, we computed the norm ($norm = \sqrt{x^2 + y^2 + z^2}$) of accelerometer and gyroscope vector for each of our IMUs. Magnetometer values were discarded as their values provide information related to the absolute location of the hand whereas gesture recognition should work regardless of the hand location. Furthermore, all data dimensions are normalised with zero mean and a standard deviation of 1.

5.2 Window Length and Step-size

As basic unit for classification we use sliding windows, i.e. data windows of fixed sample size that constitute snapshots of the continuous data stream. Features are computed per window. Sliding windows are a well-established method of feature extraction used in many domains (speech to text [37], activity recognition [21, 22, 31], etc..). Their advantage is that the extracted features can be used with almost any algorithm [8]. They typically have two configuration parameters: size and step. For parameter selection, we cross validated the data with several window sizes (140, 160, 180, 200 samples, where 1 second contains 85-87 samples).

As for the labels, we consider one window to have a gesture label only if the window contains the whole gesture. Otherwise we label it as idle class. We used steps of 20, 30, 40 and 50 samples and again used cross validation to select a value for this parameter. The details of cross validation are given in Section 5.4.

5.3 Feature Engineering

The recorded data set contains the following dimensions: (x, y, z, *norm*) values of gyroscope, (x, y, z, *norm*) values of accelerometer, values of pressure sensors, values of bending sensors. For each data dimension we used the following descriptive statistic as features: minimum, maximum, range, average, standard deviation and signal energy from the sliding windows. Minimum and maximum values of the bend sensors should contribute to capturing the static part (posture) of the gesture. On the other hand, the derived values from the norm value of the gyroscope and the accelerometer should capture orientation independent motions aspects, as the norm is just the intensity of the accelerometer or gyroscope in any direction.

For the gyroscope and accelerometer values, we also used the spectrum features, namely the amplitude of Fast Fourier Transform (FFT) coefficients for the signal in the given window. For FFT, one has to decide which and how many coefficients are used. Typically, either the n first or the n largest (by amplitude) coefficients are used [28]. Figure 4 shows that in our case, on average, the amplitudes of FFT coefficients decrease monotonically. This means that the first coefficients are the largest ones, which in turn means that the lower frequencies are dominant. Therefore, using the amplitudes of the first n coefficients is a good way to proceed. For selecting a suitable number of first FFT coefficients, we use again cross-validation to choose amongst the following options: $n \in \{5,6,10,15\}$.

In total we extract 78 statistical features from bending sensors, 30 statistical features from pressure sensors, 336 statistical features and $n \times 56$ FFT features for IMUs, where n is the number of first FFF components used for the window. It is worth pointing out that the majority of the features come from motion sensors (IMUs).

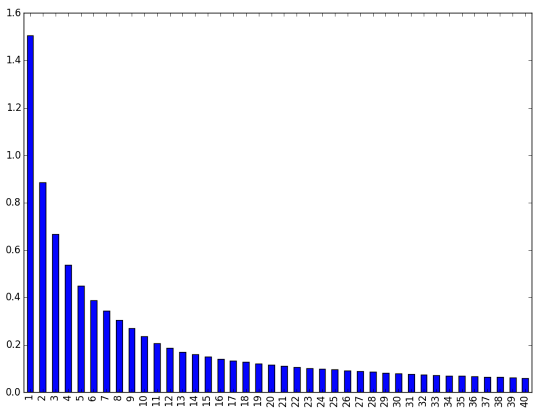

Figure 4: Average amplitudes (over all signals) of the first FFT coefficients (excluding the zeroth) for all 200 frame windows

To avoid correlated features, we calculated correlations between features and automatically removed the features that highly correlate with each other (with absolute Pearson correlation index more or equal to 0.99). Finally, extracted features are normalised with zero mean and a standard deviation of 1.

5.4 Procedure for Algorithm and Parameter Selection

In this section, we describe our procedure to select the best performing algorithm and parameters such as length and step-size for sliding windows, and the number of FFT coefficients to be used as features. We evaluated window sizes of $140, 160, 180, 200$, and window step-sizes of $20, 30, 40, 50$, as well as a number of FFT coefficients of $5, 6, 10, 15$. The choice of these options is discussed in Sections 5.2 and 5.3.

As previously mentioned, similar to activity recognition solutions, we emphasise the motion sensors and follow an approach (sliding windows) that is frequently used in activity detection. Therefore, we chose classification algorithms that have proven to provide robust performance on activity recognition using wearable sensors [21, 22, 46, 47], namely:

- K Nearest Neighbours (KNN)
- Linear Discriminant Analysis Classifier (LDAC)
- Support Vector Machines (SVM) with a linear kernel
- Logistic Regression (LR)

According to the survey presented in [3], discriminative classification algorithms are very effective in identifying features that mostly contribute to discriminations between activities using wearable sensors. Therefore, our discriminative classification algorithms (in our case SVM and LR) should work very well in case the gesture is well captured by extracted features of windows. LDAC is suitable when a linear transformation (LDA) of the data yields in linearly separable classes (in the transformed space). On the other hand, KNN uses the notion of distance in feature space and it can perform good even when linear separability is not possible.

Considering the large number of features we have (724 when using 5 FFT components, 1284 when using 15 of them), we were concerned about overfitting. Therefore, we employed dimensionality reduction techniques prior to training. We used Principal Component Analysis (PCA) which applies an orthogonal linear transformation of the data, in an unsupervised manner, resulting in a maximised variance of data in the transformed space. On the other hand, we also used the supervised linear transformations, namely Linear Discriminant Analysis (LDA) and also it's state-of-the-art alternative Spectral Regression Discriminant Analysis (SRDA) [4]. The latter methods utilise class labels for minimising the within-class variance and maximising between-class variance (in the transformed space). An extensive analysis on how the used algorithms and dimensionality reduction work, including the mathematics behind it can be found in [2]. In our cross-validation process, the classifiers have been trained in both ways: without any dimensionality reduction and with prior dimensionality reduction transformations.

For each window and step-size, we prepared the dataset as follows: We selected only those windows from the complete dataset that are "unambiguous windows", i.e. windows that contain either no gesture, or a full gesture (see Figure 3, where window 1 contains no gesture, windows 2 and 5 contain a partial gesture and are not part of the training data set, and windows 3 and 4 contain a full gesture). The rationale was that the classifier should only learn the full gestures, not parts of gestures.

Moreover, the classes in our data are unbalanced as the majority of the windows are labeled as "idle class" (no gesture). Balancing was achieved by the following procedure: First we calculate the average number of windows per gesture which we will denote by k. Then we removed, before splitting to train and test set, all idle class windows except k random number of idle class windows from the data set. It is worth mentioning that, the balancing was used only during the algorithm and parameter selection process but not during the algorithm evaluation process (Section 6).

The respective training data set was then 80% randomly chosen windows, and the test data set the remaining 20%. We used the following procedure to select the winner combination of configuration (parameters) and algorithm:

First, we compute the performance of each combination "configuration/algorithm" by 5-fold cross validation over the training data set.

Then, for each configuration we compute the average performance over all algorithms to select the winner configuration. The winner algorithm would then be the best-performing algorithm for this configuration. The rationale for this procedure was that we wanted to have the configuration to be as robust as possible in relationship to an algorithm in order to avoid overfitting, i.e. we did not want to select a configuration that only works for a single algorithm.

5.5 Algorithm and Parameter Selection Results

The procedure of selecting and parametrising a classification algorithm that we described above in Section 5.4 yielded the following: The best configuration is the one with a window length of 200 frames, step-size of 20 and 15 FFT coefficients with an average cross-validation (across all compared algorithms) score of 95.6%. Another configuration with less computationally intensive parameters, namely window length of 200 frames, window steps for the sliding windows of 50 and only 5 FFT coefficients had an average cross validation score of 95.3%. Considering that the score difference is minimal whereas the computation efficiency is higher, we selected the latter configuration. For this configuration, the best performing algorithm was LDA+LR with an cross-validation f_1 score of 99.8%. Here, initially LDA was used to perform dimensionality reduction to 32 components and then a logistic regression algorithm was trained and tested on the dimensionally reduced data.

6 Algorithm Evaluation

In this section, we report on accuracies on the full dataset for the selected algorithm and configuration, which constitutes a realistic scenario of continuous data stream analysis.

