TACTILE SPEECH COMMUNICATION:

DESIGN AND EVALUATION OF HAPTIC CODES FOR PHONEMES WITH GAME-BASED LEARNING

A Thesis

Submitted to the Faculty

of

Purdue University

by

Juan S. Martinez

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science in Electrical and Computer Engineering

May 2019

Purdue University

West Lafayette, Indiana

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF DISSERTATION APPROVAL

Dr. Hong Z. Tan, Chair

School of Electrical and Computer Engineering, Purdue University

Dr. Alexander Quinn

School of Electrical and Computer Engineering, Purdue University

Dr. Charlotte M. Reed

Research Laboratory of Electronics, Massachusetts Institute of Technology

Approved by:

Dr. Pedro Irazoqui

Head (interim) of Electrical and Computer Engineering

A todos aquellos que han participado en este gran esfuerzo. Mamá, Papá y Nicolás, este trabajo es el producto de su apoyo incondicional, que me ha impulsado a alcanzar mis metas y a la distancia ha contribuido a mi desarrollo personal y profesional. Nada de esto hubiera sido posible sin el amor de familia que siempre me han brindado. A ustedes, gracias.

ACKNOWLEDGMENTS

I would like to thank Professor Hong Z. Tan for her mentoring, advise, and her supervision in a very friendly and charismatic manner. Working with Professor Tan has been very rewarding and a very valuable experience, both academically, and personally. I would also like to thank Professor Alexander Quinn and Dr. Charlotte M. Reed for serving in my committee. Special thanks are extended to Dr. Reed for her advice, various contributions to this thesis, and the past work that led to the contributions of the thesis.

I would also like to thank all the people from the HIRL lab at Purdue that participated in experiments and gave advise and valuable feedback on all the work done. Many thanks to the speech project team: Jaehong Jung, Gina M. Clepper, and Ni Kang for their contributions to the project.

Finally, I would also like to acknowledge Google LLC, Facebook Inc., and Colfuturo for supporting all the work that led to the development of this thesis.

TABLE OF CONTENTS

Page

LIST OF TABLES vii

LIST OF FIGURES viii

ABSTRACT xi

1 Introduction 1

2 Background 4

2.1 The Tactile Sense 6

2.2 Tactile Speech Communication 11

2.3 Classical Psychophysics 13

2.4 Adaptive Procedures 16

2.5 Tactile Illusions 19

2.6 Serious Games 23

2.7 Information Theory 25

3 A Review of Prior Research on a Phonemic-Based Tactile Display for Speech Communication 28

3.1 TAPS (TActile Phonemic Sleeve) 29

3.2 Haptic Code Design 31

3.3 Software Interface 32

3.4 Experimental Procedures 35

3.4.1 Exploratory Experiments 36

3.4.2 Phoneme-based and Word-based Learning Protocols 38

3.4.3 Phoneme Identification Experiment 43

4 The “Haptos” Game 47

4.1 Introduction 47

4.2 Proposed Framework 49

Page

4.3 Haptos Design 51
4.4 First Version of Haptos 54
4.5 Exploratory Tests 63
4.5.1 Procedure 63
4.5.2 Results 64
4.6 Conclusion 68
5 Improved Haptic Code Design for Faster Transmission Rates with TAPS . . 70
5.1 Introduction 70
5.2 Statistical Studies of Spoken English 71
5.3 Design of Improved Haptic Codes 76
5.4 Evaluation of Improved Haptic Codes 83
5.4.1 A new version of the Haptos game 83
5.4.2 Learning Materials 84
5.4.3 Procedure 86
5.4.4 Data Analysis 89
5.4.5 Results 91
6 Conclusions and Future Work 103
REFERENCES 107
A Labels for Phoneme IPA Symbols 111
B Learning Material in the First Version of Haptos 113
C Software Library Architecture 115
D Improved Haptic Codes Description 121

LIST OF TABLES

Table Page

2.1 Classification of mechanoreceptors based on receptive field and adaptation rate . . . 7

2.2 Mechanoreceptor functions based on frequency . . . 8

2.3 Adaptive method comparison: up-down and one-up-two-down . . . 18

4.1 Summary of quest points in the first version of Haptos . . . 62

5.1 The top ten most frequent phonemes out of 72,210 phonemes (derived from Table II of [51]) . . . 72

5.2 The top ten most frequent phonemes out of 103,887 phonemes (derived from Table 3 of [52]) . . . 75

5.3 Haptic symbols by training zones . . . 84

5.4 Learning levels for the 49-symbol experiment . . . 87

5.5 Statistics for the learning pace of participants for consonants and vowels . 96

5.6 Results from the pooled tests on level L10 with the new haptic symbols . . 97

A.1 Phoneme labels for consonants . . . 111

A.2 Phoneme labels for vowels . . . 112

B.1 Learning material for the first version of Haptos . . . 113

D.1 Description of improved consonant codes . . . 122

D.2 Description of improved vowel codes . . . 129

D.3 Description of chunk codes for the most frequent co-occurring pairs of phonemes . . . 136

LIST OF FIGURES

Figure Page

2.1 Background topics involved in the main tasks of the thesis work 4

2.2 Location of mechanoreceptors in a cross section of the skin (taken from Fig.12.2 of [4, Chapter 12]) 7

2.3 Relative displacement threshold as a function of frequency for the 4 types of mechanoreceptors (adapted from Fig.8 of [5]) 8

2.4 Innervation density of the four mechanoreceptor types in the hand (taken from Fig.6 (B) of [6]) 9

2.5 Relative displacement threshold as a function of frequency for different contact areas (taken from Fig.7 of [10]) 10

2.6 The Tadoma method (taken from [11]) 11

2.7 Example of a psychometric function 14

2.8 Example of the up-down adaptive method (taken from Fig.4 of [27]). "+" marks denote a positive response and "—" marks denote a negative response.17

2.9 Temporal and spatial illustrations of the cutaneous rabbit effect and the funneling effect 20

2.10 Temporal and spatial illustration of the apparent motion 21

2.11 Input hypothesis model in second language acquisition (SLA) theory (modified from Fig.1 of [40]) 24

3.1 Tactor layout of the phonemic-based tactile display 30

3.2 Gauntlets used in phonemic-based studies 30

3.3 Haptic code design program implemented in MATLAB 33

3.4 Results of Experiment I in exploratory studies (taken from Fig. 5 of [1]) . 37

3.5 Progress on participants in the phoneme-based learning protocol (taken from Fig.4 of [2]) 40

3.6 Progress on participants in the word-based learning protocol (taken from Fig.6 of [2]) 42

Figure Page

3.7 Equivalent number of words correctly identified in the word-based approach (taken from Fig.7 of [2]) 42

3.8 Mean response times of participants in phoneme identification experiments (taken from Fig.7 of [3]) 46

4.1 General framework for role-playing game design in language learning . . . 49

4.2 Haptos design in terms of the role-playing game framework for language learning 52

4.3 First version of the Haptos game 55

4.4 Map of the first version of the Haptos game 56

4.5 Training School in the first version of the Haptos game 57

4.6 Quest puzzle to unlock quest 2 in the first version of Haptos 59

4.7 Quest 4 in the first version of Haptos 60

4.8 Fishing complementary quest in the first version of Haptos 61

4.9 Training time profiles of participants in the first version of Haptos 65

4.10 Performance of participants in the first version of Haptos (chance level = 33.33%) 67

4.11 Equivalent number of words correctly identified in the tests of the first version of Haptos. Tests had a chance level of 33.3% 68

5.1 Normalized heat map of frequencies of occurrence of pairs of phonemes (derived from Table IV of [51]) 73

5.2 Top ten most frequent pairs of co-occurring phonemes in [51] 74

5.3 Visual representation of the improved set of consonants 77

5.4 Visual representation of the improved set of vowels 80

5.5 Visual representation of abbreviated pairs of phonemes 82

5.6 The village in the second version of Haptos 83

5.7 Training school in the second version of Haptos 85

5.8 The lab: Testing zone in the second version of Haptos 88

5.9 Training behaviour of participants in experiments with the new haptic symbols 91

5.10 Response time behaviour of participants in level tests with the new haptic symbols 93

Figure Page

5.11 Performance of participants in successful level tests with the new haptic symbols. Error bars display a ± 1 standard deviation 94

5.12 Equivalent number of symbols correctly identified for groups of phonemes in level tests with the new haptic symbols 95

5.13 Visualization tool used to analyze the pooled confusion matrix in the L10 tests with the new haptic symbols 98

5.14 Confusions between consonants and between consonants and chunks in the pooled results of the L10 tests with the new haptic codes 99

5.15 Confusions between vowels, consonants and chunks in the pooled results of the L10 tests with the new haptic codes 100

C.1 Deployment diagram of the software library 116

C.2 Simplified class diagram of the Motu singleton component 116

C.3 Sequence diagram for playing a sentence in the software library 117

C.4 Sequence diagram for playing a sequence of phonemes in the software library 118

C.5 Sequence diagram for playing an arbitrary matrix in the software library 120

D.1 Tactor layout of the phonemic-based tactile display 121

ABSTRACT

MSECE, Purdue University, May 2019. Tactile Speech Communication: Design and Evaluation of Haptic Codes for Phonemes with Game-based Learning. Major Professor: Hong Z. Tan.

This thesis research was motivated by the need for increasing speech transmission rates through a phonemic-based tactile speech communication device named TAPS (TActile Phonemic Sleeve). The device consists of a 4-by-6 tactor array worn on the forearm that delivers vibrotactile patterns corresponding to English phonemes. Three studies that proceeded this thesis evaluated a coding strategy that mapped 39 English phonemes into vibrotactile patterns. This thesis corresponds to a continuation of the project with improvements summarized in two parts. First, a design and implementation of a training framework based on theories of second language acquisition and game-based learning is developed. A role playing game named Haptos was designed to implement this framework. A pilot study using the first version of the game showed that two participants were able to master a list of 52 words within 45 minutes of game play. Second, an improved set of haptic codes was designed. The design was based on the statistics of spoken English and included an additional set of codes that abbreviate the most frequently co-occurring phonemes in duration. The new set included 39 English phonemes and 10 additional abbreviated symbols. The new codes represent a 24 to 46% increase in word presentation rates. A second version of the Haptos game was implemented to test the new 49 codes in a learning curriculum distributed over multiple days. Eight participants learned the new codes within 6 hours of training and obtained an average score of 84.44% in symbol identification tests with error rates per haptic symbol below 18%. The results demonstrate the feasibility of employing the new codes for future work where the ability to receive

longer sequences of phonemes corresponding to phrases and sentences will be trained and tested.

1. INTRODUCTION

Tactile speech communication is an area of research that studies different mechanisms to transmit information from speech though vibrotactile stimulation. The applications of this type of communication systems are numerous. A few examples include: communication aids for people with severe auditory impairments, tactile supplements for speech perception, secure systems to encrypt speech messages, and new communication systems that do not depend on the visual and/or auditory sensory modalities. The challenge of transmitting the rich information of speech through the tactile sense has been tackled through various approaches. Every approach is based on a coding mechanism that maps some components from speech into haptic symbols delivered by a specific tactile device. Strategies for encoding speech include spectral-based frequency-to-place transformations, articulatory-based methods, and phonemic-based approaches. Systems have also been developed for transmitting written text, including letter-based approaches. Several problems arise when using these strategies. For example, frequency-based approaches introduce token variability as the same speech content produced by several speakers may translate into different tactile stimuli, all because the frequency content of the signals vary even though the carried information remains unchanged. Letter-based encodings of written text avoid the problem of token variability but result in relatively slow in word presentation rates. There are several advantage of a phonemic-based approach based on encoding English phonemes into distinct tactile patterns. The encoding removes the token variability problems associated with the spectral-based approach. Compared with letter-based encoding, phonemic encoding is relatively faster due to the fact that the number of phonemes is the same or less than the number of letters in any English word. The challenge, however, is that there are many more English phonemes (on the order of 40) than letters of the alphabet (26). The phonemic-based approach would work if it can be

demonstrated that up to 40 vibrotactile patterns can be created that are perceptually distinct and their mappings to phonemes can be learned within a reasonable amount of time.

Prior to this thesis research, a phonemic-based strategy was developed on a 24-tactor array named TAPS (TActile Phonemic Sleeve) worn on the forearm. Vibrotactile patterns corresponding to 39 English phonemes were designed and tested in several studies [1–3]. Phoneme recognition rates higher than 85% and word recognition rates for two-word phrases of up to 40 words per minute were achieved in these studies. The best participants in the studies were able to learn one English word per minute with a vocabulary size of 100 words. These results demonstrated that the transmission of speech through TAPS is feasible with a reasonable amount of training on the order of hours. Encouraged by the phoneme and word recognition results of these early studies, ongoing efforts focus on increasing the speech transmission rates with TAPS and the long-term training of users to receive continuous speech on their arms.

The objectives of this thesis research are to modify the first-generation vibrotactile phonemic codes for faster transmission rates, and to test the effectiveness of a game-based learning environment. First, a more engaging training mechanism was developed based on the theory of serious games and game-based second-language acquisition. This mechanism allowed naive participants to learn phoneme codes in a virtual village while immersed in a role playing game called Haptos. Secondly, modifications to phoneme codes were made in two significant ways: the most frequent phonemes in English were coded with the shortest duration, and new combination codes (called "chunks") were developed for the most frequently co-occurring pairs of phonemes. The result was an expanded set of 49 haptic codes that mapped to the 39 English phonemes and the 10 most frequently co-occurring phoneme pairs. The improved symbols represent abbreviations in time of the most frequently occurring and co-occurring phonemes and were expected to improve the word presentation rates

achievable with TAPS. The new codes were evaluated through a training and testing procedure using the Haptos game with naive participants.

This thesis document is organized as follows. Chapter 2 presents relevant background of the topics that impacted the core work of the thesis. Chapter 3 presents the technical description of TAPS, as well as all the work done on the phonemic-based approach through TAPS that proceeded this thesis research. Chapter 4 presents the design of a game-based learning mechanism to teach the phoneme encoding as a second language, as well as the design and implementation of the first version of the Haptos game. A pilot study that demonstrated the effectiveness of the training mechanism is presented. Chapter 5 contains the design of the improved haptic codes and the implementation of the second version of the Haptos game. The game was used in a training protocol where eight participants learned and were tested on the new set of 49 symbols. Results of symbol-identification tests are presented and discussed. Finally, Chapter 6 presents the conclusions of the work and the future directions of the research in tactile speech communication using TAPS and the new set of haptic symbols.

2. BACKGROUND

The work presented in this thesis required the development of three main core tasks: (i) the design and implementation of haptic codes for phonemes in a phonemic-based tactile speech communication device, (ii) the design of a game-based training paradigm, (iii) and the evaluation of the newly designed codes in phoneme identification experiments.

These tasks were approached based on several critical background topics which are reviewed in the present chapter, including: key concepts on the physiological aspects of the tactile sense, tactile speech communication, classical psychophysics, adaptive procedures, tactile illusions, serious games, and information theory. Fig. 2.1 shows a visualization of the background topics and how they are related to the main core tasks of the thesis.

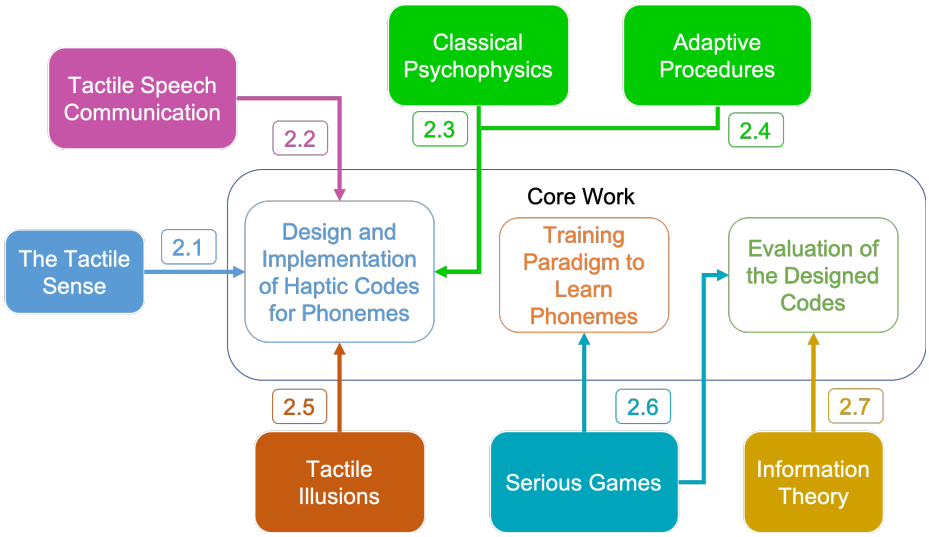


Fig. 2.1.: Background topics involved in the main tasks of the thesis work

According to the visualization, the numbered relations explain how the background topics inform the main tasks. They are described below following the numbering in Fig. 2.1.

2.1 The response of touch receptors in the skin to mechanical stimulation provides a set of design boundaries for physical parameters of vibrotactile stimuli. Parameters like frequency of vibration are determined by such boundaries and are important in the design of haptic codes for phonemes.

2.2 Tactile speech communication provides the theoretical background on different approaches to solve the problem of transmitting speech through the sense of touch. Design rules for haptic codes take into account what has been demonstrated in past research and existing effective methods for this type of communication.

2.3, 2.4 Classical psychophysics and adaptive procedures include methodologies to perform psychophysical experiments that determine intensity detection thresholds and perform actuator equalization. The results of these experiments determine the physical amplitudes of the vibrotactile stimuli that compose haptic codes for phonemes.

2.5 Tactile illusions include illusory movement sensations and mislocalization effects that can be exploited to design a set of rich and distinct haptic codes for phonemes.

2.6 Serious games provide an interesting approach to learning second languages with engaging and effective training techniques. The process of learning the haptic codes was performed based on a game-based training protocol. This approach allows experimenters to continuously monitor the learning process of participants through their progress in the virtual tasks of the game, which are designed to improve communication skills in the second language.

2.7 Information theory provides a theoretical framework to evaluate the transmission of information through a communication channel. Considering participants as communication channels and the haptic codes as sources of information, the study of this transmission provided a means to assess the information-transmission capacity of the haptic codes.

Based on the relationships established, the following sections describe the relevant aspects of each topic.

2.1 The Tactile Sense

The sense of touch refers to all sensations caused by mechanical displacements of the skin. This definition is also extended to include all the sensations caused by temperature, pain and sensations that inform about limb position. The term that includes all these sensations is "somatosensation" [4, Chapter 12].

For the development of this thesis, the most relevant aspects of somatosensation are the physiological structures of touch receptors in the skin, and their corresponding response to physical characteristics of mechanical stimuli. These receptors are embedded in both the epidermis and dermis layers of the skin. The primary "tactile receptors" are called *mechanoreceptors*. These are: *Meissner corpuscles, Merkell cell neurite complexes, Pacinian corpuscles* and *Ruffini endings*. Fig. 2.2 shows the location of these receptors in a cross section of the skin of the human hand.

The four main mechanoreceptors can be characterized by two main factors: the size of their receptive field, and their adaptation rate. The size of the receptive field specifies the body area to which the receptor will respond; while the adaptation rate describes the rate at which the neural pulses produced by the receptor "adapt" to a constant static stimulus, i.e, the rate at which the high frequency impulses produced by a stimulation ceases so that pulses subside to a normal rate. Table. 2.1 shows the classification of the mechanoreceptors based on these properties.

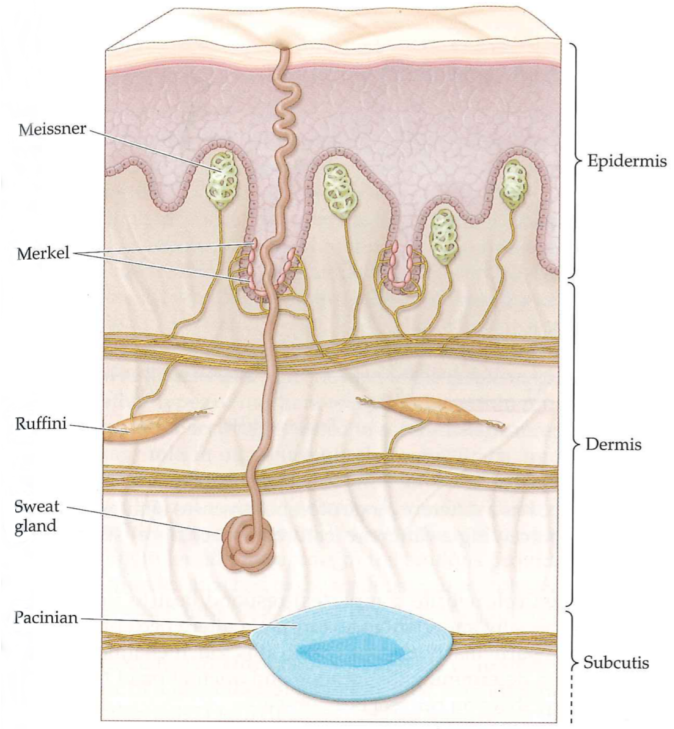


Fig. 2.2.: Location of mechanoreceptors in a cross section of the skin (taken from Fig.12.2 of [4, Chapter 12])

Table 2.1.: Classification of mechanoreceptors based on receptive field and adaptation rate

	Size of Receptive Field	
Adaptation Rate	**Small**	**Large**
Fast	FAI (Meissner)	FAII (Pacinian)
Slow	SAI (Merkel)	SAII (Ruffini)

Based on this classification, the mechanoreceptor types are responsible for the perception of different mechanical stimuli in terms of their frequency content. Such responses can be derived from the frequency behaviour of each mechanoreceptor in terms of their detection threshold. A threshold is defined as the minimum amount of

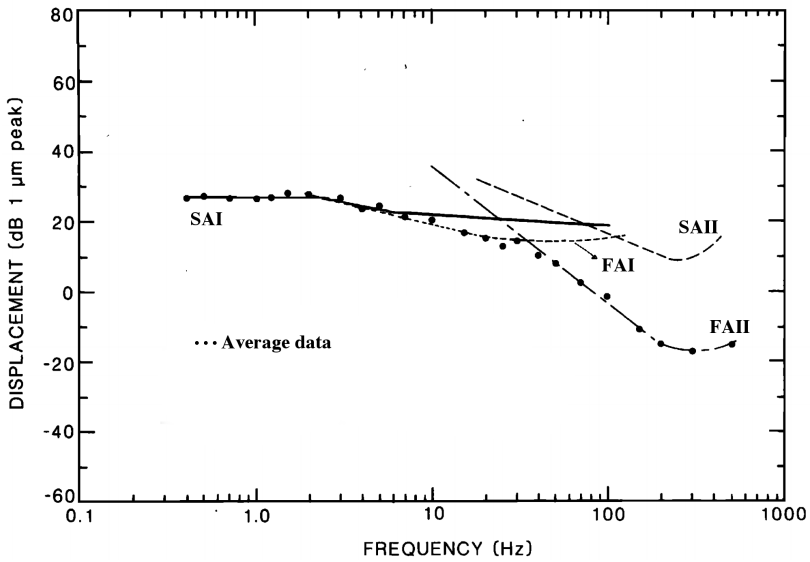


Fig. 2.3.: Relative displacement threshold as a function of frequency for the 4 types of mechanoreceptors (adapted from Fig.8 of [5])

Table 2.2.: Mechanoreceptor functions based on frequency

Mechanoreceptor type	**Sensitivity**	**Functions**
SAI	0.4 to 3Hz	Texture perception
FAI	3 to 40Hz	Low-frequency vibration detection
FAII	40 to >500Hz	High-frequency vibration detection
SAII	100 to >500Hz	Finger position, stable grasp

energy needed for a stimulus to be reliably perceived. Fig. 2.3 shows the minimally detectable displacement of the skin (relative to 1μm) produced by a vibrating stimulator as a function of frequency for each mechanoreceptor type. The figure has been adapted from the studies described in [5] from experiments in human hands. Based on Fig. 2.3, the different response ranges can be summarized in Table. 2.2 adapted from [4, Chapter 12].

The distribution of mechanoreceptors in the body also differs depending on the body part and the receptor type. For example, studies from [6] performed on a

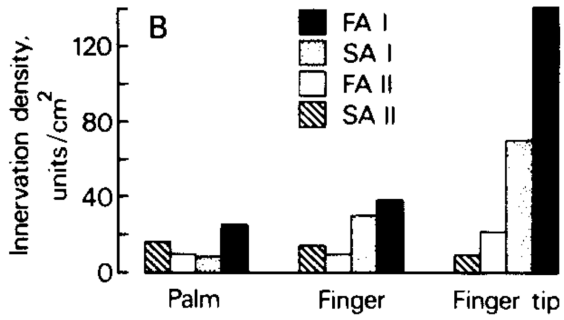


Fig. 2.4.: Innervation density of the four mechanoreceptor types in the hand (taken from Fig.6 (B) of [6])

glabrous (non-hairy) skin area of the human hand show the density of innervation of the four types of mechanoreceptors on three areas of the hand: The fingertip, the rest of the finger, and the palm. The relative densities across receptors in these regions were 4.2, 1.6 and 1.0, respectively. Fig. 2.4 shows a histogram of the innervation density in the mentioned areas. From this study, it can be observed that receptor types with small receptive fields are densely packed on the fingertips.

The density of receptors in different body parts influences the spatial resolution of the touch sense. This resolution has historically been measured as a two-point threshold, which is the smallest separation of two touch stimuli that can be identified as two instead of one. From studies performed in [7], two-point thresholds were measured for male and female participants for different body parts. Results show that the highest acuity is present at the extremities, with the fingertips having a resolution of approximately 5mm.

Although commonly used, the two-point threshold is not a measure of spatial resolution due to the ability of participants to differentiate two distinct points below the two-point threshold. Furthermore, measurements on the two-point threshold often show high variability within and between participants [8]. More accurate measures involve grating orientation experiments or gap detection experiments. An example

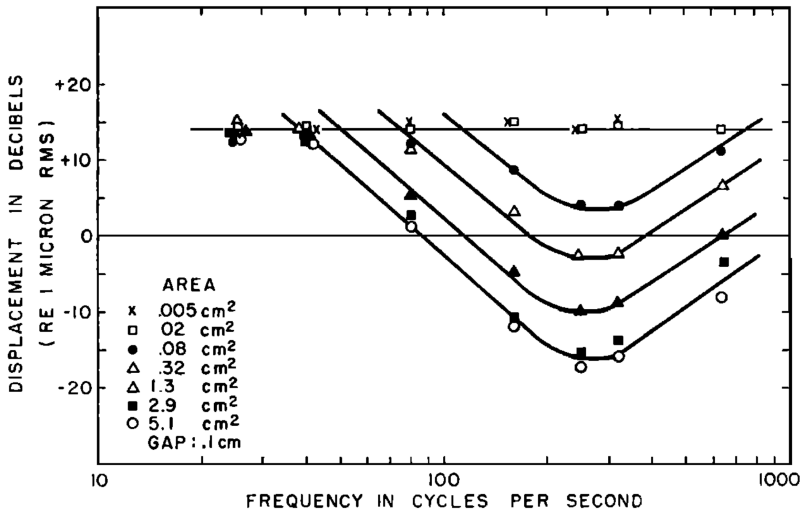


Fig. 2.5.: Relative displacement threshold as a function of frequency for different contact areas (taken from Fig.7 of [10])

of the latter are experiments described in [9], that showed resolutions at the index fingertip close to 1mm.

Another important factor that influences mechanoreceptor responses is the contact area of stimulation. Earlier experiments performed in [10] show that displacement threshold of a vibrating stimulus placed on the palm decreases as the contact area increases from 0.005cm^2 to 5.1cm^2. Fig. 2.5 shows this behaviour for different contact areas with a gap of 0.1cm between the skin and a non-vibrating rigid surface.

Overall, the different experiments and studies that characterize the response of mechanoreceptors demonstrate that the human sense of touch is capable of sensing vibrating signals as high as 700Hz. Moreover, spatial acuity of the cutaneous receptors varies with location and can achieve high resolution at body extremities. These results provide boundaries for the design of vibrotactile stimuli and offer information on the physical parameters required to excite the tactile sense.

Fig. 2.6.: The Tadoma method (taken from [11])

2.2 Tactile Speech Communication

The communication of speech through vibrotactile stimulation has always been a difficult task. Numerous approaches have been taken to map the rich and complex content of speech to vibrotactile stimuli. In this area of research, efforts are often motivated by the fact that the sense of touch is capable of transmitting such complex information. One example is the Tadoma method. This communication technique is mainly used by the deaf-blind community and consists of placing the hand on the face of a talker. During speech production, Tadoma users perceive and recognize the facial movements from the talker's face. Fig. 2.6 shows an example of the method.

Studies performed on this method have revealed that reception of connected speech is possible at rates between 60 and 80 words per minute (wpm). Furthermore, the identification accuracy of segmented consonants and vowels, isolated monosyllabic words, and key words in conversational sentences is 55%, 40% and 80%, respectively [11, 12].

Mapping of speech content to vibrotactile stimuli attempts to achieve word-per-minute rates close to those of natural methods like the Tadoma method; or even

normal speech rates on the order too 80 to 100 words per minute. Different approaches have studied the effectiveness of different mappings. For instance, popular approaches involve the extraction of spectral properties of speech that are later mapped to distinct locations of stimulation on the skin. For example, a vocoder was used in [13] to decompose speech into 23 frequency channels that activated 60 Hz sinusoidal signals in 23 vibrators. Tests with deaf subjects showed that recognition rates of words in isolation increased linearly with practice time. One of the top participants in the study mastered a list of 127 words within 12 weeks of training. A similar approach followed in [14] extracted the fundamental frequency of speech signals for it to be displayed as an eight-channel spatio-temporal stimulus. Experiments were conducted to investigate how vibrotactile stimuli could convey stress and intonation features in different conditions of supplementary visual and auditory information. Results showed that tactile feedback supplements speech reception by conveying sentential intonation contours.

Approaches that translate letters to tactile stimuli have also been explored [15,16]. For example, researchers in [16] presented the development of 2 synthetic alphabets that translate Roman characters to skin-stretch patterns delivered to a skin-stretch feedback device with 2 degrees of freedom. The first alphabet (named "Two-Dashes") was composed by minimal and distinct stretch patterns, while the second alphabet (named "Unistrokes") included patterns that draw a simplified, spatial version the Roman characters. Researchers studied the efficacy of the alphabets compared to regular Morse codes delivered on the same apparatus. Results proved that the Morse code alphabet was better recognized than the synthetic mappings. Nevertheless, correct recognition probabilities close to 80% were achieved with the other alphabets by the end of the experiment. As another strategy to encode text, studies in [17] compared the reception of Morse code through motional, vibrotactile and auditory stimulation. The mapping of speech to tactile sensations was made through Morse code and the use of a square-modulated 200 Hz sinusoidal signal. The presence and duration of the signal indicated the presence and duration of a dot or a dash. Results

collected from a word-recognition task showed that experienced Morse code operators could achieve communication rates of up to 25 words per minute.

Other studies have come to the general consensus that tactile aids are useful as supplements for other types of speech acquisition methods (e.g. lipreading) or environmental sounds [18–21]. For example, experiments in [22] studied sensory integration of tactile and auditory cues in speech-reading identification. Researchers showed that the use of envelope cues in the form of tactile stimuli improves speech-reading performance. These studies are supported by experimental evidence on sensory integration between tactile and auditory stimuli [23, 24].

Overall, the aforementioned approaches to tactile speech communication have demonstrated that tactile feedback is an effective supplement to other methods of speech perception. Nevertheless, speech communication itself through these approaches is not yet feasible. One practical method would require a robust mapping between speech and tactile stimuli that could be learned in reasonable time and generate high rates of communication (close to 60 words per minute or higher, as in the Tadoma method).

2.3 Classical Psychophysics

The study of perception in the context of haptic stimuli unfolds within the field of psychophysics. This area of psychology studies the relationship between sensations in the psychological domain and stimuli in the physical domain [25, Chapter 1]. A central area in this science is the study of the sensitivity limits of the human senses.