6.1 Algorithm Performance on Continuous Sensor Data

As realistic algorithm performance, we consider its performance on the following dataset: All windows are used in the test data set, which includes those with a partial gesture windows. A partial gesture window is the one that contains only a portion of a gesture (see window 2 and 3 in Figure 3). Moreover, there is no balancing (neither in the training nor test data set) but class weighting is used when training in order to prevent bias towards the larger classes. This corresponds to the data that would be available in a real world continuous sensor data stream. For windows that contain a partial gesture, we assume the algorithm prediction is correct when the classification outcome is either the idle class or the correct gesture class that is partially in the window. We refer to this strategy as dual labelling in the test set.

On this test set, the LDA+LR algorithm with a window size of 200, a step size of 50, and with only the 5 first FFT gestures in the feature set, performs with an 98.5% f_1 score. The confusion matrix is given in Table 2 and the receiver operating characteristic (ROC) curve is visualised in Figure 5. Here, from 9581 windows, 9440 were classified correctly. From the correctly classified windows, 1618 contained full gestures, 2802 partial gestures and 5020 contained no gesture at all (belonged to the idle class). On the other hand, 141 windows were misclassified from which 6 contained full gestures, 106 partial gestures and 29 came from the idle class.

6.2 Algorithm Performance without FFT Components

Removing FFT calculations during the gesture extraction can speed up the processing the data stream. The rationale details for such an optimisation is discussed in Section 7.2 below. Removing the FFT components from all accelerometer and gyroscope dimensions results in a recognition f_1 score (when considering dual labelling of the ambiguous windows in test set) of 98.2%.

7 Discussion

As our results reveal, we achieve a high classification accuracy in general. As presented in Figure 5, the prediction confidence is also high. It is important to stress that our results were achieved by including in the test set the windows that contain partial gestures. In a live gesture recognition system, there is no way of excluding them. More specifically, in a live scenario, we need to get a sliding window over a stream of data, as visualised in Figure 3, and since we don't know when a gesture starts and when it ends, we can't know beforehand whether a partial gesture is in a window. In the test set, we used the dual labelling strategy which delivers an accuracy of 98.5% (see Table2 and Figure 5). We argue that dual labelling is acceptable for a live system: In the end, it is just a matter of how fast the recognition system realises that the gesture is being performed. In case it predicts the correct gesture class (the one that is partially contained in window), then we can recognise the gesture even before it is completed. Otherwise, if the classifier predicts it to be a idle class window, then the next window (or previous one) will be a window with the full gesture in it and will be classified correctly.

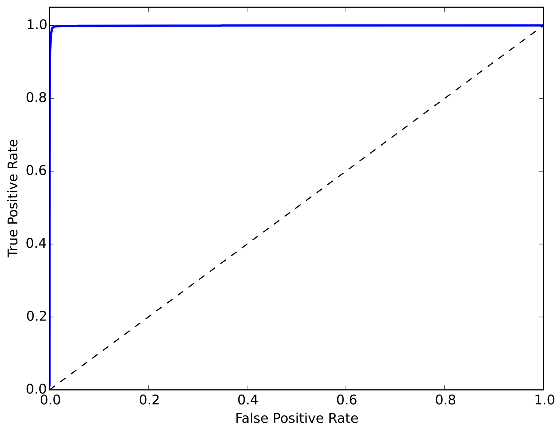

Figure 5: ROC curve for weighted average of all classes

7.1 Analysis of Confusions between Classes

Although overall performance is really good, it is not perfect: some windows are misclassified. In this section, we analyse the confusions between classes. As presented in Section 6, there are 141 misclassifications out of 9440 windows, of which 106 misclassified windows contain partial gestures. We have only 6 misclassifications from windows that contained full gestures. In the following discussion, we focus on windows that contain only partial gestures on them. Several such windows are being misclassified as another gesture and this phenomenon affects mainly the following 10 classes: Down, Swipe Left, Push Away, Swipe Left, Grasp 2, Point at Self, Swipe Left, Go Away, Grasp 1 and Zoom.

A further investigation shows that, there are some systematic misclassification for such windows. For the **"Down"** gesture, nine misclassifications come from windows that partially contain the "Shoulder pat" gesture, five from ones that partially contain the "Continue" gesture and six from windows that contain partial "Swipe left" gestures. When looking on how the gestures movements are performed, the "Shoulder pat" starts and ends very similarly to the "Down" gesture. So, in case the start or the end of the gesture is missing in a window, it seems quite logical that such a confusion can happen. "Continue" could either start as "Down" or as a "Swipe Left" gesture (depending from person to person). Looking into each of those windows reveals that seven out of nine confusions with "Shoulder pat"contain the beginning (on average 40%) of the of the "Shoulder pat" gesture whereas only two out of

	Nothing	(1) One	(2) Two	(3) Three	(4) Four	(5) Five	Thumbs up	Thumbs down	Point to self	Shoot	Scissor	Cutthroat	Continue	Knocking	Waving	Come here	Go away	Push away	Never mind	Talking	Calling	Walking	Shoulder pat	Point	Swipe left	Swipe right	Swipe up	Down	Turn	Zoom	Grasp 1	Grasp 2
Nothing	6024	2	1	5		4	3	1	7	1	1	4	2		1		6	12	2		2		3	5	15	3	4	28	2	7	6	8
(1) One		107																														
(2) Two			113																													
(3) Three				110																									1			
(4) Four					116																											
(5) Five						97																										
Thumbs up							108																									
Thumbs down								116																								
Point to self									117																							
Shoot										89																						
Scissor											126																					
Cutthroat												110																				
Continue													96												1							
Knocking														123																		
Waving															133																	
Come here																111																
Go away																	93															
Push away																		111														
Never mind																			80													
Talking																				107												
Calling												1									128											
Walking																						110										
Shoulder pat																							113									
Point																								112								
Swipe left																									103							
Swipe right																										95						
Swipe up																											108					
Down	1																								1			107				
Turn																													127			
Zoom																														95		
Grasp 1																															138	1
Grasp 2																																117

Table 2: Confusion matrix of classification using dual labelling in test set. Note that zero values (no misclassification) have been removed from the table for better readability

nine contain the end (on average 66%) of the gesture. However for both cases, the prediction confidence is not particularly high: 66% respectively 46%. Confusions of the "Down" gesture with the gesture "Continue" are predicted with an average confidence is 73% and all of them happen in windows that contain only the beginning (on average 30%) of the "Continue" gesture. Similarly, all five windows that are confused with the "Swipe left" gesture contain only the beginning of the "Swipe left" gesture (on average 29%) and the prediction confidence is 55%. Only one window that contains the end of the gesture (51%) is confused with "Down", with a confidence of 73%.

"Swipe Left" class has nine confusions with windows that partially contain "Continue", all of them containing only the end (35%) of the gesture. Two more confusions happen with windows that partially contain "Never Mind" gesture and all of them contain only the last part of the gesture (70%). Interestingly, those confusions have a relative high prediction confidence (average: 90%). Three more confusions happen with windows that contain only the last part (32%) of the "Waving" gesture with a relatively quite low average prediction confidence (58%).

For the **"Push away"**, four misclassifications come from windows with partial "Continue" gesture. Three of them contain the beginning (44%) of the gesture and are misclassified with an average confidence of 44%, whereas the other one contains the last

Classified	Full/Partial Gesture	Confidence
Correctly	Full	98%
Correctly	Partial	89%
Incorrectly	Full	67%
Incorrectly	Partial	57%

Table 3: Average classification confidence depending on whether the window contains the whole gesture

(73%) part of the gesture is predicted with a confidence of 26%. Two other confusions happen with partial "Never mind" gestures. The both contain only the last part (39%) of the gesture and are predicted with average confidence of 46%.

It therefore seems that the gestures: "Down", "Push away", "Continue" and "Shoulder pat" are very similar either at the beginning or at the end of the gesture. For the other gestures there does not seem to be a systematic misclassification as the confusions are distributed among most of the other classes.

To address the partial window misclassification problem, we investigate whether there is a difference in prediction confidence for such misclassification in comparison to other windows. From Table 3 we can see that the misclassified windows that contain only partial gestures have quite a low average prediction confidence

(57%). One way to minimise the number of misclassifications would therefore be to accept only predictions with higher confidence. According to Table 3, such a setting would mainly affect misclassified windows and especially those with partial gestures on them. This would be acceptable in a live recognition system as when a gesture is performed, there will be window(s) with the whole gesture on it where prediction confidence would be higher. Therefore, the practical classification performance would not be affected since the gesture would still be detected either before or afterwards. The only practical implication is how early a gesture can be recognised.

7.2 Computing efficiency

There are several performance indicators to consider when providing a gesture recognition system for human computer interaction. Besides the accuracy, which is the most important, recognition speed (delays hurt the user experience [45]) and power consumption important. Recognition speed and power consumption directly relate to computational cost. Both are of particular relevance if the gesture recognition is planned to be carried out in a mobile or embedded device, which seems very practical especially for wearable sensor based systems for gesture recognition. the Section 5.4, we mentioned that we chose the sliding window step of 50 and the number of FFT coefficients to be 5, even though the configuration was not the best one in terms of cross-validation score. The rationale behind reducing the number of FFT coefficients was that it reduces the number of features and therefore the computational costs and therefore power consumption. Choosing a step-size of 50 frames instead of 20, enables us to perform predictions less often (only after each 50 frames). On the other hand of course, a larger window step means a larger delay: In our case, 50 frames window size means a delay of 0.625 seconds, which we deemed acceptable for the time being. In the end, this is a trade-off in system design of course.

Removing all FFT components (as described in Section 6.2), on the other hand, results in a more significant reduction of computational costs. During the feature extraction process, descriptive statistical calculations like min, max, average, energy and standard deviation can be calculated with a time complexity of $O(n)$, where n is the number of points (window size). On the contrary, the computation of FFT has a time complexity of $O(nlogn)$ and it is performed for every data dimension d that contains accelerometer or gyroscope values ($d = 56$). In addition, for every such calculation the amplitudes of the first k coefficients need to be extracted with a time complexity of $O(k)$. By having only descriptive statical features that are calculated in $O(n)$, the feature extraction process results in a total complexity of $O(fn)$, where f is the number of features. It is worth mentioning that removing FFT components also results in 280 less features. Moreover, descriptive statistic can easily be computed incrementally for a sliding window in a data stream [42] (see the online algorithm[6]). To our knowledge, there is no way of calculating FFT incrementally in a sliding window.

Removing FFT also reduces the complexity in the classification process, though only by a constant factor. First, LDA performs a matrix multiplication (of dimensions $f \times c$) to transform data into the new space. For a single window (of dimensions $1 \times f$), this transformation has a time complexity of $O(fc)$, where c is the dimension of the new space, which in our case equals the number of classes. Note that $c \leq f$. After this transformation, the multinomial logistic regression is applied for classification. Given the reduced dimensions of the window ($1 \times c$), it has a time complexity of $O(c^2)$ (where c is the number of classes), making the whole classification complexity of $O(fc)$. By removing FFT components we reduce f by 280 which impacts the computational complexity of classification. Note that we do not discuss the time complexity of training process as the training can be done offline and therefore it does not play a role in the live system performance.

[6] https://en.wikipedia.org/wiki/Algorithms_for_calculating_variance

8 Conclusion

In this paper, we presented a gesture recognition system built for recognising 31 natural and interaction-oriented hand gestures. Our feature extraction is based on statistics and spectral properties of a sliding window over the data stream. We show that our features are highly discriminative for natural hand gestures and we achieve an accuracy of 98.5% with our gesture recognition system, which relies on linear discriminant analysis for dimensionality reduction and logistic regression for classification. Moreover, accuracy does not significantly suffer (98.2%) when the computationally expensive FFT features are removed. The main contribution of this paper lies in showing that all selected gestures can be recognised very well, given the sensors on the custom data glove and selected features extracted using sliding window approach.

This result is relevant for gesture-based interfaces, as it means that continuous gesture detection based on continuous sensing is accurate enough; and can be implemented in a computationally efficient manner. Computational efficiency is particularly important in wearable systems, considering the mobile nature of such systems. One direction of future work following up this line of argumentation will be an implementation of the recognition system on smartphone (wirelessly connected to the glove), making it a complete mobile solution. Further directions of future work on computing efficiency will include an exploration of which sensors are irrelevant and could be completely removed without degrading the recognition accuracy. Less sensors means less features, and hence more computational efficiency, but also fewer sensors to supply with power. Spreading out from the core of gesture recognition, it will of course be interesting to design interactions with computer systems using such natural and interaction-oriented gestures that can be recognised well.

Acknowledgements

Work on this paper is partially funded by MoreGrasp (H2020-ICT-2014-1 643955). The Know-Center is funded within the Austrian COMET Program - Competence Centers for Excellent Technologies - under the auspices of the Austrian Federal Ministry of Transport, Innovation and Technology, the Austrian Federal Ministry of Economy, Family and Youth and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency FFG.

We thank Eduardo Veas for his advice for building the data glove and planning the experiment study, and Christoffer Öjeling for his work on the data glove and his help in preparing and carrying out the experiment.

References

[1] H. Birk, T. B. Moeslund, and C. B. Madsen. Real-time recognition of hand alphabet gestures using principal component analysis. In *In 10th Scandinavian Conference on Image Analysis*, 1997.

[2] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[3] A. Bulling, U. Blanke, and B. Schiele. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Comput. Surv.*, 46(3):33:1–33:33, Jan. 2014.

[4] D. Cai, S. Member, X. He, J. Han, and S. Member. Srda: An efficient algorithm for large-scale discriminant analysis. *IEEE Transactions on Knowledge and Data Engineering*, pages 1–12, 2008.

[5] Q. Chen, N. D. Georganas, and E. Petriu. Real-time vision-based hand gesture recognition using haar-like features. In *Instrumentation and Measurement Technology Conference Proceedings, 2007. IMTC 2007. IEEE*, pages 1–6, May 2007.

[6] E. Coupet, F. Moutarde, and S. Manitsaris. Gesture recognition using a depth camera for human robot collaboration on assembly line.

Procedia Manufacturing, 3:518 – 525, 2015. 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, {AHFE} 2015.

[7] N. Dardas and N. D. Georganas. Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques. *Instrumentation and Measurement, IEEE Transactions on*, 60(11):3592–3607, Nov 2011.

[8] T. Dietterich. Machine learning for sequential data: A review. In T. Caelli, A. Amin, R. Duin, D. de Ridder, and M. Kamel, editors, *Structural, Syntactic, and Statistical Pattern Recognition*, volume 2396 of *Lecture Notes in Computer Science*, pages 15–30. Springer Berlin Heidelberg, 2002.

[9] P. Garg, N. Aggarwal, and S. Sofat. Vision based hand gesture recognition. *World Academy of Science, Engineering and Technology*, 49(1):972–977, 2009.

[10] P. Glomb, M. Romaszewski, S. Opozda, and A. Sochan. Choosing and modeling the hand gesture database for a natural user interface. In *Proceedings of the 9th International Conference on Gesture and Sign Language in Human-Computer Interaction and Embodied Communication*, GW'11, pages 24–35, Berlin, Heidelberg, 2012. Springer-Verlag.

[11] J. Han, S. Ahn, and G. Lee. Transture: Continuing a touch gesture on a small screen into the air. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, pages 1295–1300, New York, NY, USA, 2015. ACM.

[12] M. M. Hasan and P. K. Mishra. Hand gesture modeling and recognition using geometric features: a review. *Canadian Journal on Image Processing and Computer Vision*, 3(1):12–26, 2012.

[13] W. Higgins. A comparison of complementary and kalman filtering. *Aerospace and Electronic Systems, IEEE Transactions on*, AES-11(3):321–325, May 1975.

[14] C.-L. Huang and W.-Y. Huang. Sign language recognition using model-based tracking and a 3d hopfield neural network. *Mach. Vision Appl.*, 10(5-6):292–307, Apr. 1998.

[15] J.-F. Jego, A. Paljic, and P. Fuchs. User-defined gestural interaction: A study on gesture memorization. In *3D User Interfaces (3DUI), 2013 IEEE Symposium on*, pages 7–10, March 2013.

[16] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, 82(1):35–45, 1960.

[17] M. Kavakli, M. Taylor, and A. Trapeznikov. Designing in virtual reality (desire): A gesture-based interface. In *Proceedings of the 2Nd International Conference on Digital Interactive Media in Entertainment and Arts*, DIMEA '07, pages 131–136, New York, NY, USA, 2007. ACM.

[18] H. Kenn, F. V. Megen, and R. Sugar. A glove-based gesture interface for wearable computing applications. In *Applied Wearable Computing (IFAWC), 2007 4th International Forum on*, pages 1–10, March 2007.

[19] E. C. Kiruthika, E. Coimbatore, and N. Kumar. Survey on hand gesture recognition. *International Journal of Engineering Research and Technology*, 3(2):943–946, 2014.