In the concept of haptic perception, the sensitivity limits and characteristics of the sense of touch are of particular interest. Such limits are mainly the *absolute threshold (AL)* and the *difference threshold (DL)*. Briefly mentioned in 2.1, AL is defined as the minimum amount of stimulus energy necessary to produce a sensation in 50% of the trials conducted in a psychophysical experiment. Similarly, DL is defined as the minimum amount of change in a stimulus required to produce a just

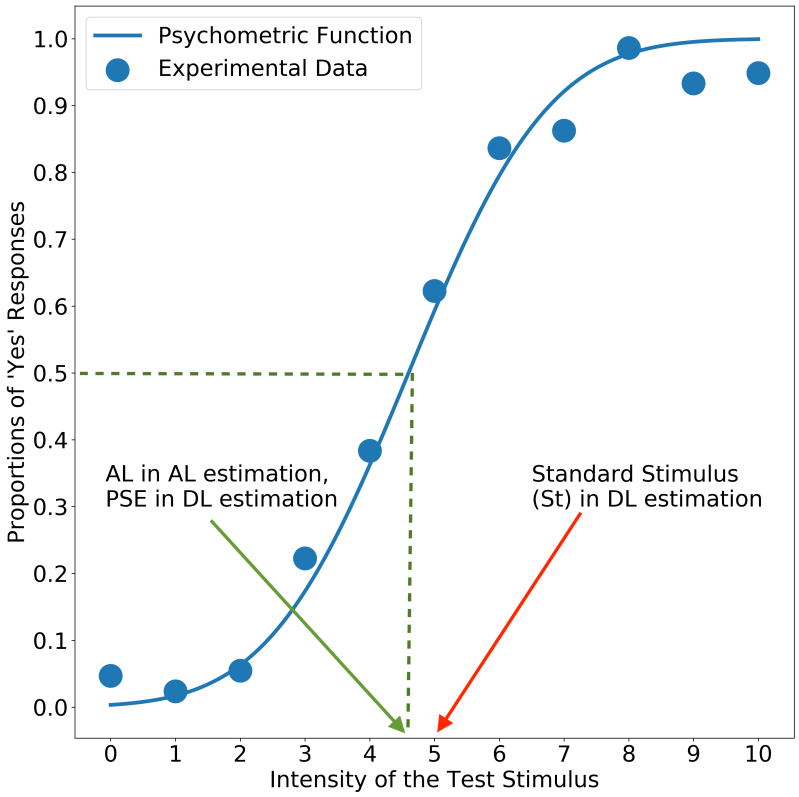


Fig. 2.7.: Example of a psychometric function

noticeable difference (jnd) in the sensation in, for example, 50% of the trials of an experiment [25].

To study these thresholds, three methods were developed by the German physicist G. T. Fechner in the 19th century. The methods were named: *the method of constant stimuli, the method of limits,* and *the method of adjustment.* Each one of these methods takes into account the random nature of the biological systems. This means that in an experiment in which a participant is meant to report if s/he perceives a stimulus or not, presenting the same stimulus over several trials may yield different amounts of "yes" and "no" answers. An important assumption of such methods is that the proportion of "yes" or "positive" responses is a monotonic function of the stimulus level and is often approximated by a cumulative Gaussian distribution function. This function is called a *psychometric function* and is illustrated in Fig. 2.7.

In the figure, data points are scattered in an increasing behaviour and approximated by a cumulative Gaussian distribution. The green lines mark an example of a threshold level of 50% in the y axis level, which corresponds to a stimulus value of ≈ 4.6 in the x axis. This x-axis point corresponds to the AL estimate in a detection experiment (i.e., AL estimation).

Considering a discrimination experiment (i.e., DL estimation), the x-axis point that corresponds to the 50% threshold level in the y axis is called the *point of subjective equality*, or *PSE*, and represents the value at which the *test stimulus*, over a large number of trials, is perceived subjectively as equal to a *standard or reference stimulus* (St). In Fig. 2.7, the standard stimulus is denoted by the red arrow and corresponds to a fixed reference stimulus that is compared to other comparison stimulus (Co) by participants. DL estimation experiments present pairs (Co, St) and the participant's task is to judge what stimulus from a presented pair produces a sensation of greater magnitude [25]. In these experiments, the fitted standard deviation of the Gaussian distribution can be taken as the DL estimate. Another important quantity is the *constant error* or *CE*, computed as $|PSE - S_t|$ and reflects stochastic fluctuations in the experiment that influence the measurements.

The procedure to estimate AL and DL for all classical methods can be found in [25, Chapter 2]. For the developments of this thesis, it is convenient to describe the procedures of the method of adjustment, both for AL estimation and DL estimation. In the case of AL estimation, the stimulus intensity is set by the experimenter to a limit far above or far below the threshold. Then, the participant is asked to decrease the intensity until it is undetectable, or to increase the intensity until it is just detectable, respectively. The participant is asked to perform the decreasing and increasing series by equal number of times, and the AL is estimated as the average of all the final settings [25, 26].

For the case of DL estimation, the participant is asked to adjust the intensity of a comparison stimulus until it feels equal to a standard stimulus (S_t). The intensity of the comparison stimulus is recorded after each series of adjustment. For all the

values of the adjusted stimulus, a plot of frequency of responses against stimulus intensity will approach a normal distribution (if a fair amount of trials are performed). The sample mean of the data ($\bar{X}$) is the *PSE* and the standard deviation σ is the *DL* [25, 26]. The latter can be obtained from a fitted Gaussian distribution of the experimental data or by analytical calculations such as:

$$\sigma = \sqrt{\frac{N \sum X^2 - \left(\sum X\right)^2}{N^2}}, \tag{2.1}$$

where X denotes a single stimulus intensity at the end of a series and N is the total number of series [25].

The method of adjustment provides a practical procedure to adjust the intensities of actuators compared to a reference stimulus. Even though this does not strictly follow a psychophysical experiment, the equalization process would ensure that the sensations produced by actuators arranged in an array would all be similar.

2.4 Adaptive Procedures

Compared to classical psychophysics, adaptive procedures are more efficient methods for threshold estimation. The main characteristic of these methods is that the stimulus intensity level (ϕ) on a specific trial is determined by the preceding stimulus level and responses [27]. In general, the procedures are based on experiments that try to locate a value of ϕ that corresponds to a specific "positive response" level of the participant. For the cases of "yes-no" experiments (e.g, experiments where the participant is asked "is the signal present?"), the threshold is defined at a particular percentage level of "yes" responses.

To estimate different values of ϕ at different performance levels, an experiment starts by presenting an initial stimulus level ϕ_{start} and increasing or decreasing its level depending on the participant's responses. Points at which the stimulus changes direction from increasing to decreasing or vice versa are called *reversals*, and the amount by which the stimulus increases or decreases is called *step size*. Questions to

be asked in this experiment design are: (1) When should the stimulus level change direction?, (2) What performance level to track?, (3) What is the step size for the stimulus level to change?, (4) When does the experiment end?, and (5) How is the threshold estimate calculated at the end? [28, Chapter 11]. Answers to all these questions vary greatly with a significant number of alternatives per question. One example of an adaptive design is the simple *up-down method* or *staircase method*, which estimates the threshold corresponding to the 50% point of the psychometric function. As described in [27], the method answers the questions posted above by: (1) Changing the stimulus direction after one opposite response relative to the previous response, (2) Tracking the 50% performance level, (3) Using a constant step size, (4) Testing until six to eight reversals are obtained, and (5) Running probit analysis or averaging peaks and valleys of reversal points to estimate the 50% point. An example of the method (taken from [27]) is shown in Fig. 2.8.

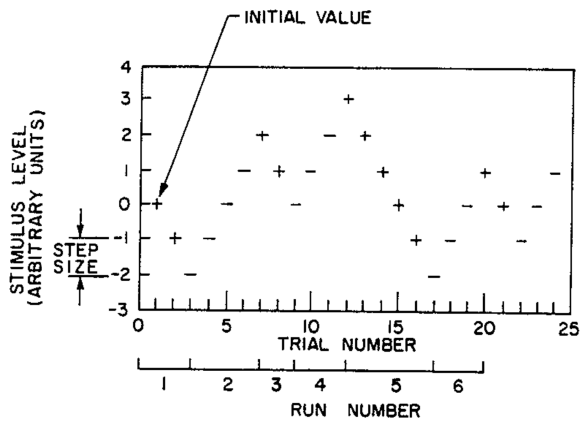


Fig. 2.8.: Example of the up-down adaptive method (taken from Fig.4 of [27]). "+" marks denote a positive response and "—" marks denote a negative response.

From these simple settings, the method can be adapted to use different step sizes depending on the number of reversals and data analysis techniques. Moreover, the strategy adopted to respond to question (1) would define estimated thresholds at percentile points different from the 50% point. These other methods are called *transformed up-down methods*. For example, Table. 2.3 provides a comparison of the simple

Table 2.3.: Adaptive method comparison: up-down and one-up-two-down

Method	**Decrease After**	**Increase after**	**Percentile point at convergence**
up-down	+	—	50%
one-up two-down	++	+ — or —	70.7%

up-down method and another modification called the *one-up two-down method*. The table is adapted from Table I of [27] where a more complete description of other transformed methods is available. Entries in the table marked as "—" denote a negative response of a trial, and "+" marks indicate a positive response.

Another modification of interest to these methods is the presentation of multiple intervals on each trial. For example, the *three-interval, two-alternative, one-up two-down method* is a modification of the one-up two-down method in which every trial of the experiment consist of three stimulus presentations. In every trial, two intervals contain a reference stimulus and one randomly selected interval contains the test stimulus. The task of the participant on every trial is to select the interval that contained the test stimulus. In the context of a detection experiment, the reference stimulus could be the absence of a signal, and the test stimulus could be the signal itself. On every subsequent trial, the intensity of the test stimulus would be increased or decreased according to the decision rules described in Table. 2.3 for the one-up two-down method.

Determining the absolute identification threshold through these methods is critical for the design of tactile stimuli. With this knowledge, the amplitude of a signal can be specified relative to the threshold of a participant so that the intensity of the stimulus will always be perceived as the correct intensity designed by the experimenter, regardless of the participant. The units *dBSL* (sensation level) are often used to specify these amplitudes. They indicate the intensity as the number of dB above a person's threshold.

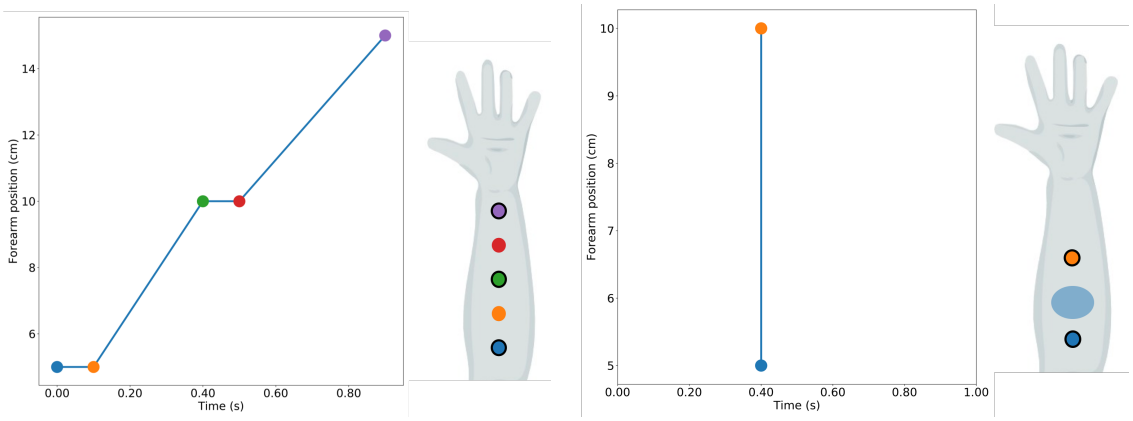
2.5 Tactile Illusions

The study of vibrotactile stimuli and its perception has led to the discovery of various illusory effects. These effects are very useful when trying to create dynamic and distinct stimulus from discrete physical vibrators. The characteristics of these illusions depend on the physical parameters of the vibration, such as spatial location, duration in time of individual vibrations, and temporal separation between two or more vibrators. The *tau* and *kappa* effects are basic examples of these type of phenomena.

As described in [29], the *tau* effect is a spatial illusion where three tactile stimuli are delivered sequentially such that the physical distance between the first and second is twice the distance between the second and third. It has been shown that, if the time interval between the second and third stimulus is twice as large as the time interval between the first and the second, the perceived distance between the second and third is twice as large as the distance between the first and second. For three stimuli, the effect is optimal when the ratio of the two time intervals is not greater than 4:1. The effect has been demonstrated for distances between 30 to 85 mm on the forearm and for inter-stimulus intervals between 200 and 500 ms [29]. Similarly, the *kappa* effect is a phenomenon where judgments in temporal intervals vary with spatial configurations of tactile stimuli.

These type of illusions show how the stimulation perceived by a participant can vary temporally and spatially based on physical parameters controlled by an experimenter. Similar to the working principles of the *tau* and *kappa* effects, other illusions produce dynamic sensations of movement along an area of the skin or phantom sensations that occur at non-stimulated areas. The most relevant effects to be discussed are the *"cutaneous rabbit" effect*, the *funneling illusion*, and *apparent motion*.

The "curaneous rabbit," or sensory saltation, is an effect produced by the sequential activation of short pulses at multiple locations on the skin. It is perceived as a stimulus that "hops" along the skin surface [29, 30]. The short pulses that compose



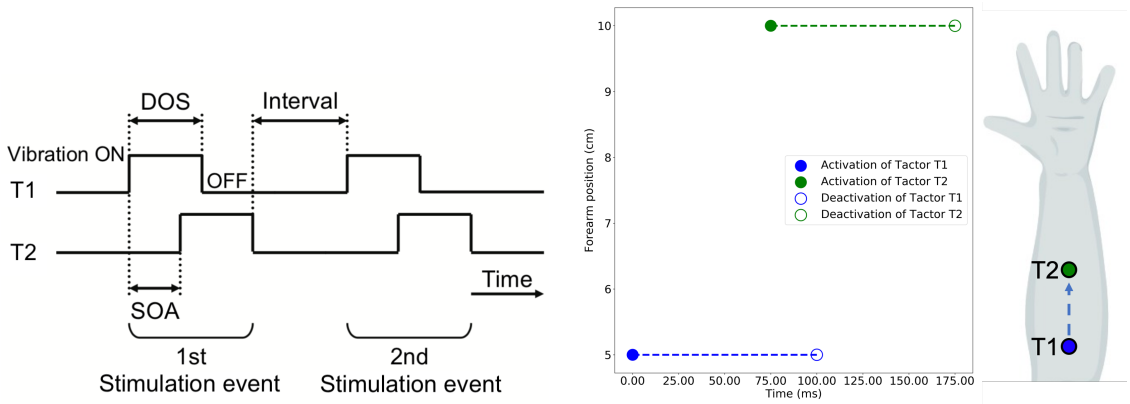
(a) Cutaneous rabbit effect

(b) Funneling illusion

Fig. 2.9.: Temporal and spatial illustrations of the cutaneous rabbit effect and the funneling effect

each stimulus vary between 3 and 6 taps for an optimal effect. Additionally, a temporal interval in the range between 20 and 250 ms between stimulus is optimal [29]. Studies in [30] demonstrated that the illusion is essentially independent of auditory temporal manipulations or from overt attention of the participants during stimulation. Fig. 2.9(a) shows the spatial and temporal properties of the "curaneous rabbit" effect. The plotted circles show the presence of vibrotactile stimuli at different instances of time and locations on the forearm. The perceived stimulus locations are divided into real physical vibrators (circles with solid lines) and phantom illusions (circles without lines).

Another type of illusion that involves the mislocalization of stimuli is the *funneling* illusion. This effect is described as a phantom stimulus perceived at an intermediate point between multiple stimuli when presented simultaneously on adjacent areas of the skin [29, 31]. The location of the phantom sensation can be controlled by the intensities of the two physical vibrators. An equal intensity produces a stimulus at the mid-point of the actuators and an unbalanced setting will locate the illusory stimulus closer to the strongest vibrator. Such an effect has been used to create a moving



(a) DOS and SOA times in apparent motion (adapted from [32])

(b) Apparent motion

Fig. 2.10.: Temporal and spatial illustration of the apparent motion

sensation between two actuators by gradually controlling their intensities over time [31]. An illustration of the spatial and temporal properties of the illusion is presented in Fig. 2.9(b). The two vibrators in the forearm are activated simultaneously and a phantom stimulus is perceived between the two.