[20] H. Koskim"aki, V. Huikari, P. Siirtola, and J. R"oning. Behavior modeling in industrial assembly lines using a wrist-worn inertial measurement unit. *Journal of Ambient Intelligence and Humanized Computing*, 4(2):187–194, 2013.

[21] H. Koskimaki, V. Huikari, P. Siirtola, P. Laurinen, and J. Roning. Activity recognition using a wrist-worn inertial measurement unit: A case study for industrial assembly lines. In *Control and Automation, 2009. MED '09. 17th Mediterranean Conference on*, pages 401–405, June 2009.

[22] N. C. Krishnan and D. J. Cook. Activity recognition on streaming sensor data. *Pervasive Mob. Comput.*, 10:138–154, Feb. 2014.

[23] A. Kulshreshth and J. J. LaViola, Jr. Exploring 3d user interface technologies for improving the gaming experience. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 125–134, New York, NY, USA, 2015. ACM.

[24] P. Kumar, J. Verma, and S. Prasad. Hand data glove: A wearable real-time device for human-computer interaction. *International Journal of Advanced Science and Technology*, 43, 2012.

[25] B.-G. Lee, B.-L. Lee, and W.-Y. Chung. Wristband-type driver vigilance monitoring system using smartwatch. *Sensors Journal, IEEE*, 15(10):5624–5633, Oct 2015.

[26] T. Martins, C. Sommerer, L. Mignonneau, and N. Correia. Gauntlet: a wearable interface for ubiquitous gaming. In *MobileHCI '08: Proceedings of the 10th international conference on Human computer interaction with mobile devices and services*, page 367, New York, New York, USA, Sept. 2008. ACM Request Permissions.

[27] P. Molchanov, S. Gupta, K. Kim, and K. Pulli. Multi-sensor system for drivers hand-gesture recognition. In *IEEE Conference on Automatic Face and Gesture Recognition*, pages 1–8, 2015.

[28] F. Mörchen. Time series feature extraction for data mining using dwt and dft. Technical report, Math and Computer Science Department, Philipps University, Marburg, Germany, 2003.

[29] K. Murakami and H. Taguchi. Gesture recognition using recurrent neural networks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '91, pages 237–242, New York, NY, USA, 1991. ACM.

[30] P. Neto, D. Pereira, J. Norberto Pires, and A. Moreira. Real-time and continuous hand gesture spotting: An approach based on artificial neural networks. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 178–183, May 2013.

[31] J. Ortiz Laguna, A. Olaya, and D. Borrajo. A dynamic sliding window approach for activity recognition. In J. Konstan, R. Conejo, J. Marzo, and N. Oliver, editors, *User Modeling, Adaption and Personalization*, volume 6787 of *Lecture Notes in Computer Science*, pages 219–230. Springer Berlin Heidelberg, 2011.

[32] J. Ou, Y. Shi, J. Wong, S. R. Fussell, and J. Yang. Combining audio and video to predict helpers' focus of attention in multiparty remote collaboration on physical tasks. In *Proceedings of the 8th International Conference on Multimodal Interfaces*, ICMI '06, pages 217–224, New York, NY, USA, 2006. ACM.

[33] C. Oz and M. C. Leu. American sign language word recognition with a sensory glove using artificial neural networks. *Eng. Appl. Artif. Intell.*, 24(7):1204–1213, Oct. 2011.

[34] G. Ren and E. O'Neill. Freehand gestural text entry for interactive tv. In *Proceedings of the 11th European Conference on Interactive TV and Video*, EuroITV '13, pages 121–130, New York, NY, USA, 2013. ACM.

[35] A. Roitberg, N. Somani, A. Perzylo, M. Rickert, and A. Knoll. Multimodal human activity recognition for industrial manufacturing processes in robotic workcells. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, ICMI '15, pages 259–266, New York, NY, USA, 2015. ACM.

[36] M. Romaszewski, P. Glomb, and P. Gawron. Natural hand gestures for human identification in a human-computer interface. In *Image Processing Theory, Tools and Applications (IPTA), 2014 4th International Conference on*, pages 1–6, Oct 2014.

[37] T. J. Sejnowski and C. R. Rosenberg. Parallel networks that learn to pronounce english text, 1987.

[38] S. Sen, V. Subbaraju, A. Misra, R. K. Balan, and Y. Lee. The case for smartwatch-based diet monitoring. In *Pervasive Computing and Communication Workshops (PerCom Workshops), 2015 IEEE International Conference on*, pages 585–590, March 2015.

[39] M. Shoaib, S. Bosch, H. Scholten, P. J. Havinga, and O. D. Incel. Towards detection of bad habits by fusing smartphone and smartwatch sensors. In *Pervasive Computing and Communication Workshops (PerCom Workshops), 2015 IEEE International Conference on*, pages 591–596, March 2015.

[40] T. Stiefmeier, D. Roggen, G. Ogris, P. Lukowicz, and P. Lukowicz. Wearable activity tracking in car manufacturing. *Pervasive Computing, IEEE*, 7(2):42–50, April 2008.

[41] T. Takahashi and F. Kishino. Hand gesture coding based on experiments using a hand gesture interface device. *SIGCHI Bull.*, 23(2):67–74, Mar. 1991.

[42] K. Tangwongsan, M. Hirzel, S. Schneider, and K.-L. Wu. General incremental sliding-window aggregation. *Proc. VLDB Endow.*, 8(7):702–713, Feb. 2015.

[43] T.-H. Tsai, C.-C. Huang, and K.-L. Zhang. Embedded virtual mouse system by using hand gesture recognition. In *Consumer Electronics - Taiwan (ICCE-TW), 2015 IEEE International Conference on*, pages

352–353, June 2015.

[44] W. Van Vlaenderen, J. Brulmans, J. Vermeulen, and J. Schöning. Watchme: A novel input method combining a smartwatch and bimanual interaction. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, pages 2091–2095, New York, NY, USA, 2015. ACM.

[45] J. P. Wachs, M. Kölsch, H. Stern, and Y. Edan. Vision-based hand-gesture applications. *Commun. ACM*, 54(2):60–71, Feb. 2011.

[46] J. Ward, P. Lukowicz, G. Troster, and T. Starner. Activity recognition of assembly tasks using body-worn microphones and accelerometers. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(10):1553–1567, Oct 2006.

[47] J. A. Ward. *Activity monitoring: Continuous recognition and performance evaluation*. PhD thesis, Swiss Federal Institute of Technology (ETH) Zürich, 2006.

[48] J. Weissmann and R. Salomon. Gesture recognition for virtual reality applications using data gloves and neural networks. In *Neural Networks, 1999. IJCNN '99. International Joint Conference on*, volume 3, pages 2043–2046 vol.3, 1999.

[49] C. Xu, P. H. Pathak, and P. Mohapatra. Finger-writing with smartwatch: A case for finger and hand gesture recognition using smartwatch. In *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications*, HotMobile '15, pages 9–14, New York, NY, USA, 2015. ACM.

[50] D. Xu. A neural network approach for hand gesture recognition in virtual reality driving training system of spg. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 3, pages 519–522, 2006.

[51] P. Zappi, T. Stiefmeier, E. Farella, D. Roggen, L. Benini, and G. Troster. Activity recognition from on-body sensors by classifier fusion: sensor scalability and robustness. In *Intelligent Sensors, Sensor Networks and Information, 2007. ISSNIP 2007. 3rd International Conference on*, pages 281–286, Dec 2007.

[52] X. Zhang, X. Chen, W.-h. Wang, J.-h. Yang, V. Lantz, and K.-q. Wang. Hand gesture recognition and virtual game control based on 3d accelerometer and emg sensors. In *Proceedings of the 14th International Conference on Intelligent User Interfaces*, IUI '09, pages 401–406, New York, NY, USA, 2009. ACM.

[53] Y. Zhao, P. H. Pathak, C. Xu, and P. Mohapatra. Demo: Finger and hand gesture recognition using smartwatch. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '15, pages 471–471, New York, NY, USA, 2015. ACM.

Bibliography

[Alles, 1970] Alles, D. S. (1970). Information transmission by phantom sensations. *IEEE Transactions on Man-Machine Systems*, 11(1):85–91.

[Barghout et al., 2009] Barghout, A., Cha, J., Saddik, A. E., Kammerl, J., and Steinbach, E. (2009). Spatial resolution of vibrotactile perception on the human forearm when exploiting funneling illusion. In *2009 IEEE International Workshop on Haptic Audio visual Environments and Games*, pages 19–23.

[Birk et al., 1997] Birk, H., Moeslund, T. B., and Madsen, C. B. (1997). Real-time recognition of hand alphabet gestures using principal component analysis. In *In 10th Scandinavian Conference on Image Analysis*.

[Bliss et al., 1970] Bliss, J. C., Katcher, M. H., Rogers, C. H., and Shepard, R. P. (1970). Optical-to-tactile image conversion for the blind. *IEEE Transactions on Man-Machine Systems*, 11(1):58–65.

[Bola et al., 2016] Bola, Ł., Siuda-Krzywicka, K., Paplińska, M., Sumera, E., Hańczur, P., and Szwed, M. S. (2016). Braille in the sighted: Teaching tactile reading to sighted adults. *PLoS ONE*.

[Braille, 1829] Braille, L. (1829). Method of writing words, music and plain songs by means of dots for use by the blind and arranged for them. *Paris, France: l'Institution Royale des Jeunes Aveugles*.

[Brewster and Brown, 2004] Brewster, S. and Brown, L. M. (2004). Tactons: Structured tactile messages for non-visual information display. In *Proceedings of the Fifth Conference on Australasian User Interface*, AUIC, pages 15–23, Darlinghurst, Australia, Australia. Australian Computer Society, Inc.

[Brown et al., 2006] Brown, L. M., Brewster, S. A., and Purchase, H. C. (2006). Multidimensional tactons for non-visual information presentation in mobile devices. In *Proceedings of the 8th Conference on Human-computer Interaction with Mobile Devices and Services*, MobileHCI '06, pages 231–238, New York, NY, USA. ACM.

[Byrd et al., 1995] Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995). A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM J. Sci. Comput.*, 16(5):1190–1208.

[Cattel, 1886] Cattel, J. M. (1886). The time taken up by cerebral operations. *Mind*, os-XI(42):220–242.

[Cauchard et al., 2016] Cauchard, J. R., Cheng, J. L., Pietrzak, T., and Landay, J. A. (2016). Activibe: Design and evaluation of vibrations for progress monitoring. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 3261–3271, New York, NY, USA. ACM.

[Cha et al., 2008] Cha, J., Rahal, L., and Saddik, A. E. (2008). A pilot study on simulating continuous sensation with two vibrating motors. In *2008 IEEE International Workshop on Haptic Audio visual Environments and Games*, pages 143–147.

[Chang et al., 2002] Chang, A., O'Modhrain, S., Jacob, R., Gunther, E., and Ishii, H. (2002). Comtouch: Design of a vibrotactile communication device. In *Proceedings of the 4th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, DIS '02, pages 312–320, New York, NY, USA. ACM.

[Chen, 2009] Chen, J. (2009). Laptop distracts truck driver, truck driver kills family of six. *GIZMO*.

[Chen et al., 2018a] Chen, J., Castillo, P., Turcott, R., Israr, A., and Lau, F. (2018a). Feeling speech on the arm. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, pages D107:1–D107:4, New York, NY, USA. ACM.

[Chen et al., 2018b] Chen, J., Turcott, R., Castillo, P., Setiawan, W., Lau, F., and Israr, A. (2018b). Learning to feel words: A comparison of learning approaches to acquire haptic words. In *Proceedings of the 15th ACM Symposium on Applied Perception*, SAP '18, pages 11:1–11:7, New York, NY, USA. ACM.

[Chen et al., 2007] Chen, Q., Georganas, N. D., and Petriu, E. (2007). Real-time vision-based hand gesture recognition using haar-like features. In *Instrumentation and Measurement Technology Conference Proceedings, 2007. IMTC 2007. IEEE*, pages 1–6.

[Cholewiak and Collins, 1995] Cholewiak, R. W. and Collins, A. A. (1995). Vibrotactile pattern discrimination and communality at several body sites. *Perception & Psychophysics*, 57(5):724–737.

[Cholewiak and Collins, 2000] Cholewiak, R. W. and Collins, A. A. (2000). The generation of vibrotactile patterns on a linear array: Influences of body site, time, and presentation mode. *Perception & Psychophysics*, 62(6):1220–1235.

[Cholewiak and Collins, 2003] Cholewiak, R. W. and Collins, A. A. (2003). Vibrotactile localization on the arm: Effects of place, space, and age. *Perception & Psychophysics*, 65(7):1058–1077.

[Cholewiak and Craig, 1984] Cholewiak, R. W. and Craig, J. C. (1984). Vibrotactile pattern recognition and discrimination at several body sites. *Perception & Psychophysics*, 35(6):503–514.

[Christina London and Rascon, 2014] Christina London, C. N. and Rascon, M. (2014). Teen distracted by cell phone hit, killed by truck. *NBC*.

[Chu and Beasley, 1997a] Chu, P. and Beasley, J. (1997a). A genetic algorithm for the generalised assignment problem. *Computers & Operations Research*, 24(1):17 – 23.

[Chu and Beasley, 1997b] Chu, P. C. and Beasley, J. E. (1997b). A genetic algorithm for the generalised assignment problem. *Comput. Oper. Res.*, 24(1):17–23.

[Craig, 1982] Craig, J. C. (1982). Temporal integration of vibrotactile patterns. *Perception & Psychophysics*, 32(3):219–229.

[Craig, 1983] Craig, J. C. (1983). Some factors affecting tactile pattern recognition. *International Journal of Neuroscience*, 19(1-4):47–57.

[Daneman, 1988] Daneman, M. (1988). How reading braille is both like and unlike reading print. *Memory & Cognition*, 16(6):497–504.

[Dardas and Georganas, 2011] Dardas, N. and Georganas, N. D. (2011). Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques. *Instrumentation and Measurement, IEEE Transactions on*, 60(11):3592–3607.

[de Jesus Oliveira and Maciel, 2014] de Jesus Oliveira and Maciel, A. (2014). Introducing the modifier tactile pattern for vibrotactile communication. In *EUROHAPTICS, 2014*.

[Dinse et al., 2005] Dinse, H. R., Kalisch, T., Ragert, P., Pleger, B., Schwenkreis, P., and Tegenthoff, M. (2005). Improving human haptic performance in normal and impaired human populations through unattended activation-based learning. *ACM Trans. Appl. Percept.*, 2(2):71–88.

[Duncan and Boynton, 2007] Duncan, R. O. and Boynton, G. M. (2007). Tactile hyperacuity thresholds correlate with finger maps in primary somatosensory cortex (s1). *Cerebral Cortex*, 17(12):2878.

[Dunkelberger et al., 2018] Dunkelberger, N., Sullivan, J., Bradley, J., Walling, N. P., Manickam, I., Dasarathy, G., Israr, A., Lau, F. W. Y., Klumb, K., Knott, B., Abnousi, F., Baraniuk, R., and O'Malley, M. K. (2018). Conveying language through haptics: A multi-sensory approach. In *Proceedings of the 2018 ACM International Symposium on Wearable Computers*, ISWC '18, pages 25–32, New York, NY, USA. ACM.

[Eid et al., 2015] Eid, M., Korres, G., and Jensen, C. B. F. (2015). Soa thresholds for the perception of discrete/continuous tactile stimulation. In *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*, pages 1–6.

[Fan et al., 2016] Fan, T., Ma, C., Gu, Z., Lv, Q., Chen, J., Ye, D., Huangfu, J., Sun, Y., Li, C., and Ran, L. (2016). Wireless hand gesture recognition based on continuous-wave doppler radar sensors. *IEEE Transactions on Microwave Theory and Techniques*, 64(11):4012–4020.

[Fisher, 1975] Fisher, D. F. (1975). Reading and visual search. *Memory & Cognition*, 3(2):188–196.

[Galvin et al., 2001] Galvin, K. L., Ginis, J., Cowan, R. S., Blamey, P. J., Clark, G. M., et al. (2001). A comparison of a new prototype tickle talkertm with the tactaid 7. *The Australian and New Zealand Journal of Audiology*, 23(1):18.

[Galvin et al., 1999] Galvin, K. L., Mavrias, G., Moore, A., Cowan, R. S., Blamey, P. J., and Clark, G. M. (1999). A comparison of tactaid ii+ and tactaid 7 use by adults with a profound hearing impairment. *Ear and hearing*, 20 6:471–82.

[Garg et al., 2009] Garg, P., Aggarwal, N., and Sofat, S. (2009). Vision based hand gesture recognition. *World Academy of Science, Engineering and Technology*, 49(1):972–977.

[Gault, 1924] Gault, R. H. (1924). Progress in experiments on tactual interpretation of oral speech. *The Journal of Abnormal Psychology and Social Psychology*.