Lastly, the *apparent motion* describes how sequential activation of a series of discrete vibrators produce a continuous and smooth stimulus moving across the skin [29, 32]. The effect is controlled by temporal characteristics of the physical actuators that create the stimuli. These parameters are the Duration of Stimulus (DOS) and the Stimulus-Onset Asynchrony (SOA). The DOS time refers to the amount of time in which an individual vibrator is activated; and the SOA time is the time difference between the successive activation of two actuators. For example, for an actuator (T1) activated at time t_1, a second adjacent actuator (T2) will be activated at time $t_2 = t_1 + SOA$. In this situation, it is often the case that $t_2 < t_1 + DOS$ meaning that the first tactor is still on when the second tactor is activated. This combination of DOS/SOA times generates the apparent motion illusion. An illustration is shown in Fig. 2.10(a) and in Fig. 2.10(b), where filled circles show activation of the actuators and empty circles show their deactivation. DOS and SOA values that generate the

illusion are codependent. As described in [29], for a DOS of 100 ms and an SOA of 70 ms, a continuous sense of movement is perceived. In Fig. 2.10(a), every stimulation event corresponds to the overlapped activation of two tactors placed at different distances on the forearm to create the illusion. The overlap can be evidenced in the time diagram of Fig. 2.10(b). The presented sequence creates a continuous movement from T1 to T2 on a stimulation event, and the time interval between successive events plays a critical role in the judgment of the direction of movement. Experiments performed in [32] proved that an interval of 400 ms or longer leads to 95% or better correct direction identification rates. This proved to be true for DOS/SOA pairs of 100/100, 200/200, 400/300 and 800/400 ms.

Additionally, other experiments performed in [32] explored the optimal combinations of DOS and SOA values to create effective circular apparent motion effects. The required number of tactors for a circular movement was also investigated. The study concluded that a circuit-completion time greater than 400 ms with 4 tactors creates reliable effects, where participants are able to determine the direction of movement with close to 100% correct-answer percentage.

All tactile illusions described can also be combined to create rich and high dimensional stimuli for haptic experiences. For example, the work presented in [33] describes the development of an algorithm that uses a sparse vibrotactile array on the back of participants to create smooth, two-dimensional tactile strokes on the skin. The algorithm named “Tactile Brush” combines the apparent motion illusion and the funneling effect. By determining the adequate SOA and intensity settings on discrete vibrators, the algorithm produces smooth movement sensations using phantom and physical actuators. The designed stimulus was evaluated in the context of a simulation game as a mechanism that provides a more immersive user experience.

2.6 Serious Games

The study of serious games involves the development and assessment of games created with additional purposes other than entertainment. Videogames have proven to be an effective means of perceptual learning (especially in the visual dimension) [34, 35], and as a consequence, a popular area of research is the application of such games to learning and training scenarios. For example, incidental categorization of auditory stimuli was shown to be possible in the context of a game that included other multimodal cues [36, 37]. This principle was extended to the study of categorization of non-native speech sounds in [38], in which native Japanese speakers learned the /r/-/l/ distinction in English through a videogame.

This aided speech categorization through videogames also motivates the study of second language acquisition in the same context. Some approaches to this task are games that implement the models and hypothesis proposed by Krashen in the Second Language Acquisition (SLA) theory [39]. In games that incorporate SLA, a learner is immersed in the second language such that key points of the theory are addressed. Some of these points are: acquisition must be implicit and subconscious, is attitude dependent, should present informal situations, use grammatical "feel", and have a stable order of acquisition [40].

Within second language acquisition theory, one important hypothesis that has been implemented in serious games is the "Input Hypothesis". The hypothesis states that the acquisition of language (provided that is sequential and divided into stages) progresses from stage i to stage $i + 1$ when the acquirer is able to understand input from stage $i + 1$ by focusing on the meaning rather than the form of the message [39]. A diagram that models how information flows according to the input hypothesis and how it is acquired in a learner's brain is shown in Fig. 2.11.

According to [40], the input hypothesis in second language learning requires abundant information from the target language to be always available, so that the user acquires an "intuitive feel" of the language rather than learning a formal grammatical

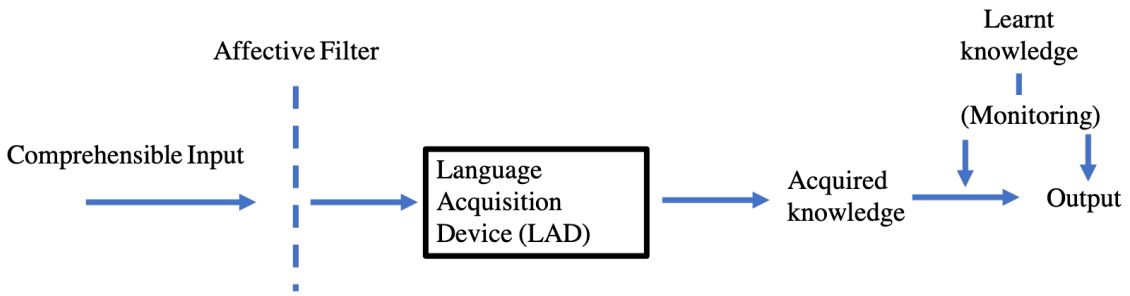


Fig. 2.11.: Input hypothesis model in second language acquisition (SLA) theory (modified from Fig.1 of [40])

structure. As shown in Fig. 2.11, this information is first filtered by an "affective" filter, which is responsible for learning inhibition. Secondly, the information is processed by a portion of the brain denoted as the "language acquisition device" (LAD). The acquired knowledge is the output of the device which is used by the learner to produce output in the target language. The knowledge is also corrected and monitored by traditionally learned grammar knowledge [40].

In this context, a suitable game design that is able to support second language acquisition is the role-playing game design. Role-playing games immerse users in a fictional world in which they assume the role of a virtual character who is part of a narrative that evolves throughout the game. In the context of second language acquisition, these type of games often include objectives in the story that are directly related to improving communication skills in the target language. In this design, effective learning is accomplished by task-based storylines, in which the virtual avatars contain a skill profile that progresses according to the learning objectives. A "good" game is then described as a design that balances entertainment and the development of cognitive tasks related to learning concepts in a new language [41–43].

One example of a role-playing game constructed under this characteristics is the game "Lost in the Middle Kingdom" presented in [40]. In this game, the user is immersed in a fantasy realm that teaches the Chinese language. The learner can interact with various objects and learn the associated words by listening to the spoken

utterances and looking at visual representations of the words. This game is also an example of the usage of pictographic representations that strengthen the relation between new words learned in the new language and previous concepts [44]. The virtual environment creates contextual scenarios that surround the user with the target language to support the acquisition of information by its meaning.

2.7 Information Theory

Information can be considered as the reduction in uncertainty about a system being studied. It is what is learned that reduces the variability in the state of such system. As denoted in [45], the concept can be explained in terms of a psychophysical experiment, in which the participant is viewed as a communication channel. Here, the stimulus contains a certain amount of input information (IS), and the participant's responses carry another amount of output information (IR). In addition, there exists a correlation between the information of the responses and the information in the input stimulus; such content is denoted as the transmitted information (IT).

Information theory provides a framework in which these three concepts can be quantified in *bits* in an *absolute identification experiment (AI)*. In this procedure, a set of k stimuli is randomly presented to a participant following a one-interval paradigm (one stimulus presentation per trial). The task of the participant is to identify the stimulus in every trial by assigning a response label to the stimulus presented. From the classical work of Shannon in [46], the quantification of both IS and IR in this experiment can be written as

$$\begin{aligned} IS &= -\sum_{i=1}^{k} P(S_i)\log_2 P(S_i) \\ IR &= -\sum_{j=1}^{k} P(R_j)\log_2 P(R_j), \end{aligned} \tag{2.2}$$

where k is the number of distinct stimuli, each one with an *a priori* probability $P(S_i)$. The responses made by the participant are denoted as R_j, each one with a probability

$P(R_j)$. The reduction in uncertainty can be calculated from the probability of the stimulus being presented before and after a specific response of the participant ($P(S_i)$ and $P(S_i|R_j)$, respectively). This results in the correlation between the information in the stimulus and the information obtained per trial, which represents the information transferred. This quantity is then calculated as

$$IT(S_i, R_j) = \log_2 \frac{P(S_i|R_j)}{P(S_i)}. \tag{2.3}$$

To account for the total number of trials, the average IT is then determined by

$$IT = \sum_{i=1}^{k} \sum_{j=1}^{k} P(S_i, R_j) IT(S_i, R_j). \tag{2.4}$$

The probabilities $P(S_i)$ depend on the type of randomization used in the experiment. It is often preferred to use randomization with replacement to ensure that the probabilities remain constant throughout the experiment.

In a practical scenario, the results from an AI experiment are collected in a stimulus-response confusion matrix. This structure is a square matrix of size $k \times k$ with rows and columns labeled after all the stimuli and responses, respectively. The entry n_{ij} of the matrix represents the number of trials that stimulus i was labeled as response j. From the confusion matrix, the maximum-likelihood estimate for IT is

$$IT_{est} = \sum_{i=1}^{k} \sum_{j=1}^{k} \left(\frac{n_{ij}}{n} \right) \log_2 \left(\frac{n_{ij} \cdot n}{n_i \cdot n_j} \right), \tag{2.5}$$

where $n_i = \sum_{j=1}^{k} n_{ij}$, $n_j = \sum_{i=1}^{k} n_{ij}$ and $n = \sum_{i=1}^{k} \sum_{j=1}^{k} n_{ij}$. This estimated information transfer is valuable in determining the number of categories or elements that a participant can correctly identify without any error. This quantity is given by $\lfloor 2^{IT_{est}} \rfloor$. The amount of bits transferred also has a limit, which is denoted as the *channel capacity*. As discussed in [45], the capacity in various experiments that involve uni-dimensional stimuli is surprisingly consistent and lies in the range of 2.3 to 3.2 bits.

The maximum-likelihood estimate of IT, IT_{est}, is biased and subject to statistical fluctuations. Moreover, for limited number of trials, it tends to over estimate IT. As discussed in [47], several issues arise when the number of trials n is not large enough (i.e., when $n < 5k^2$). For experiments in which the number of trials collected is significantly lower than $5k^2$ but the error rate is low, it is sometimes better to compute a conservative lower-bound for IT instead of IT_{est}. The bound depends on the error rate e which can be derived from the percent correct score (pc) as

$$\begin{aligned} pc &= \frac{\sum_{i=1}^{k} n_{ii}}{n}, \\ e &= 1 - pc. \end{aligned} \tag{2.6}$$

Thus, as presented in [47], the conservative lower-bound estimate for IT can be calculated as:

$$IT_{pc} = (1 - 2e) \cdot \log_2 k \tag{2.7}$$

3. A REVIEW OF PRIOR RESEARCH ON A PHONEMIC-BASED TACTILE DISPLAY FOR SPEECH COMMUNICATION

The core work of this thesis (initially presented in Fig. 2.1 of Chapter 2) results from the motivation of creating a tactile speech communication system that can be learned within a reasonable amount of time. Between the summer of 2017 and fall of 2018, a team of researchers from Purdue University, Massachusetts Institute of Technology and Facebook developed a phonemic-based speech communication system for the forearm. Training and testing were conducted with the system on over 60 participants. The results show that the best participants can acquire one English word per minute using a vocabulary of up to 500 English words.

This thesis research was made possible with the hardware and software systems developed during the 12-month skunkworks project. It is therefore necessary to provide an overview of all the components that this thesis research depends on, including the haptic codes for English phonemes, the hardware and software interfaces, the learning and testing procedures, and the major results. The materials presented in this chapter are largely based on the following three publications:

- C. M. Reed, H. Z. Tan, Z. D. Perez, E. C. Wilson, F. M. Severgnini, J. Jung, **J. S. Martinez**, Y. Jiao, A. Israr, F. Lau, K. Klumb, F. Turcott, and R. Abnousi, "A phonemic-based tactile display for speech communication," *IEEE Transactions on Haptics*, vol. 12, pp. 2–17, 2019. [3]

- J. Jung, Y. Jiao, F. M. Severgnini, H. Z. Tan, C. M. Reed, A. Israr, F. Lau, and F. Abnousi, "Speech communication through the skin: design of learning protocols and initial findings," *Proceedings of HCI International*, 2018, pp.15−20, 2018. [1]

- Y. Jiao, F. M. Severgnini, **J. S. Martinez**, J. Jung, H. Z. Tan, C. M. Reed, E. C.Wilson, F. Lau, A. Israr, R. Turcott, K. Klumb, and F. Abnousi, "A comparative study of phoneme-and word-based learning of english words presented to the skin," *Proceedings of EuroHaptics 2018*, pp. 623−635, 2018. [2]

The author of this thesis was a co-author on two of the publications. The research unique to this thesis is presented later in Chapters 4 and 5.

3.1 TAPS (TActile Phonemic Sleeve)

The hardware used consists of a 4-by-6 tactor array worn on the forearm as a sleeve. The device (named TAPS under TActile Phonemic Sleeve) was composed of 24 wide-bandwidth 32.2 mm × 26.3 mm × 9.0 mm Tectonic TEAX13C02-8/RH audio exciters (https://www.tectonicaudiolabs.com/resources/) used as tactors. Each tactor had a nominal impedance of 8 Ω, a flat frequency response between 50 Hz and 2 kHz, and a peak resonance at approximately 600 Hz. Tactors were arranged to form two rows of six tactors along the longitudinal direction of both volar and dorsal sides of the forearm. This configuration involved a total of 24 tactors distributed in six longitudinal and four transverse rows on the forearm. Fig. 3.1 shows the tactor layout of the configuration. In the figure, a yellow oval labeled with "TX" references tactor number X, where X is a number between 1 and 24. Tactors T1-T12 reside on the dorsal (back) side of the arm, and T13-T24 are on the volar side (underside) of the arm.

Participants wore the device on the left forearm as a gauntlet. The gauntlets used by participants varied between experiments [1, 2] and [3], but were all based on velcro attachments to individual tactors. The images in Fig. 3.2 show examples of the gauntlets used by the participants in the different studies.

The set of 24 tactors were driven by 12, 3.7 W class D Maxim MAX98306 stereo audio amplifiers sourced from Adafruit (https://www.adafruit.com/product/987). The amplifiers were grouped by custom-made connection boards that connected all ampli-

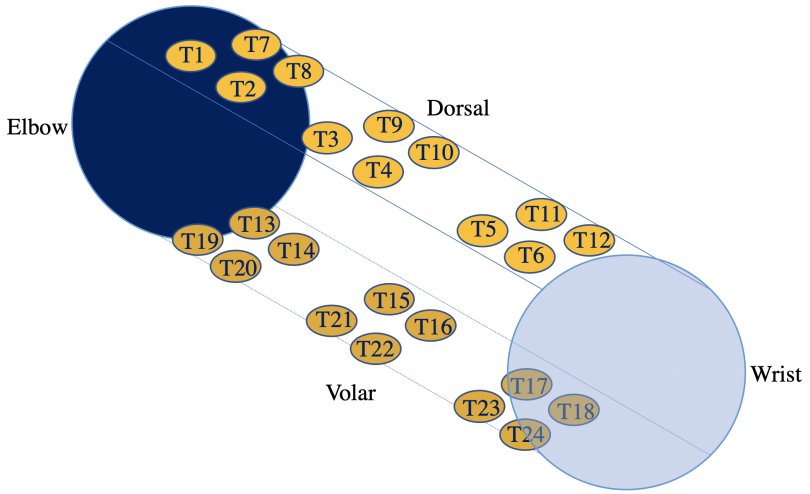
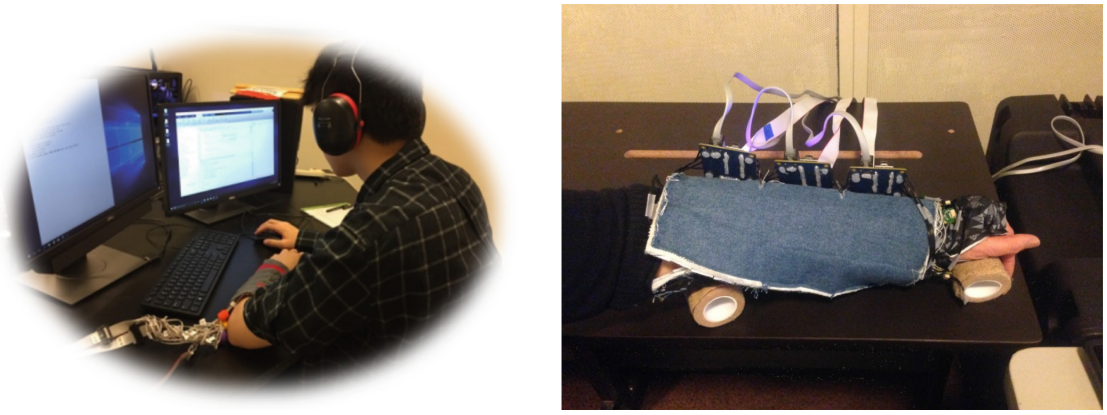


Fig. 3.1.: Tactor layout of the phonemic-based tactile display



(a) Participant wearing the gauntlet in [1, 2]

(b) Participant wearing the gauntlet in [3]

Fig. 3.2.: Gauntlets used in phonemic-based studies

fier inputs to 3 DB-25 cable connectors. The DB-25 cables were connected to the 24 output channels of a MOTU 24 Ao audio interface (http://motu.com/products/avb/24ai-24ao) through 3, 8-channel ports, one for each DB-25 cable. The audio interface was connected to a Windows computer through a USB port.