[Geldard, 1957] Geldard, F. A. (1957). Adventures in tactile literacy. *American Psychologist*, 12(3):115–124.

[Geldard, 1960] Geldard, F. A. (1960). Some Neglected Possibilities of Communication. *Science*, 131(3413):1583–1588.

[Geldard, 1975] Geldard, F. A. (1975). *Sensory saltation: Metastability in the perceptual world.* Lawrence Erlbaum, Oxford, England.

[Geldard and Sherrick, 1972] Geldard, F. A. and Sherrick, C. E. (1972). The cutaneous "rabbit": A perceptual illusion.

[Gescheider, 1966] Gescheider, G. A. (1966). Resolving of successive clicks by the ears and skin. *Journal of Experimental Psychology*, 71(3):378.

[Gescheider et al., 2010] Gescheider, G. A., Wright, J. H., and Verrillo, R. T. (2010). *Information-processing channels in the tactile sensory system: A psychophysical and physiological analysis.* Psychology Press.

[Gill, 2003] Gill, J. (2003). Guidelines: Pictograms, icons and symbols, royal national institute of the blind, uk, 2003.

[Glomb et al., 2012] Glomb, P., Romaszewski, M., Opozda, S., and Sochan, A. (2012). Choosing and modeling the hand gesture database for a natural user interface. In *Proceedings of the 9th International Conference on Gesture and Sign Language in Human-Computer Interaction and Embodied Communication*, GW'11, pages 24–35, Berlin, Heidelberg. Springer-Verlag.

[Godde et al., 2000] Godde, B., Stauffenberg, B., Spengler, F., and Dinse, H. R. (2000). Tactile coactivation-induced changes in spatial discrimination performance. *Journal of Neuroscience*, 20(4):1597–1604.

[Goff, 1967] Goff, G. D. (1967). Differential discrimination of frequency of cutaneous mechanical vibration. *Journal of Experimental Psychology*, 74(2, Pt.1):294–299.

[González and Mark, 2004] González, V. M. and Mark, G. (2004). "constant, constant, multi-tasking craziness": Managing multiple working spheres. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 113–120, New York, NY, USA. ACM.

[Gunther, 2001] Gunther, E. (2001). *Skinscape: A tool for composition in the tactile modality*. PhD thesis, Massachusetts Institute of Technology.

[Gunther et al., 2002] Gunther, E., Davenport, G., and O'Modhrain, S. (2002). Cutaneous grooves: Composing for the sense of touch. In *Proceedings of the 2002 Conference on New Interfaces for Musical Expression*, NIME '02, pages 1–6. National University of Singapore.

[Guyton, 1991] Guyton, A. C. (1991). Sensory receptors: Neuronal circuits for processing information. *Basic Neuroscience: Anatomy & Physiology*.

[Han et al., 2015] Han, J., Ahn, S., and Lee, G. (2015). Transture: Continuing a touch gesture on a small screen into the air. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, pages 1295–1300, New York, NY, USA. ACM.

[Hasan and Mishra, 2012] Hasan, M. M. and Mishra, P. K. (2012). Hand gesture modeling and recognition using geometric features: a review. *Canadian Journal on Image Processing and Computer Vision*, 3(1):12–26.

[Heller and Schiff, 2013] Heller, M. A. and Schiff, W. (2013). *The psychology of touch*. Psychology Press.

[Hellmich, 2004] Hellmich, N. (2004). Most people multitask, so most people don't sit down to eat. *USA Today*.

[Hodzic et al., 2004] Hodzic, A., Veit, R., Karim, A. A., Erb, M., and Godde, B. (2004). Improvement and decline in tactile discrimination behavior after cortical plasticity induced by passive tactile coactivation. *Journal of Neuroscience*, 24(2):442–446.

[Hoggan et al., 2007] Hoggan, E. E., Anwar, S., and Brewster, S. A. (2007). *Mobile Multi-actuator Tactile Displays*. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Hughes et al., 2011] Hughes, B., Gemmert, A. W. V., and Stelmach, G. E. (2011). Linguistic and perceptual-motor contributions to the kinematic properties of the braille reading finger. *Human Movement Science*, 30(4):711 – 730. Special Issue: Progress in Graphonomics: A Perceptual Motor Skill Perspective.

[Israr et al., 2012] Israr, A., Kim, S.-C., Stec, J., and Poupyrev, I. (2012). Surround haptics: Tactile feedback for immersive gaming experiences. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '12, New York, NY, USA. ACM.

[Israr and Poupyrev, 2011] Israr, A. and Poupyrev, I. (2011). Control space of apparent haptic motion. In *2011 IEEE World Haptics Conference*, pages 457–462.

[Jiao et al., 2018] Jiao, Y., M. Severgnini, F., Martinez, J. S., Jung, J., Tan, H. Z., Reed, C. M., Wilson, E. C., Lau, F., Israr, A., Turcott, R., Klumb, K., and Abnousi, F. (2018). A comparative study of phoneme- and word-based learning of english words presented to the skin. In Prattichizzo, D., Shinoda, H., Tan, H. Z., Ruffaldi, E., and Frisoli, A., editors, *Haptics: Science, Technology, and Applications*, pages 623–635, Cham. Springer International Publishing.

[Kaczmarek et al., 1991] Kaczmarek, K., Webster, J., Bach-y Rita, P., and Tompkins, W. J. (1991). Electrotactile and vibrotactile displays for sensory substitution systems. *Biomedical Engineering, IEEE Transactions on*, 38(1):1–16.

[Kaul and Rohs, 2017] Kaul, O. B. and Rohs, M. (2017). Haptichead: A spherical vibrotactile grid around the head for 3d guidance in virtual and augmented reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 3729–3740, New York, NY, USA. ACM.

[Kirman, 1974a] Kirman, J. H. (1974a). Tactile apparent movement: The effects of interstimulus onset interval and stimulus duration. *Perception & Psychophysics*, 15(1):1–6.

[Kirman, 1974b] Kirman, J. H. (1974b). Tactile perception of computer-derived formant patterns from voiced speech. *The Journal of the Acoustical Society of America*, 55(1):163–169.

[Kirman, 1983] Kirman, J. H. (1983). Tactile apparent movement: The effects of shape and type of motion. *Perception & Psychophysics*, 34(1):96–102.

[Lamel et al., 1989] Lamel, L. F., Kassel, R. H., and Seneff, S. (1989). Speech database development: Design and analysis of the acoustic-phonetic corpus. In *Speech Input/Output Assessment and Speech Databases*.

[Larson, 2004] Larson, K. (2004). The science of word recognition. *Advanced Reading Technology, Microsoft Corporation*.

[LEDERMAN, 1991] LEDERMAN, S. J. (1991). Skin and touch. *Encyclopedia of Human Biology*, 7:51–63.

[Lederman and Klatzky, 1993] Lederman, S. J. and Klatzky, R. L. (1993). Extracting object properties through haptic exploration. *Acta Psychologica*, 84(1):29 – 40. Tactile Pattern Recognition.

[Lee and Starner, 2010] Lee, S. C. and Starner, T. (2010). Buzzwear: Alert perception in wearable tactile displays on the wrist. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10. ACM.

[Levenshtein, 1966] Levenshtein, V. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady*, 10:707.

[Li et al., 2016] Li, H., Yang, W., Wang, J., Xu, Y., and Huang, L. (2016). Wifinger: Talk to your smart devices with finger-grained gesture. In *Proceedings of the 2016*

ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp '16, pages 250–261, New York, NY, USA. ACM.

[Liao et al., 2016] Liao, Y.-C., Chen, Y.-L., Lo, J.-Y., Liang, R.-H., Chan, L., and Chen, B.-Y. (2016). Edgevib: Effective alphanumeric character output using a wrist-worn tactile display. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16, pages 595–601, New York, NY, USA. ACM.

[Lien et al., 2016] Lien, J., Gillian, N., Karagozler, M. E., Amihood, P., Schwesig, C., Olson, E., Raja, H., and Poupyrev, I. (2016). Soli: Ubiquitous gesture sensing with millimeter wave radar. *ACM Trans. Graph.*, 35(4):142:1–142:19.

[Lofvenberg and Johansson, 1984] Lofvenberg, J. and Johansson, R. (1984). Regional differences and interindividual variability in sensitivity to vibration in the glabrous skin of the human hand. *Brain Research*, 301(1):65 – 72.

[Luzhnica et al., 2016a] Luzhnica, G., Simon, J., Lex, E., and Pammer, V. (2016a). A sliding window approach to natural hand gesture recognition using a custom data glove. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 81–90.

[Luzhnica et al., 2017] Luzhnica, G., Stein, S., Veas, E., Pammer, V., Williamson, J., and Smith, R. M. (2017). Personalising vibrotactile displays through perceptual sensitivity adjustment. In *Proceedings of the 2017 ACM International Symposium on Wearable Computers*, ISWC '17, pages 66–73, New York, NY, USA. ACM.

[Luzhnica and Veas, 2017] Luzhnica, G. and Veas, E. (2017). Vibrotactile patterns using sensitivity prioritisation. In *Proceedings of the 2017 ACM International Symposium on Wearable Computers*, ISWC '17, pages 148–155, New York, NY, USA. ACM.

[Luzhnica and Veas, 2018a] Luzhnica, G. and Veas, E. (2018a). Investigating interactions for text recognition using a vibrotactile wearable display. In *23rd International Conference on Intelligent User Interfaces*, IUI '18, pages 453–465.

[Luzhnica and Veas, 2018b] Luzhnica, G. and Veas, E. (2018b). Skin reading meets speech recognition and object recognition for sensory substitution. In *Proceedings* *of the 2018 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2018 ACM International Symposium on Wearable Computers*, UbiComp '18, New York, NY, USA. ACM.

[Luzhnica and Veas, 2019a] Luzhnica, G. and Veas, E. (2019a). Background perception and comprehension of symbols conveyed through vibrotactile wearable displays. In *24rd International Conference on Intelligent User Interfaces*, IUI '19, New York, NY, USA. ACM.

[Luzhnica and Veas, 2019b] Luzhnica, G. and Veas, E. (2019b). Optimising the encoding for vibrotactile skin reading. In *ACM CHI Conference on Human Factors in Computing Systems*, CHI '19, New York, NY, USA. ACM.

[Luzhnica et al., 2016b] Luzhnica, G., Veas, E., and Pammer, V. (2016b). Skin reading: Encoding text in a 6-channel haptic display. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers*, ISWC '16, pages 148–155, New York, NY, USA. ACM.

[Luzhnica et al., 2018] Luzhnica, G., Veas, E., and Seim, C. (2018). Passive haptic learning for vibrotactile skin reading. In *Proceedings of the 2018 ACM International Symposium on Wearable Computers*, ISWC '18. ACM.

[Makous et al., 1995] Makous, J. C., Friedman, R. M., and Vierck, C. J. (1995). A critical band filter in touch. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 15 4:2808–18.

[Michel et al., 2011] Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., , Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A., and Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014):176–182.

[Millar, 2003] Millar, S. (2003). *Reading by touch.* Routledge.

[Millar, 2004] Millar, S. (2004). *Reading by Touch in Blind Children and Adults*, pages 437–457. Springer Netherlands, Dordrecht.

[Milnes et al., 1996] Milnes, P., Stevens, J. C., Brown, B. H., Summers, I. R., and Cooper, P. G. (1996). Use of micro-controller in a tactile aid for the hearing impaired. In *Proceedings of 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 1, pages 413–414 vol.1.

[Murakami and Taguchi, 1991] Murakami, K. and Taguchi, H. (1991). Gesture recognition using recurrent neural networks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '91, pages 237–242, New York, NY, USA. ACM.

[Neto et al., 2013] Neto, P., Pereira, D., Norberto Pires, J., and Moreira, A. (2013). Real-time and continuous hand gesture spotting: An approach based on artificial neural networks. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 178–183.

[News, 2017] News, F. (2017). Florida woman, 19, dies in car crash after sending text to her boyfriend. *GIZMO*.

[Nicolau et al., 2013] Nicolau, H., Guerreiro, J. a., Guerreiro, T., and Carriço, L. (2013). Ubibraille: Designing and evaluating a vibrotactile braille-reading device. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '13, pages 23:1–23:8, New York, NY, USA. ACM.

[Nicolau et al., 2015] Nicolau, H., Montague, K., Guerreiro, T., Rodrigues, A., and Hanson, V. L. (2015). Holibraille: Multipoint vibrotactile feedback on mobile devices. In *Proceedings of the 12th Web for All Conference*, W4A '15, pages 30:1–30:4, New York, NY, USA. ACM.

[Novich, 2015] Novich, S. D. (2015). *Sound-to-Touch Sensory Substitution and Beyond*. PhD thesis, Rice University.

[Novich and Eagleman, 2015] Novich, S. D. and Eagleman, D. M. (2015). Using space and time to encode vibrotactile information: toward an estimate of the skin's achievable throughput. *Experimental Brain Research*, (10).

[Osman and Potts, 1989] Osman, I. and Potts, C. (1989). Simulated annealing for permutation flow-shop scheduling. *Omega*, 17(6):551 – 557.

[Park et al., 2016] Park, J., Kim, J., Oh, Y., and Tan, H. Z. (2016). Rendering moving tactile stroke on the palm using a sparse 2d array. In *Haptics: Perception, Devices, Control, and Applications:International Conference, EuroHaptics 2016*, Cham. Springer International Publishing.

[Phillips et al., 1994] Phillips, A. J., Thornton, A. R. D., Worsfold, S., Downie, A., and Milligan, J. (1994). Experience of using vibrotactile aids with the profoundly deafened. *International Journal of Language and Communication Disorders*, 29(1):17–26.

[Pleger et al., 2001] Pleger, B., Dinse, H. R., Ragert, P., Schwenkreis, P., Malin, J. P., and Tegenthoff, M. (2001). Shifts in cortical representations predict human discrimination improvement. *Proceedings of the National Academy of Sciences*, 98(21):12255–12260.

[Pleger et al., 2003] Pleger, B., Foerster, A.-F., Ragert, P., Dinse, H. R., Schwenkreis, P., Malin, J.-P., Nicolas, V., and Tegenthoff, M. (2003). Functional imaging of perceptual learning in human primary and secondary somatosensory cortex. *Neuron*, 40(3):643–653.

[Prescott et al., 2011] Prescott, T. J., Diamond, M. E., and Wing, A. M. (2011). Active touch sensing.

[Pu et al., 2013] Pu, Q., Gupta, S., Gollakota, S., and Patel, S. (2013). Whole-home gesture recognition using wireless signals. In *Proceedings of the 19th Annual International Conference on Mobile Computing & Networking*, MobiCom '13, pages 27–38, New York, NY, USA. ACM.

[Purves et al., 2008] Purves, D., Augustine, G. J., Fitzpatrick, D., Hall, W. C., LaMantia, A.-S., McNamara, J. O., and White, L. E. (2008). Neuroscience. 4th. *Sunderland, Mass.: Sinauer. xvii*, 857:944.

[Rahal et al., 2009] Rahal, L., Cha, J., and Saddik, A. E. (2009). Continuous tactile perception for vibrotactile displays. In *2009 IEEE International Workshop on Robotic and Sensors Environments*, pages 86–91.

[Ranjbar, 2008] Ranjbar, P. (2008). Vibrotactile identification of signal-processed sounds from environmental events presented by a portable vibrator: A laboratory study. *Iranian Rehabilitation Journal*, 6.

[Rayner, 1998] Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124:372–422.

[Rayner et al., 2001] Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., and Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest*, 2(2):31–74. PMID: 26151366.

[Rayner et al., 2010] Rayner, K., Slattery, T. J., and Bélanger, N. N. (2010). Eye movements, the perceptual span, and reading speed. *Psychonomic Bulletin & Review*, 17(6):834–839.

[Reed and Delhorne, 2003] Reed, C. M. and Delhorne, L. A. (2003). The reception of environmental sounds through wearable tactual aids. *Ear and hearing*, 24(6):528–538.

[Reed et al., 2018] Reed, C. M., Tan, H. Z., Perez, Z. D., Wilson, E. C., Severgnini, F. M., Jung, J., Martinez, J. S., Jiao, Y., Israr, A., Lau, F., Klumb, K., Turcott, R., and Abnousi, F. (2018). A phonemic-based tactile display for speech communication. *IEEE Transactions on Haptics*, pages 1–1.

[Reeves, 1995] Reeves, C. R. (1995). A genetic algorithm for flowshop sequencing. *Computers & Operations Research*, 22(1):5 – 13. Genetic Algorithms.

[Reicher, 1969] Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 81(2):275.

[Reuter et al., 2014] Reuter, E.-M., Voelcker-Rehage, C., Vieluf, S., and Godde, B. (2014). Effects of age and expertise on tactile learning in humans. *European Journal of Neuroscience*, 40(3):2589–2599.