3.2 Haptic Code Design

As explained in [3], a phonemic-based approach was taken to encode all 39 English phonemes, which was used to construct words by concatenating the phonemes in each word. The research efforts of the study focused on the design of 39 distinct haptic symbols and the mapping to the 39 English phonemes. The phonemes consist of 24 consonants and 15 vowels. Due to the difficulty associated with the display of some of the symbols in the IPA (International Phonetic Alphabet), a custom-designed labeling system was used, where each phoneme was represented by one or two uppercase letters. Table. A.1 and Table. A.2 in appendix A show the correspondence between the labels and the IPA symbols for consonants and vowels, respectively. As described in [3], the design of the codes was based on three main considerations:

1. The perceptual properties of the tactile sense
2. Adequate mapping of the articulatory properties of phonemes to tactile stimuli
3. The distinctiveness of the haptic codes

To address these concerns, sinusoidal signals of 60 and 300 Hz were used as base stimulation patterns on all the tactors of TAPS. Within the set of consonants, manner of articulation, place of articulation and voicing were encoded in the frequency characteristics of the signals, their duration, and their location. The set of phonemes was divided into plosives (P, B, T, D, K, G), nasals (M, N, NG), semivowels (H, W, R, L, Y), fricatives (F, V, TH, DH, S, Z, SH, ZH) and affricatives (CH, J). Signals of 100 ms distinguished plosives from the other types of consonants (with a duration of 400 ms); and a frequency of 60 Hz was used to distinguish nasals from plosives, fricatives and affricatives. Place of articulation was encoded along the longitudinal direction such that phonemes produced at the front of the mouth were coded at the wrist, those in the middle of the mouth were presented in the middle of the arm, and those produced at the back of the mouth were delivered at the elbow. The rules for consonant coding were violated in a few specific cases to produce distinct haptic patterns (e.g., S and

Z do not follow the place of articulation rules). Finally, amplitude modulation was used to distinguish voiced from unvoiced phonemes. A complete description of all the consonants with their physical characteristics is available in Table 1 of [3].

The design of vowel codes mainly followed the third constraint. For this case, stimuli that moved on the arm were designed for each of the 15 vowels based on different haptic illusions. The main effects that take part in this coding are the apparent motion effect (sensation of continuous and smooth movement produced by sequential activation of discrete vibrators [29, 32]), sensory saltation ("hoping" stimulus moving across the skin [29, 30]), and the funneling effect (phantom stimulus perceived between multiple stimuli [29, 31]). A description of these illusions is given in section 2.5 of Chapter 2.

The 15 vowels were divided into six long vowels (EE, AH, OO, AE, AW, ER), four short vowels (UH, IH, EH, UU), and five diphthongs (AY, I, OW, OE, OY). Long and short codes were coded as long and short smooth apparent motions, respectively. Two diphthongs (OW and OY) were coded with sensory saltation effects with 3 taps at each of 3 successively stimulated tactors. The smooth apparent motion movements and sensory saltation movements ran through the arm in the longitudinal direction on both the dorsal and volar sides. Some of the diphthongs traversed the arm in the course of their movement, occasionally evoking a funneling effect. A complete description of the codes for vowels and the corresponding sensations is provided in Table 2 of [3].

3.3 Software Interface

Custom MATLAB software was used to design all the 39 codes as digital waveforms and to deliver them to TAPS by through the MOTU 24 Ao interface. The software created 24-channel programmable waveforms with varying amplitude, duration frequency, amplitude modulation, and timing properties that were later played using the playrec utility (http://www.playrec.co.uk/). The software included programs to

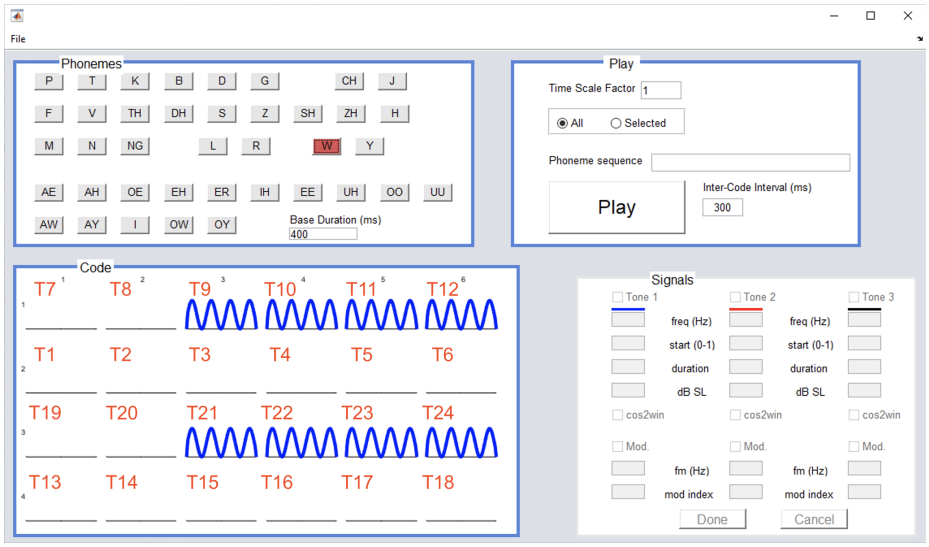


Fig. 3.3.: Haptic code design program implemented in MATLAB

design haptic codes, perform detection threshold measurements, adjust tactor intensities, and perform phoneme and word identification training and testing sessions. Haptic code design was accomplished through a design program included in the custom MATLAB software. The program stored the set of 39 haptic symbols in a file that could be used by the other programs. Fig. 3.3 shows the interface of the design program. According to Fig. 3.3, the program is divided into four panels labeled by Phonemes, Play, Code, and Signals. In the Phonemes panel, the user is able to see the 39 phonemes as buttons with the custom labels explained in 3.2. In this panel, the design of a phoneme code started by clicking on the phoneme to be designed and by specifying a base duration in milliseconds for the signal.

In the Play panel, the signals for the selected phoneme could be played on TAPS by clicking the "Play" button. The signals played according to a time factor specified at the top of the panel. This parameter controlled the playback speed of the signals to be played, with 1 indicating the unscaled, normal speed. The user is also able to play the signals for all 24 tactors by selecting "All" or play the signals of previously selected tactors by choosing "Selected" (see next paragraph). Furthermore, a sequence of

phonemes could be played by typing their labels as a sequence separated by commas. The sequence played with an inter-code interval specified in milliseconds.

In the Code panel, the 24 tactors are represented by a 4×6 matrix. The tactors in Fig. 3.3 are labeled after the layout specified in Fig. 3.1 to show their correspondence. To assign a signal to a tactor, a user can click on the corresponding cell of the matrix and configure signal parameters from the Signals panel. Signals assigned to tactors could be played individually by using the "Selected" option on the Play panel.

In the Signals panel, a maximum of 3 tones could be assigned to the selected tactor from the Code panel. Every tone is a sinusoidal signal that can be configured in terms of its frequency, starting point in time, duration and amplitude. The starting point and duration are specified as a percentage of the base duration (specified by the user in the Phonemes panel) using a number between 0 and 1. The signal amplitude is specified in sensation level (SL), a dB value above a participant's detection threshold. Finally, tones can be modulated in amplitude by using a cosine squared window or another sinusoidal signal at a lower frequency.

By specifying the amplitudes of signals in units of sensation level (dB SL), the signals were ensured to provide a consistent perceived intensity across participants. In this sense, detection thresholds need to be measured for any participant who would learn the haptic symbols assigned to phonemes. Furthermore, in order to ensure a consistent intensity across tactors, an equalization of the whole array of actuators needs to be performed. Two additional programs were developed to perform detection threshold measurements and tactor equalization. Detection threshold measurements were performed by a program named *TactThresh* and equalization was performed by a program named *TactAdjust*.

TactThresh implemented a three-interval, two-alternative, one-up two-down adaptive procedure. The program allowed the user to select a tactor from the array as the main actuator and configure its duration and frequency. The initial amplitude in dB relative to the maximum output of the MOTU device could also be configured, as well as the initial step size, the number of reversals to perform, a second step size

effective after a certain number of reversals had been achieved, the reversal number at which the step size changed, and the total number of reversals to be collected before terminating the trials.

TactAdjust allowed users to adjust the intensity levels of all tactors relative to a specified tactor. The program implemented a method of adjustment, in which the participant would select a reference tactor and would adjust the intensities of all other actuators until they were perceived as equally intense as the reference stimulus. The duration, frequency and amplitude of stimulation of the reference actuator could be configured by the experimenter.

Lastly, testing and training programs were also developed in MATLAB. These programs used the outcomes of the haptic design program, TactAdjust, and TactThresh to generate customized (in stimulus level) phoneme signals for every participant. Training programs (referred to as *FreePlay* programs) allowed participants to play phonemes or words with visual representations of the phonemes in the tactor array. A 4×6 array of colored circles represented the array and showed duration, frequency, modulation, and motion of the signal associated with a phoneme. Training runs with correct-answer feedback were included in such scenarios. In addition, testing programs performed word-identification or phoneme-identification experiments for a fixed number of trials with no correct-answer feedback. Such tests logged metrics for every test, which included the stimulus, the response times per trial, and percent-correct scores.

3.4 Experimental Procedures

With the required hardware and software and the design of a set of 39 haptic codes, different experiments were conducted to assess the distinctiveness of the stimuli set, as well as the feasibility of the new encoding to convey speech information effectively and rapidly.

3.4.1 Exploratory Experiments

The first exploratory experiments were conducted in [1]. A preliminary pilot study was conducted with one participant who learned phonemes and words through 21 days and 12 experimental sessions (excluding weekends and day breaks). The experiment included 6 phonemes (EE, AY, OO, D, M, S) and 24 words made of 10 CV (consonant-vowel) words and 14 CVC (consonant-vowel-consonant) words created from the same set of phonemes. The study proved that the participant was able to achieve percent-correct scores higher than 90% for all the tests of phonemes and words. This encouraged researchers to design 2 experiments to further investigate the learnability of the haptic codes. Before the experiments began on the first day, threshold detection experiments were conducted and intensity calibration of the device was performed using the MATLAB programs developed for these purposes. Details on parameters used for the TactThresh and TactAdjust programs are described in section 3.4 of [1].

Experiment I in [1] included 4 naive participants who learned 10 phonemes and 51 words in a 6-day experiment. The phonemes learned were EE, AY, OO, D, M, S, W, DH, K and I and the words learned included 23 CV words and 28 CVC words (10 CV and 14 CVC words from the previous pilot study). On every day, participants learned phonemes and words through 5 minutes of practice in a FreePlay program followed by 5 minutes of a trial-by-trial correct-answer feedback test. Percent-correct scores were logged for every test conducted. The daily sessions were divided as:

1. Day 1: Learning 6 phonemes
2. Day 2: Learning 24 words made up of the first 6 phonemes
3. Day 3: Learning of 4 new phonemes and testing of all 10 phonemes
4. Day 4: Learning of 27 new words
5. Days 5 and 6: Learning of all 51 words

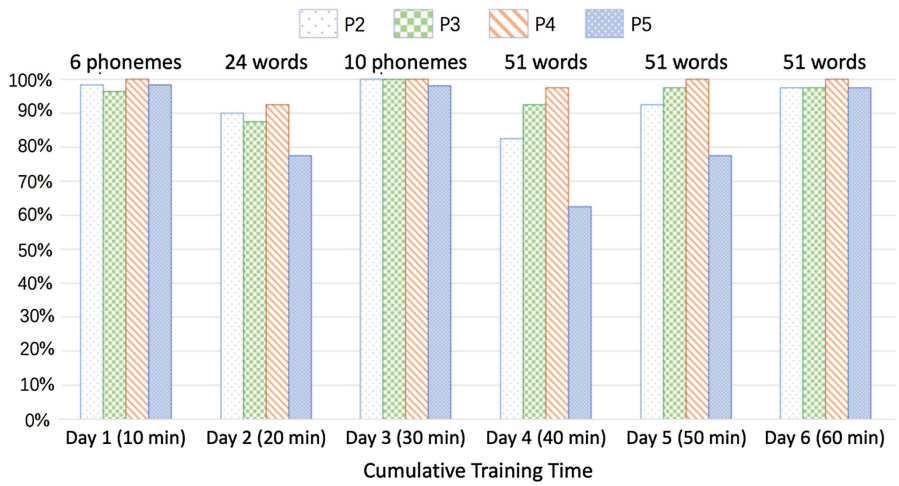


Fig. 3.4.: Results of Experiment I in exploratory studies (taken from Fig. 5 of [1])

The results from Experiment I of [1] are presented in Fig. 3.4. The figure shows that participants were able to recognize up to 51 words within 1 hour of training and were able to achieve a performance higher than 90% on the last day of tests. All learners showed that transitioning from phonemes to words represented a challenge. Nevertheless, all participants were able learn the word sets with increasing performance throughout the days.

A second experiment (denoted as Experiment II in [1]) assessed the memory consolidation theory in learning the full set of 39 phonemes. The theory was expected to potentially enhance the learning process by allowing learned material to consolidate in the memory of participants from one day of testing to another [48]. The experiment also generated insights on the time required to learn the set of codes with a feasible performance level. This experiment tested one naive participant for a total of 14 consecutive days.

The set of 39 phonemes was divided into 8 groups learned cumulatively throughout 8 consecutive days. On every day, the participant was first tested on the set of the previous phonemes before learning a new group. The learning stage was conducted using a FreePlay program. This process was then followed by a cumulative test on all

phonemes learned so far with trial-by-trial correct-answer feedback. The process of testing the previous group, learning the new group, and testing all phonemes learned so far lasted 10 minutes per day. The participant had to achieve a percent-correct score of 90% or higher to move to the next group of phonemes. From day 9 to day 14, the participant was tested on 4 runs of 50 trials each with the full set of 39 phonemes. Percent-correct scores were logged for every test conducted.

The role of memory consolidation was evidenced in the results of Experiment II, where performance on a given set of phonemes tested at the start of a new day always improved or remained the same as the day before. Moreover, the participants from Experiment II demonstrated that the encoding was simple enough to learn between 4 and 6 phonemes every day. On the last 6 days of tests with 39 phonemes, the learner achieved an average performance of 93.8 $\pm$ 3.8% correct score. Individual scores per day can be found in Fig. 8 of [1].

Overall, the experiments in [1] showed that participants can learn the full set of 39 phonemes in less than 80 minutes and a set of 51 words within 60 minutes. The experimental designs based on learning and testing on consecutive days involved memory consolidation and proved that a distributed protocol of spending a short period of time each day over many days enhances learning.

3.4.2 Phoneme-based and Word-based Learning Protocols

With the previous knowledge that the encoding of 39 English phonemes is entirely learnable in a short amount of time, the question on how to effectively train participants to recognize phonemes and words still remained unanswered. Experiments in [2] compared two approaches to this problem: a phoneme-based approach, and a word-based approach. Participants in the phoneme-based approach engaged in a distributed learning protocol throughout several days to learn the set of 39 phonemes prior to learning a list of 100 words. The participants in the word-based approach learned increasing sets of words through several days until 100 words were achieved.

The set of words consisted of 50 consonant-vowel (CV) words and 50 consonant-vowel-consonant (CVC) or vowel-consonant-vowel (VCV) words. Table 1 of [2] shows the list of words divided into 8 groups.

Based on the evidence of memory consolidation presented in [1], the phoneme-based and word-based approaches consisted of a 10-day curriculum with 10 minutes of learning per day. Every day, a test on the cumulative stimulus set was conducted. Prior to the start of the experiments, threshold experiments were conducted and intensity calibration of the tactors was performed using the MATLAB programs used in [1] and described in the previous section. Details on the parameters of the software are explained in section 2.2 of [2].