[Robles-De-La-Torre and Hayward, 2001] Robles-De-La-Torre, G. and Hayward, V. (2001). Force can overcome object geometry in the perception of shape through active touch. *Nature*, 412:445 EP –.

[Romaszewski et al., 2014] Romaszewski, M., Glomb, P., and Gawron, P. (2014). Natural hand gestures for human identification in a human-computer interface. In *Image Processing Theory, Tools and Applications (IPTA), 2014 4th International Conference on*, pages 1–6.

[Rönnberg et al., 1998] Rönnberg, J., Andersson, U., Lyxell, B., and Spens, K.-E. (1998). Vibrotactile speech tracking support: Cognitive prerequisites. *The Journal of Deaf Studies and Deaf Education*, 3(2):143–156.

[Schneider et al., 2015] Schneider, O. S., Israr, A., and MacLean, K. E. (2015). Tactile animation by direct manipulation of grid displays. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, UIST '15, pages 21–30, New York, NY, USA. ACM.

[Schneider et al., 1982] Schneider, W., Dumais, S. T., and Shiffrin, R. M. (1982). Automatic/control processing and attention. Technical report, ILLINOIS UNIV CHAMPAIGN HUMAN ATTENTION RESEARCH LAB.

[Schneider and Shiffrin, 1977] Schneider, W. and Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. detection, search, and attention. *Psychological review*, 84(1):1.

[Scott and Filippo, 1977] Scott, B. L. and Filippo, C. L. D. (1977). Progress in the development of a tactile aid for the deaf. *The Journal of the Acoustical Society of America*, 62(S1):S76–S76.

[Seim et al., 2014a] Seim, C., Chandler, J., DesPortes, K., Dhingra, S., Park, M., and Starner, T. (2014a). Passive haptic learning of braille typing. In *ACM International Symposium on Wearable Computers*, ISWC '14.

[Seim et al., 2017] Seim, C., Doering, N., Zhang, Y., Stuerzlinger, W., and Starner, T. (2017). Passive haptic training to improve speed and performance on a keypad. *IMWUT*, 1(3).

[Seim et al., 2015a] Seim, C., Estes, T., and Starner, T. (2015a). Towards passive haptic learning of piano songs. In *2015 IEEE World Haptics Conference (WHC)*, pages 445–450.

[Seim et al., 2015b] Seim, C., Hallam, J., Raghu, S., Le, T.-A., Bishop, G., and Starner, T. (2015b). Perception in hand-worn haptics: Placement, simultaneous stimuli, and vibration motor comparisons. Technical report, Georgia Institute of Technology.

[Seim et al., 2016] Seim, C., Reynolds-Haertle, S., Srinivas, S., and Starner, T. (2016). Tactile taps teach rhythmic text entry: Passive haptic learning of morse code. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers*, ISWC '16. ACM.

[Seim et al., 2014b] Seim, C. E., Quigley, D., and Starner, T. E. (2014b). Passive haptic learning of typing skills facilitated by wearable computers. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '14, pages 2203–2208, New York, NY, USA. ACM.

[Seo and Choi, 2010] Seo, J. and Choi, S. (2010). Initial study for creating linearly moving vibrotactile sensation on mobile device. In *2010 IEEE Haptics Symposium*, pages 67–70.

[Sherrick and Cholewiak, 1986] Sherrick, C. and Cholewiak, R. (1986). Handbook of perception and human performance, vol. i: Sensory processes and perception.

[Smith, 1969] Smith, F. (1969). Familiarity of configuration vs discriminability of features in the visual identification of words. *Psychonomic Science*, 14(6):261–262.

[Spence, 2014] Spence, C. (2014). The skin as a medium for sensory substitution. *Multisensory Research*, 27(5-6):293–312.

[Sterr et al., 2003] Sterr, A., Green, L., and Elbert, T. (2003). Blind braille readers mislocate tactile stimuli. *Biological Psychology*, 63(2):117 – 127.

[Stickgold and Walker, 2013] Stickgold, R. and Walker, M. P. (2013). Sleep-dependent memory triage: evolving generalization through selective processing. *Nature neuroscience*, 16(2):139.

[Summers et al., 2005] Summers, I. R., Whybrow, J. J., Gratton, D. A., Milnes, P., Brown, B. H., and Stevens, J. C. (2005). Tactile information transfer: A comparison of two stimulation sites. *The Journal of the Acoustical Society of America*, 118(4):2527–2534.

[Tan et al., 1997] Tan, H. Z., Durlach, N. I., Rabinowitz, W. M., Reed, C. M., and Santos, J. R. (1997). Reception of morse code through motional, vibrotactile, and auditory stimulation. *Perception & Psychophysics*, 59(7):1004–1017.

[Ternes and MacLean, 2008] Ternes, D. and MacLean, K. E. (2008). Designing large sets of haptic icons with rhythm. In Ferre, M., editor, *EuroHaptics*, pages 199–208, Madrid, Spain. Springer.

[Uleman and Bargh, 1989] Uleman, J. S. and Bargh, J. A. (1989). *Unintended thought*. Guilford Press.

[Van Vlaenderen et al., 2015] Van Vlaenderen, W., Brulmans, J., Vermeulen, J., and Schöning, J. (2015). Watchme: A novel input method combining a smartwatch and bimanual interaction. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, pages 2091–2095, New York, NY, USA. ACM.

[Vega-Bermudez and Johnson, 2001] Vega-Bermudez, F. and Johnson, K. O. (2001). Differences in spatial acuity between digits. *Neurology*, 56(10):1389–1391.

[Walker et al., 2002] Walker, M. P., Brakefield, T., Morgan, A., Hobson, J., and Stickgold, R. (2002). Practice with sleep makes perfect: Sleep-dependent motor skill learning. *Neuron*, 35(1):205 – 211.

[Weisenberger et al., 1991] Weisenberger, J., Broadstone, S., and Kozma-Spytek, L. (1991). Relative performance of single-channel and multichannel tactile aids for speech perception. *Journal of rehabilitation research and development*, 28(2):45–56.

[Whalen, 1991] Whalen, D. H. (1991). Infrequent words are longer in duration than frequent words. *The Journal of the Acoustical Society of America*, 90(4):2311–2311.

[White et al., 1970] White, B. W., Saunders, F. A., Scadden, L., Bach-Y-Rita, P., and Collins, C. C. (1970). Seeing with the skin. *Perception & Psychophysics*, 7(1):23–27.

[Woodworth, 1937] Woodworth, R. (1937). Experimental psychology. new york: Holt, 1938. *Department of Psychology Dartmouth College Hanover, New Hampshire.*

[Xu et al., 2011] Xu, C., Israr, A., Poupyrev, I., Bau, O., and Harrison, C. (2011). Tactile display for the visually impaired using teslatouch. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, pages 317–322, New York, NY, USA. ACM.

[Xu et al., 2015] Xu, C., Pathak, P. H., and Mohapatra, P. (2015). Finger-writing with smartwatch: A case for finger and hand gesture recognition using smartwatch. In *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications*, HotMobile '15, pages 9–14, New York, NY, USA. ACM.

[Xu, 2006] Xu, D. (2006). A neural network approach for hand gesture recognition in virtual reality driving training system of spg. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 3, pages 519–522.

[Yannier et al., 2015] Yannier, N., Israr, A., Lehman, J. F., and Klatzky, R. L. (2015). *FeelSleeve*: Haptic feedback to enhance early reading. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI 2015, 2015*, pages 1015–1024.

[Yuan et al., 2005] Yuan, H., Reed, C. M., and Durlach, N. I. (2005). Tactual display of consonant voicing as a supplement to lipreading. *The Journal of the Acoustical Society of America*, 118(2):1003–1015.

[Zhang et al., 2009] Zhang, X., Chen, X., Wang, W.-h., Yang, J.-h., Lantz, V., and Wang, K.-q. (2009). Hand gesture recognition and virtual game control based on 3d accelerometer and emg sensors. In *Proceedings of the 14th International Conference on Intelligent User Interfaces*, IUI '09, pages 401–406, New York, NY, USA. ACM.

[Zhao et al., 2018] Zhao, S., Israr, A., Lau, F., and Abnousi, F. (2018). Coding tactile symbols for phonemic communication. In *CHI Conference on Human Factors in Computing Systems.*

[Zhao et al., 2015] Zhao, Y., Pathak, P. H., Xu, C., and Mohapatra, P. (2015). Demo: Finger and hand gesture recognition using smartwatch. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '15, pages 471–471, New York, NY, USA. ACM.