For the phoneme-based experiments, a total of 12 naive participants learned the 39 English phonemes and 100 words through the following 10-day curriculum taken from [2]:

- Day 1: 6 consonants - P, T, K, B, D, G
- Day 2: 12 consonants - Day 1 + F, V, TH, DH, S, Z
- Day 3: 18 consonants - Day 2 + SH, ZH, CH, J, H, W
- Day 4: all 24 consonants - Day 3 + M, N, NG, R, L, Y
- Day 5: 8 vowels - EE, IH, AH, OO, UU, AE, AW, ER
- Day 6: 15 vowels - Day 5 + AY, I, OW, OE, OY, UH, EH
- Day 7: all 39 phonemes ($>$ 90% correct required before learning words)
- Day 8: 50 VC/CV words if $>$ 90% correct achieved with 39 phonemes
- Days 9-10: all 100 words, after 1 day with 50 VC/CV words

Within the allowed 10-minute training period per day, participants learned the new group of phonemes or words (depending on the day of the curriculum) by repeatedly playing the new stimuli on a specific FreePlay program for this purpose. Additionally,

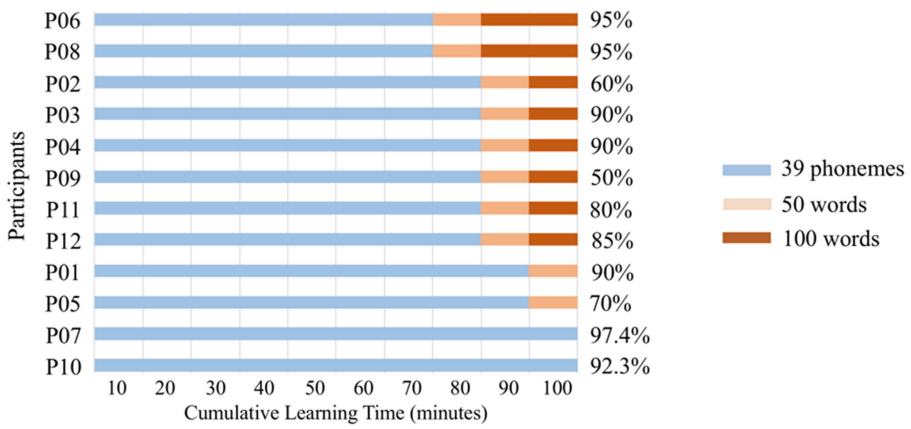


Fig. 3.5.: Progress on participants in the phoneme-based learning protocol (taken from Fig.4 of [2])

participants could answer mock trials of the test with trial-by-trial correct answer feedback. After the 10 minutes elapsed, a 20 trial test without correct answer feedback was conducted. Participants who achieved $> 90\%$ on daily tests were allowed to move on to the next group of phonemes/words. Percent-correct scores were logged for every test conducted.

For the phoneme-based approach, results of the individual progress of participants are shown in Fig. 3.5. The results show the performance level of all participants at the last day of the experiment (Day 10) for the task faced on that day. According to the figure, 8 participants completed the curriculum by learning 100 words within 10 days (100 minutes). For the other 4 participants, 2 learned only 50 words and 2 learned the set of 39 phonemes but couldn't be tested on words.

Overall, participants were able to learn the set of 39 phonemes with a $> 92\%$ accuracy within 100 minutes. Moreover, the average performance of participants who learned the sets of 100 and/or 50 words was $> 80\%$ at the end of the 100-minute period.

For the case of the word-based approach, a new group of 12 naive participants took part in a 10-day curriculum to learn sets of words with increasing difficulty (both

in terms of number of words and phonemes included in the words). The curriculum divided the set of 100 words into 8 groups as described in Table 1 of [2]. The daily procedure based on groups designed in [2] can be summarized as follows:

- Group 1: 13 words from 6 phonemes
- Group 2: Group 1 + 13 words = 26 words from 10 phonemes
- Group 3: Group 2 + 12 words = 38 words from 15 phonemes
- Group 4: Group 3 + 13 words = 51 words from 20 phonemes
- Group 5: Group 4 + 12 words = 63 words from 25 phonemes
- Group 6: Group 5 + 12 words = 75 words from 30 phonemes
- Group 7: Group 6 + 13 words = 88 words from 35 phonemes
- Group 8: Group 7 + 12 words = 100 words from 39 phonemes

As in the phoneme-based approach, the 10 minutes of training were divided by the participants into free exploration of the stimulus set and trials of a mock test with correct-answer feedback. Participants were required to achieve a $> 80\%$ performance on the 20-trial daily tests with no correct-answer feedback to move on to the next group. Percent-correct scores were logged for every test conducted.

Results for the word-based approach are shown in Fig. 3.6. The figure places a red cross on the day that a participant reached the 80% criteria to move on to the next group. The results show that performance varied greatly among participants and that there is a noticeable gap between the top two performers (P13 and P17) and the rest of the group. Overall, participants took 21.5 minutes on average to learn and pass each group [2].

An additional performance metric derived for this approach was the "equivalent number of words correctly identified", which was calculated as the percent-correct score of a test multiplied by the number of words included in the group of the test.

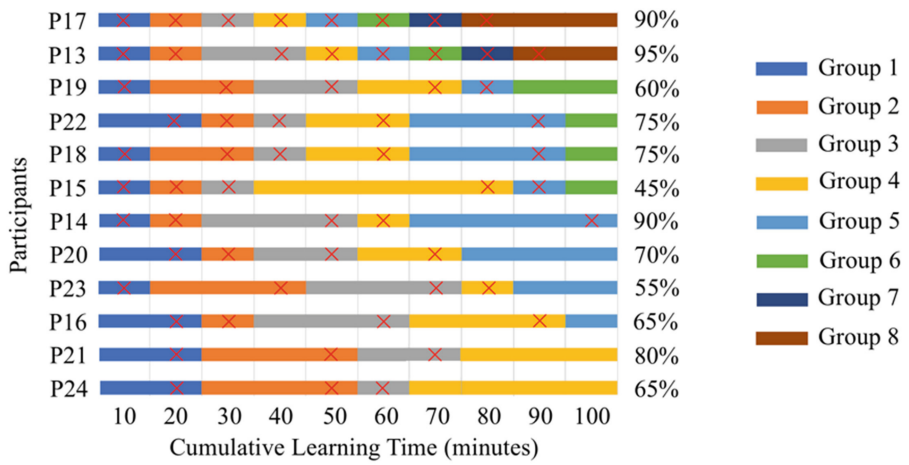


Fig. 3.6.: Progress on participants in the word-based learning protocol (taken from Fig.6 of [2])

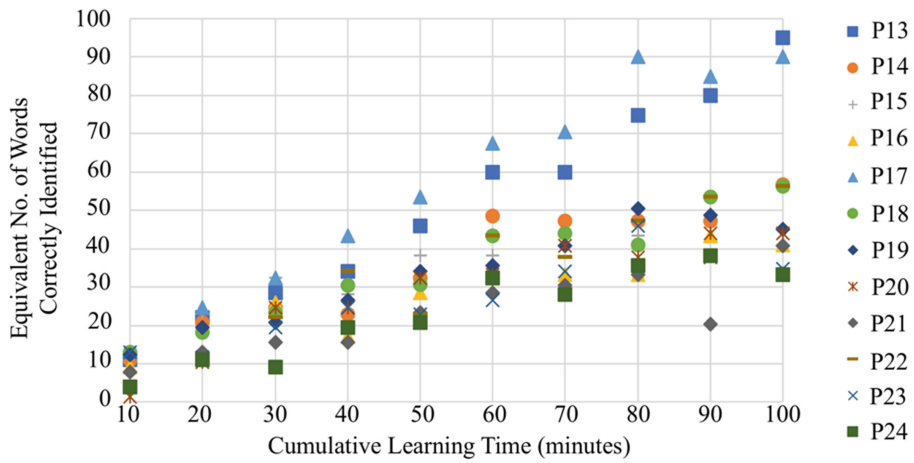


Fig. 3.7.: Equivalent number of words correctly identified in the word-based approach (taken from Fig.7 of [2])

Fig. 3.7 shows this metric for every participant as a function of the cumulative learning time. This metric shows how the two top performers can learn words at roughly 1 word per minute in a continuously increasing pattern, whereas the rest of the group plateaus at roughly 50 words by the end of the experiment.

In general, the studies performed in [2] demonstrated that participants are able to learn the set of 39 phonemes coded with the design rules described in the previous section, verifying the distinctiveness of the haptic codes. By examining the performance of participants that achieved 100 words on both approaches and the number of participants that reached this level, it was concluded that the phoneme-based approach is a safer and more reliable strategy to recognize words after sufficient training on phonemes [2].

3.4.3 Phoneme Identification Experiment

Parallel to the studies of [1, 2], a phoneme identification experiment with 10 participants was conducted in [3]. The experiment followed a similar procedure which started with the measurement of the detection threshold of participants with the TactThresh program, and the calibration of the amplitudes of the tactors using the TactAdjust program. The configuration parameters for both programs are described in sections 4.4.1 and 4.4.2 of [3].

As in previously described studies, the 39 phonemes were learned by participants following a distributed paradigm with increasing sets of consonants and vowels. The participants ran learning and testing procedures on 8 groups of phonemes. As presented in Table 4 of [3], the groups of haptic symbols were:

- C1: 6 consonants - P, B, T, D, K, G
- C2: 12 consonants - C1 + F, V, TH, DH, S, Z
- C3: 18 consonants - C2 + CH, J, SH, ZH, H, W
- C4: 24 consonants - C3 + M, N, NG, R, L, Y
- V1: 6 vowels - EE, AH, OO, AE, AW, ER
- V2: 10 vowels - V1 + UH, IH, EH, UU,
- V3: 15 vowels - V2 + AY, I, OW, OE, OY

- CV39: C4 + V3

For every group, participants started a session through a learning stage that included the use of a FreePlay program to learn phonemes. The program allowed participants to play tactile signals from a given group and were able to see a visual representation of a selected phoneme.

The learning stage was followed by an identification paradigm that consisted of a one-interval, forced-choice identification test with trial-by-trial correct-answer feedback. On every trial, a randomly selected phoneme was presented and the participant's task was to select one of the alternatives from the set being tested. Upon error, the correct and incorrect selections were highlighted and the participant was given the choice of comparing both stimuli by individually playing and/or viewing a visual representation of the signals. An unlimited time for replaying the stimuli was given. For sets C1, V1, and CV39, stimuli were selected randomly with replacement. For all other sets, half of the trials were devoted to new phonemes added to the set, and the other half to previously-learned phonemes. The number of trials in the test increased according to the number of symbols in the set and was conducted until performance in the range of 80-90% was accomplished.

Following this identification paradigm stage, a testing phase was conducted. The same identification paradigm was used except that no feedback was provided. The test was conducted on all participants for group CV39. At least two runs of 78 trials each were collected per participant. On every run, each of the phonemes was presented twice in a random order. Stimulus-response confusion matrices were constructed for every test to derive percent-correct scores and to analyze the confusion patterns between pairs of phonemes. Additionally, the participants' response times were recorded for every test trial in a run. This time was counted as the duration between the offset of a stimulus and the onset of a participant's response.

Results from [3] show that performance varied among participants. Out of the 10 participants tested, 7 achieved percent-correct scores between 85 and 97% on the CV39 tests, while the remaining 3 scores between 71 and 76%. The total time required

to complete the training ranged from 50 to 230 minutes. Analyses on the participants' behaviour on the training protocol revealed that playing phonemes on the tactile device was a more popular interaction than viewing their visual representation on the visual display.

For the tests conducted on the CV39 group, the two top-scoring test runs per participant were pooled into a confusion matrix to construct a data set of 1560 trials. In the matrix, an entry m_{ij} shows the number of trials that the phoneme i was identified as phoneme j by the participant. The overall percent-correct score derived from this matrix was 85.77%. A visualization of this matrix further revealed that, with errors per haptic symbol (the percentage of off-diagonal entries of a given row calculated as $e_i = 100 \times \sum_{j=1, i \neq j}^{n} m_{ij} / \sum_{j=1}^{n} m_{ij}$) in the range between 7.5% and 25%, confusion patterns existed within vowels and consonants but never between the two groups.

For the case of vowels, confusions were below 12% and the majority of the confused pairs consisted of stimuli with the same duration and direction of movement but different properties of location and type of movement. Two specific cases of confused pairs were UH-EH and OW-OY that shared the same type of movement and duration but different directions and locations in the array. Confusions among consonants were below 17.5% and included several clusters of confused phonemes with different patterns. Confusions of modulation within phonemes of the same duration and location were evidenced (e.g., P-B, T-D, K-G), as well as errors between phonemes of different locations and frequencies (e.g., TH-N, DH-N).

Finally, response time were analyzed as a function of the number of symbols in the stimuli set. The mean response time across the 10 participants are shown in Fig. 3.8. The figure shows how response times tend to linearly increase with the size of the stimuli set when the number of phonemes is plotted on a log scale.

Overall, participants in the experiments of [3] demonstrated the learnability of the haptic codes with a high accuracy and a reasonable amount of training time (from 1 to 4 hours). The design of the haptic symbols was effective in the sense that major

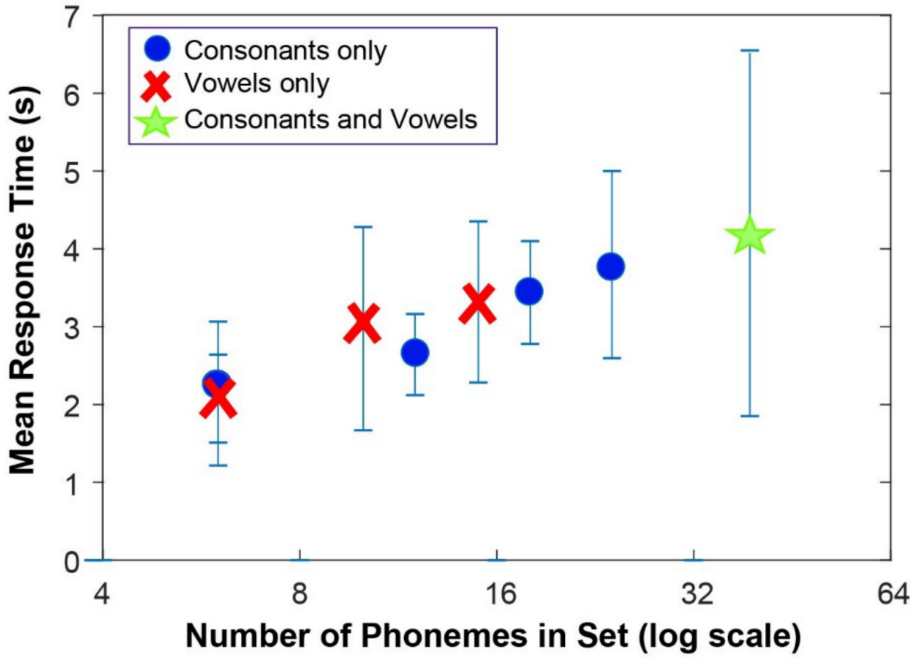


Fig. 3.8.: Mean response times of participants in phoneme identification experiments (taken from Fig.7 of [3])

classes of phonemes were not significantly confused, i.e, consonants and vowels are not confused with rates higher than 7.5%. The accuracy achieved by participants represent a promising direction for the use of the 24-tactor array and the encoding of the 39 English phonemes in real life conversations.

4. THE "HAPTOS" GAME

4.1 Introduction

The studies reviewed in Chapter 3 proved that the design rules to encode 39 English phonemes into tactile signals were effective in producing high recognition accuracy in phoneme identification experiments with a reasonable learning time. Furthermore, evidence of memory consolidation proved that a distributed learning paradigm through several days is an effective strategy in reducing the total time devoted to explicit learning. Moreover, [2] proved that a training process on phonemes prior to words is a more reliable strategy to achieve a better and more robust performance on word identification tasks.

The daily sessions on all the experiments relied on the simple process of reviewing and memorizing the tactile patterns. Training and testing procedures in the designated MATLAB software helped participants to visualize phonemes and words as they were played. Nevertheless, participants hardly used the visual representation of the tactor array to learn the different tactile patterns. In fact, participants from the experiments in [3] used the "Play" option (to feel the haptic symbols) of the FreePlay program 16.5 times more than the "Show" option (to see the visual rendering of the haptic symbols) in the training. In this regard, the learning mechanism used on all procedures of [1–3] consisted of text-based interfaces and lacked engagement tactics and effective mnemonic aids to help participants associate tactile patterns and phonemes in a more robust way.

A possible solution to such a problem is to broaden the scope of the project and view the learning process as similar to that of acquiring a new language (a *haptic language*). The *gamification* approach is a possible solution to creating a more engaging learning environment that is conducive to language acquisition. This method

relies on the use of serious games to implement second language acquisition theories in scenarios that balance learning and entertainment [39–41]. One strategy adopted by such games is to embed communication skills into the skill profile of a virtual avatar that represents the learner [41]. As long as the skills in the virtual world are related to the skills in the real world, the increasing level of virtual skills will result on the development of language skills. Such improvements can be task-based, in which actions to solve puzzles or talk to virtual characters are means to improve the learner's skills. Furthermore, context-based scenarios provide language-learning opportunities based on a specific subject that the task involves [40]. For example, talking with a virtual character about the weather results in an improvement on weather-related words.

A role-playing game provides a suitable structure to implement this strategy. In role-playing games, the learner is represented by a virtual avatar immersed in a digital world. The game follows a storyline in which the avatar engages in different adventures that improve the language abilities by solving different tasks. Using this design, second language learning games immerse the learners in a virtual world where they have to learn and use the second language in order to correctly follow the storyline, improve the avatar's skills and ultimately finish the game with a high degree of proficiency in acquiring the second language.

This chapter describes the design and exploratory evaluation of "Haptos", a role-playing game that teaches participants how to understand the "haptic language" −English words represented by the concatenation of phonemes in tactile patterns. A general design framework for this class of games is presented, as well as a preliminary implementation of the Haptos game, and an exploratory evaluation of the game with two participants.

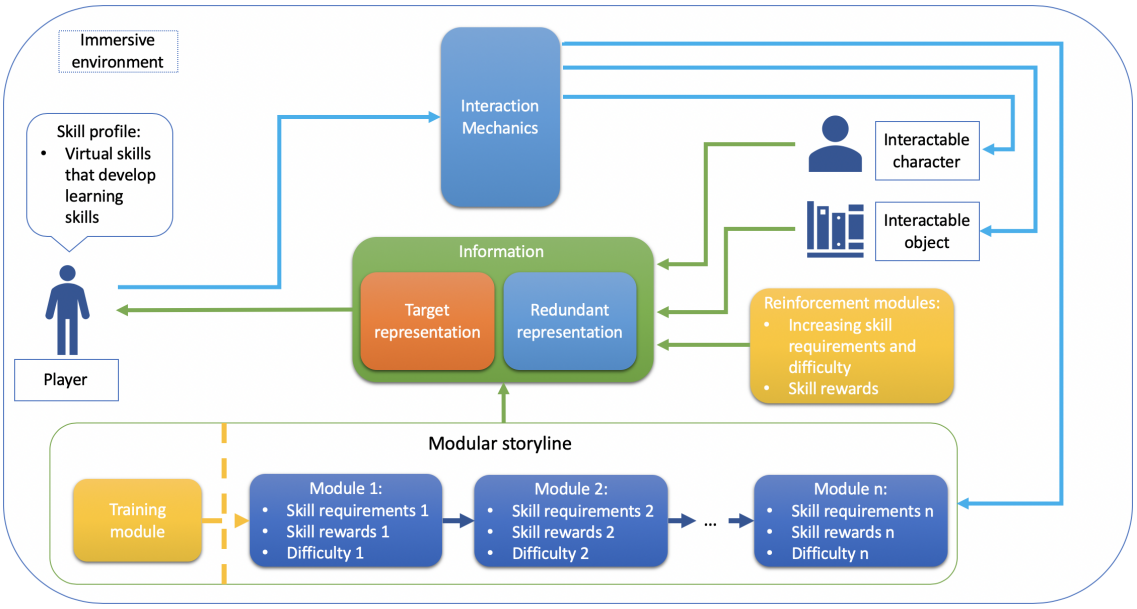


Fig. 4.1.: General framework for role-playing game design in language learning

4.2 Proposed Framework

The design of a role-playing game to address second language learning for the haptic language started from a broader point of view. A design of a general framework for role-playing games with the purpose of language learning was proposed and is presented in Fig. 4.1. The framework is based on representing the user with a virtual avatar. As specified in [41], the character has a skill profile with components that are directly related to the skills to be developed in the real world. In the context of language learning, communication skills should be mapped directly to the avatar's skills.

The avatar is immersed in a virtual world and is surrounded by the target language to be learned. This scenario is composed of interactive virtual characters and objects that exchange information with the learner in the second language. All information delivered to the learner is represented by a packet that has two main components: a target representation, and a redundant representation. The target representation refers to the target modality in which the second language is transmitted. For the

case of the haptic language, this representation refers to the tactile modality. The redundant representation refers to the same information transmitted in the target representation but in a different sensory modality (e.g., visual and/or auditory cues). The use of redundant information supports the association of knowledge between an acquired language, and its representation in the second language. It also provides mnemonic mechanisms to learn new words in the target language.

Furthermore, interactions are mediated by the interaction mechanics of the game. This module defines rules based on human-computer interaction that implement the "simple, natural, friendly and consistent" principles defined in [49]. These principles state that the interaction between the learner and the game should be as simple as possible, be in line with the player's life experience and cognitive abilities, generate an output that is helpful to the users' understanding, and maintain a consistency between the player's input and the game's output.

By defining an immersive scenario composed of characters, objects, interaction mechanisms, and a structured representation of information, the core game narrative can be developed based on a modular storyline that supports incremental skill development. The storyline is divided into modules with increasing difficulty and requirements. Such requirements are based on the player's profile. For example, a player attain a certain level associated with a communication skill that is then represented by a numeric value. Then, every module will require the learner to achieve a certain level on that specific skill to allow the learner to enter the module. Every module will represent a challenge to the learner in terms of a language-based task. After overcoming the challenge, the skill profile of the learner would be incremented following the game rules. The learner will then be allowed to enter the next module and progress on the storyline.

Additional modules are also included in the design. First, an optional training module is added to the beginning of the storyline and is available to the learners at all times. This element provides repetitive training on the learning material with no skill requirements or rewards. The learner is allowed to use the training module

with no time limitations. Therefore, the module is meant to be used as a self-paced learning mechanism. Furthermore, reinforcement modules are added to the virtual environment. These modules are not part of the storyline as they don't hold any sequential relation with the storyline plot. They are meant to be used as additional training that encourages the learner to gain extra experience. Unlike the training module, reinforcement modules have increasing difficulties according to the current learner's skill profile. They also offer rewards that may help learners progress on the storyline modules. For example, a learner may pass a module i but the rewards from that module may not be sufficient to enter module $i + 1$. The learner will then be encouraged to try reinforcement modules until the requirements for module $i + 1$ are met.

4.3 Haptos Design

With the proposed framework in Fig. 4.1, a role playing game that supports haptic language learning was designed as an "instance" of the framework. The game was named "Haptos" and its proposed design is shown in Fig. 4.2. The game develops in the context of a fantasy village named "Haptos". The village is home to a special type of villagers that have the ability of communicating through the skin by using the haptic language. Unfortunately, the village is under the control of an evil corporation that performs experiments on villagers to exploit and reproduce their communication abilities in other human beings. In this scenario, the participants are represented by Nathan, the main character of the game. Nathan is a lost scientist who found the village by accident and is now trapped inside by the ruling entity. The goal of the participants is to guide Nathan through different quests to free the villagers and return home.

The design specifies a skill profile composed of three elements: an "arm" level, an inter-code interval (ICI), and a certain number of "quest points". The arm level is related to the number of phonemes and words that a participant is able to recognize at

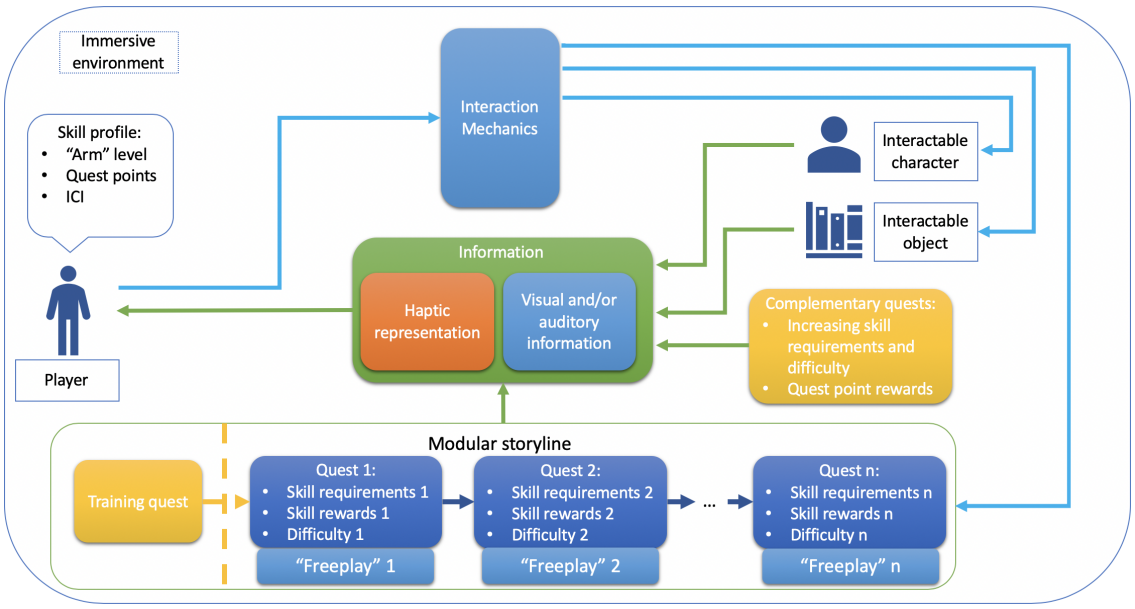


Fig. 4.2.: Haptos design in terms of the role-playing game framework for language learning

a certain point. This level increases as the participant passes several tests evaluated in the storyline modules (named as quests in the design). The inter-code interval is a specified time interval in milliseconds between individual phonemes in the phoneme sequence that compose a word. Finally, the quest points are required points to "buy" storyline quests. Players have to acquire a certain number of points specified by each quest in order to try to solve the quest.

The quests are the main building blocks of the storyline. All quests have to be unlocked by solving a small puzzle prior to entering the quest. Every quest involves a "mini game" in which the users have to recognize different sets of words in the haptic language in order to obtain quest points and learn the language. Within a quest, the user is allowed to enter a "Freeplay" mode in which a set of words can be freely explored by the player without time limitations. The words correspond to new testing words introduced in the quest. Whenever ready, the player would enter a "Solve" mode in which a cumulative word identification test without correct answer

feedback would be conducted. The test would include all testing words learned in the past and the new words introduced in the quest. Players would pass the quest by reaching a specified percent correct score in the solve mode test. Upon success, the players are rewarded with a certain amount of quest points but the accumulated points would be decreased by the amount of points required to enter and solve the quest. Also, the arm level would be increased by 1 unit.

The village also incorporates a training quest that allows players to freely explore phonemes and words. All training material is available to the players regardless of their arm level. The training quest also includes a testing mode in which participants could try a mock test on all the learning materials without correct answer feedback. This quest allows participants to train for new quests and to reinforce learned skills before solving storyline quests.

Furthermore, the village features a set of complementary quests. These quests include smaller "minigames" with a solve mode that includes the cumulative learning material learned by the player up to the avatar's arm level. By solving these smaller games, the participants acquire additional quest points needed to enter new storyline quests and progress in the game. This is a mechanism of reinforced practice between storyline quests.

Finally, the village also includes different virtual objects and characters that would establish a one-way communication with the player using the haptic language. In order to find quest locations and explore the village, the player would read dialogues produced by virtual characters or objects with a certain number of missing words. Such words would be played as sequences of phonemes to be recognized by the participants.

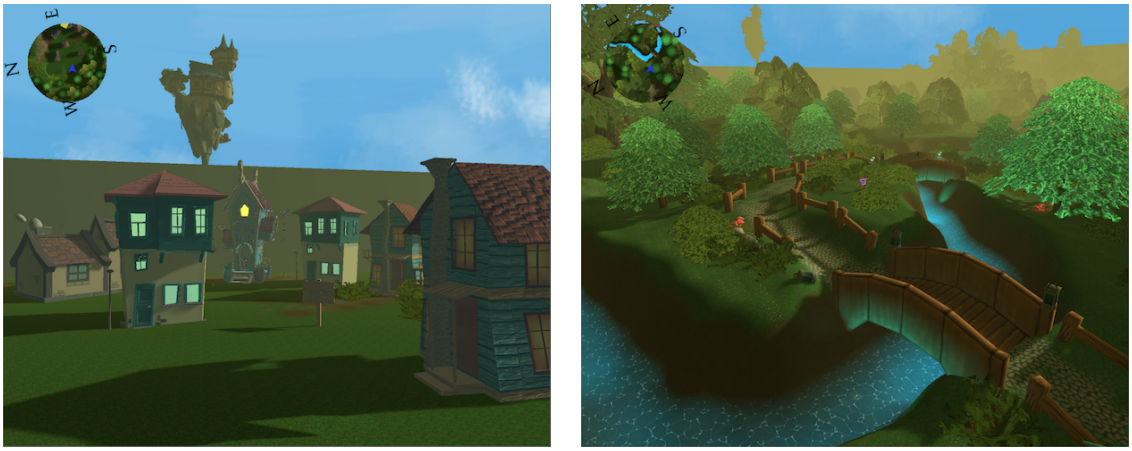
All these quests and objects create an immersive environment in which the player is surrounded by the haptic language. The game would then require the player to learn the language in quests and contextual scenarios in order to correctly follow the storyline.

4.4 First Version of Haptos

A first version of the Haptos game was implemented using the Unity 3D game engine (https://unity3d.com/) in the Windows 10 platform. All the hardware components used in the game to deliver tactile stimuli were the same as those described in section 3.1 of Chapter 3. In order to use this hardware in a Unity application, a special library written in C/C++ code was developed. The architecture of the library is described in Appendix C.

The game was designed to train participants on learning material that consisted of all 39 phonemes and a total of 200 words. This material was divided into 8 groups learned by participants through 8 storyline quests. Each group consisted of a set of 4 to 6 new phonemes, 12 to 14 training words, and 12 or 13 testing words. All words were chosen to be consonant-vowel-consonant, vowel-consonant, or consonant-vowel words. The groups are described in Table. B.1. The material was divided into training and testing sets in order to train participants with a certain degree of generalization. In this regard, the training words were available to users in the training quest and in the complementary quests, but the testing materials were only available in the storyline quests and were used to perform word identification tests without correct-answer feedback that determined if a participant passed or failed a quest.

For the first pilot study of the game, only the first 4 groups of words and phonemes were implemented. This includes a total of 12 consonants, 8 vowels, 49 training words, and 52 testing words. Furthermore, the skill profiles of the participants included the arm level and an inter-code interval time of 150 ms. This interval was kept constant throughout the game and was not considered as an additional difficulty setting for the quests. In other words, the inter-code interval did not decrease as participants progressed in the game. Moreover, no visual and/or auditory information was presented along with haptic information. Therefore, only tactile cues were conveyed to observe



(a) First version of the Haptos village

(b) Forest area of the village

Fig. 4.3.: First version of the Haptos game

how well participants learned the material and how the gaming paradigm helped the process of learning.

The 3D game was constructed using several scenes loaded from a main village scenario. This map allowed participants to freely explore the village and the surrounding forest using a first person controller. Fig. 4.3 shows a capture of the village and the forest that contained all the quest locations.

In the village, interactive virtual objects were marked with a hand symbol. The player approached these objects to initiate different types of interaction. If the player approached a quest, the quest scene would be loaded. If they interacted with an object for information, a dialog would open with important information and missing words that could be delivered as tactile stimuli. In order to navigate in the game, the participants controlled the first person controller with the W-S-A-D keys on a keyboard and the computer mouse. Locations of special quests were marked on a map that participants could access using the E key. Fig. 4.4 shows a screen capture of the map with different locations. The cyan squares marked quest locations, the pink squares marked interactive objects that included puzzles to unlock quests, the

Fig. 4.4.: Map of the first version of the Haptos game

red squares marked complementary quests, the yellow square marked the training quest, and the blue arrow represented the user.

The training quest was named in the game as the "Training School". This scenario was available to players at all times and was built as a dark room that included 3 forms of interaction: a "Freeplay" mode of phonemes, a "Freeplay" mode of words, and a cumulative training test. The Freeplay modes allowed participants to explore phonemes and words from the training material by playing individual words or phonemes. The cumulative training test was a 20-trial, 3-alternative word identification test without correct-answer feedback, implemented as a cooking game. Fig. 4.5 shows captures of the Freeplay modes and a trial of the training test.

When entering the school, the participants could enter the different options by click on the menu options of Fig. 4.5(a). "Phoneme Play" opened the Freeplay mode of phonemes, "Word Play" opened the Freeplay mode of words, and "Test your Skills" loaded the training test scene.

(a) Main menu of the Training School

(b) Freeplay mode for phonemes (the phoneme OE was clicked)

(c) Freeplay mode for words (the word "wide" was clicked)

(d) One trial of the cumulative training test

Fig. 4.5.: Training School in the first version of the Haptos game

In the Freeplay mode of phonemes, the participants could navigate through the phonemes in the learning materials and click on phonemes to play them on the TAPS device. Playback of the signals was accompanied by an animation of the phoneme on a virtual arm (see the 3D arm to the left of Fig. 4.5(b)). This animation showed the movement of every vowel as well as the animation of modulation for consonants. Animations followed a color coding based on the location of stimulation and the frequency of vibration of each stimulus. Animations for consonants and vowels can

be viewed at https://youtu.be/Fr0-XucKGEY and https://youtu.be/CYfqcdnvMyE, respectively.

In the Freeplay mode for words, participants could view the groups of training words and freely click and play the different words with a predefined inter-code interval time of 150 ms. When clicking a word, the phoneme transcription of the word was displayed at the bottom of the screen (see Fig. 4.5(c)). The training school used all the training words in Table. B.1.

Finally, by entering the cumulative training test, the participants faced a 20-trial word identification test with no correct-answer feedback. As shown in Fig. 4.5(d), every trial would play a word randomly from the full set of training words. Randomization was performed with replacement. Upon playing the word, the participant was prompted with 3 possible responses, of which the two incorrect choices always maintained a degree of similarity with the correct response. For example, a stimulus for the word "Say" is similar to "Seam" and "Sigh" because all the words start with the same phoneme S. The words were presented with vegetables unrelated to the words. When clicking a response, the corresponding vegetable would enter the pot as part of a recipe. The test informed the participants that they passed the test if their percent correct score was at least 90%, in which case a completed dish would be presented as the result of the recipe. Regardless of the test result, the score was shown at the end of the test. Participants could train in the training school without time restrictions before trying to solve a storyline quest.

In storyline quests, participants were required to approach the pink squares in the map shown in Fig. 4.4 and interact with a hand symbol. The symbol triggered a puzzle game that "unlocked" the quest after being solved. Fig. 4.6 shows an example of the hand symbol and the puzzle presented to unlock quest 2. The puzzle required participants to read a text in the screen that explained the storyline and how the mini-game in the quest contributed to the plot. In the text, some of the words were shown missing and replaced by buttons. When clicking on the buttons, participants felt the missing words on their arm through TAPS and their task was to type the

(a) Hand symbol to open the quest puzzle

(b) Quest puzzle for quest 2

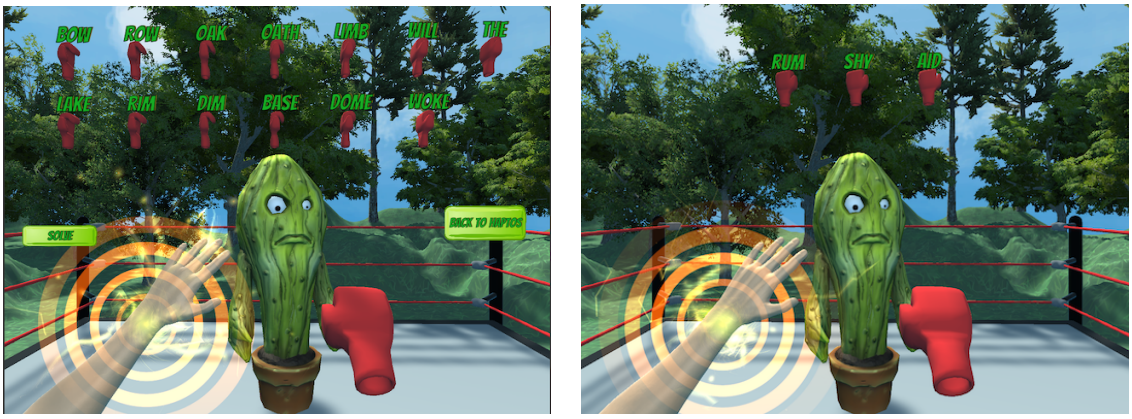
Fig. 4.6.: Quest puzzle to unlock quest 2 in the first version of Haptos

missing words in the designated blanks. These words were taken from the training words of the group that the quest belonged to.

Puzzles allowed participants to understand words based on the text context that surrounded the word. Although they represent a very informal experience, these interactions contributed to the learning experience of the participant through contextual cues that accompanied the tactile stimuli. The puzzles are examples of the interactive objects described in the proposed framework, whose purpose is to aid language learning in contextual scenarios.

Upon correctly identifying the missing words, the quest would be unlocked and showed on the map as a cyan square. Participants would then be allowed to approach the quest and review the quest requirements to enter the mini-game. Quests had different quest point requirements as well as a required arm level. If a participant met the necessary condition, s/he was allowed to enter the quest mini-game.

These mini-games were designed to contain two modes of interaction: “explore mode” and a “solve mode”. As an example, Fig. 4.7 shows the explore mode and solve mode of quest Q4, which was based on a boxing game. The explore mode refers to the “Freeplay” module attached to every quest in the design of Fig. 4.2 and allowed



(a) Explore mode

(b) Test trial in the solve mode

Fig. 4.7.: Quest 4 in the first version of Haptos

participants to feel all the new testing words included in the quest (words described in Table. B.1). The solve mode featured a 20-trial, forced choice, 3-alternative cumulative test on all the testing words learned on previous quests, including the ones from the quest that was about to be solved. Every quest was a mini-game with a specific theme that fitted into the story plot explained by the quest puzzle.

As in the mock test implemented in the training school, every trial of the solve mode test showed a correct word and 2 incorrect alternatives that hold a certain degree of similarity with the correct word. In this regard, the software tried to present words that coincide in at least one phoneme. This similarity created an additional difficulty in the test, added to the fact that the test included all testing words learned in previous quests. In the specific example of Fig. 4.7, the participant would punch the cactus character after selecting the correct response, but the cactus would punch the participant's avatar when an incorrect response was selected.

Participants passed the quest if a percent-correct score of 90% or higher was achieved in the test. In that case, the participant's arm level was increased by 1 unit, a certain amount of quest points were rewarded to the participant, and the participant's accumulated points decreased by the amount required to enter the quest.

(a) Word presented in a trial

(b) Trial possible responses

(c) Incorrect answer

(d) Correct answer

Fig. 4.8.: Fishing complementary quest in the first version of Haptos

In many cases, the net amount of points left after solving a quest were not sufficient to enter the next storyline module. Therefore, participants were required to solve one or more complementary quests to earn the necessary points.

Complementary quests were scattered throughout the map and were marked as red squares. They could also be accessed by participants at any time and included a single solve mode within a mini-game that was not related to the storyline plot. In the first version of the game, a fishing mini-game was implemented as the unique complementary module on the map. Fig. 4.8 shows screen captures of the fishing game that presented random words in a 10-trial test with 3 possible answers.

As in all the tests designed, the alternatives presented had a certain degree of similarity between words, where the majority of the options had at least one phoneme in common. In Fig. 4.8, the alternatives corresponded to 3 fishing rods in a bridge located on top of a river. If a participant selected an incorrect answer, an angry-looking fish was caught in the selected rod (see Fig. 4.8(c)). If a correct answer was selected, a blue and calm fish was presented as the caught fish (see Fig. 4.8(d)). The test ran for 10 trials and the participant passed the test with a score of 90% or higher.

In general, complementary quests were meant to be used as an extra source of quest points. Each quest rewarded participants with quest points equal to 100 points multiplied by the participants arm level. The number of times that participants had to solve the fishing complementary quest depended on the requirements of the next storyline quest to be solved, and the points rewarded by the recently passed quest. Table. 4.1 shows a summary of the required points for a quest, points rewarded by a quest, the net points left to the user after passing a quest, and the arm level of the user after passing a quest, for the first 4 storyline quests implemented.

Table 4.1.: Summary of quest points in the first version of Haptos

Quest	**Required Points**	**Rewarded Points**	**Net Points**	**Level after Quest**
Q1	100	200	200	2
Q2	400	200	200	3
Q3	500	300	300	4
Q4	1100	400	400	5

According to the requirements of the first 4 quests in Table. 4.1, the first version of the game required the participant to solve at least one complementary quest before being able to enter a storyline quest up to quest Q3. When passing Q3, the participants had to solve at least two complementary quests to enter quest Q4.

4.5 Exploratory Tests

4.5.1 Procedure

With the first implementation of the game, an exploratory study was conducted to assess the behaviour of participants in the game and informally report its effectiveness in teaching the haptic language. The main objective of the study was to discover how participants made use of the training school and complementary quests in order to overcome the challenges of the first four quests, which included a total of 12 consonants, 8 vowels, 49 training words, and 52 testing words.

Two participants were recruited for this study. The first participant (P01) was a 23 year-old male undergraduate student at Purdue University. The participant is a native English speaker who speaks Arabic and French as second languages. The second participant (P02) was a 37 year old male graduate student at Purdue University. He is a native Korean speaker who learned English at 14 years of age as a second language.

At the beginning of the experiments, detection thresholds for both participants were measured using the TactThresh program with a three-interval, two-alternative, one-up two-down adaptive procedure. The configurations of step sizes, number of reversals and electrical characteristics of the stimulus used were the same as in [3] and the past studies reviewed in Chapter 3. Followed by the threshold measurements, participants equalized the intensities of all the tactors using the TactAdjust program with all the program settings from the aforementioned studies.

After completing the initial measurements and equalization, the participants were informed about the purpose of the study, the game mechanics and the main goal of the game. All the necessary explanations about the use of the training school and the quests were given so that participants could use the game independently.

The participants played the game continuously until they passed the first four storyline quests in a self-paced manner. No time limit was imposed on any quest during either training or testing and the participants were not interrupted at any

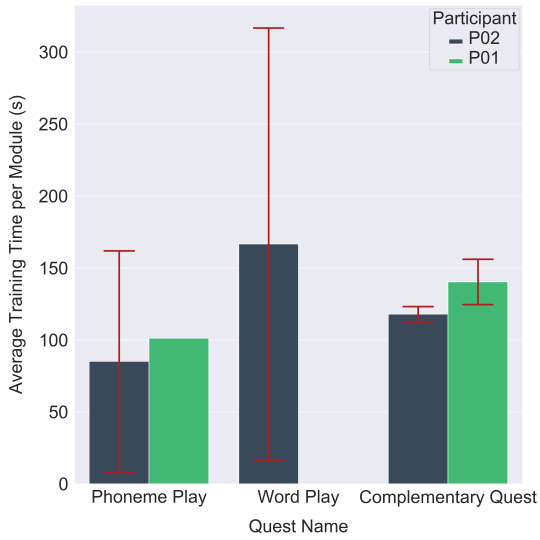
point while they played. A few interventions were made when a participant needed a detailed explanation about an interaction or had any questions about a quest.

In terms of data analysis, a detailed report of the behaviour of the participants was kept. The game was also configured to log the progress of the users in terms of their cumulative training time, time spent in the training school, time spent in the two modes of every quest, and performance on all the storyline quests. In the analysis, the total time spent in the Freeplay modes of the school and the time spent in the complementary quests were added to the cumulative training time of the participants. No phoneme identification experiments were conducted in the first version of the game.

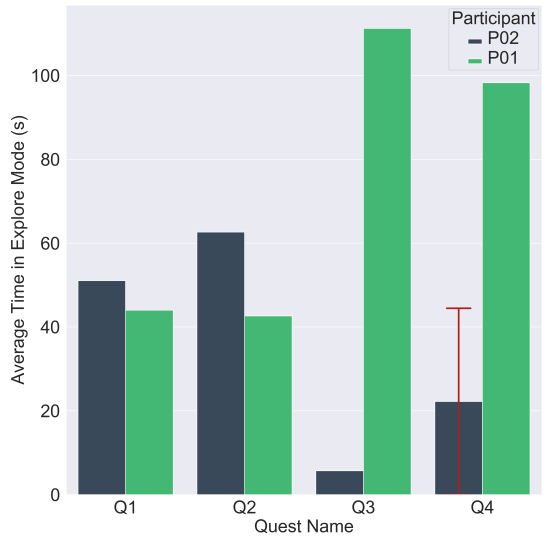
4.5.2 Results

The participants demonstrated that the game dynamics were easy to learn. Participants understood quickly how the training words in the school were related to the puzzles that unlocked quests. Furthermore, they were able to finish all four quests easily within one hour. The game mechanics proved to be highly learnable, both in the navigation aspects and in the tasks to be performed on the quests. Participants were also constantly aware of their progress as they monitored their arm level and quest points. They were able to understand the requirements of storyline quests and the amount of times that the complementary quest had to be solved in order to enter a specific module.

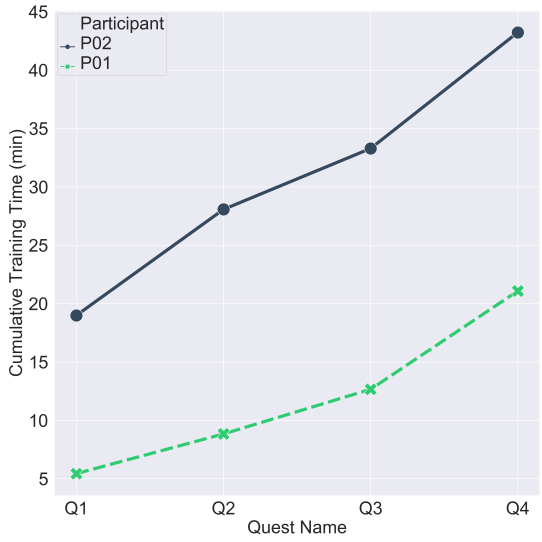
The training profiles for each participant were derived from different training time plots shown in Fig. 4.9. Fig. 4.9(a) shows the learning time in training scenarios (scenarios are: the Freeplay mode for phonemes (Phoneme Play), the Freeplay mode of words (Word Play), and the complementary quest), Fig. 4.9(b) shows the average time spent on the explore mode of all the quests, and Fig. 4.9(c) shows the cumulative learning time at the points where the different quests were solved. According to the average training time spent on the three scenarios in Fig. 4.9(a) and the training



(a) Average time spent on training modules in the first version of Haptos. Red error bars display a ± 1 standard deviation. The data is averaged across all cases in which each participant used one of the three training scenarios



(b) Average time spent on quest explore modes in the first version of Haptos. Data is averaged across all cases in which participants used the explore mode of every quest. Red error bars show a ±1 standard deviation



(c) Cumulative training time

Fig. 4.9.: Training time profiles of participants in the first version of Haptos

time spent in explore modes in Fig. 4.9(b), participants showed different training strategies. Participant P02 mostly focused on learning the words and discovering the phonemes through playing words in the school, whereas participant P01 was only interested in learning the individual phonemes and used the complementary quest and the explore mode of the storyline quests to learn the words. This is further evidenced in Fig. 4.9(b) which shows how P01 spent significantly more time than P02 on the explore mode in quests Q3 and Q4. Moreover, Fig. 4.9(c) shows a significant difference between the learning curves. Participant P01 didn't require a significant amount of training to master all quests, but participant P02 needed repeated interactions with training words in the training school in order to familiarize himself with the phonemes. This can presumably be explained by the fact that P02 reported difficulties to associate the sequences of phonemes with their corresponding words. Nevertheless, both participants were able to complete the quests with less than 45 minutes of training.

In terms of performance results, Fig. 4.10 shows the percent correct scores of both participants for the four quests. The data show the results for all successful quests (quests where the participants achieved a $> 90\%$ percent-correct score). As shown in Fig. 4.10, the participants were able to master the sets of words with performance levels higher than the minimum criteria on most of the cases. It is worth noting that every test had a chance level of 33.3% and that even though the criteria was relatively high (90%), only P02 had to repeat one of the quests (Q4) once before successfully passing the corresponding test.

To further analyze the performance, an additional plot of the equivalent number of words correctly identified was derived from the data. As described in [2], this metric is obtained by multiplying the percent correct score of a test by the number of words in the stimuli set. Fig. 4.11 shows this result for the four quests tested. A first assessment of the data shown on Fig. 4.11 fitted a simple linear regression to both of the data sets of the participants. The regressions revealed that P01 learned roughly 2.15 words per minute and P02 learned 1.69 words per minute. The linear impact of

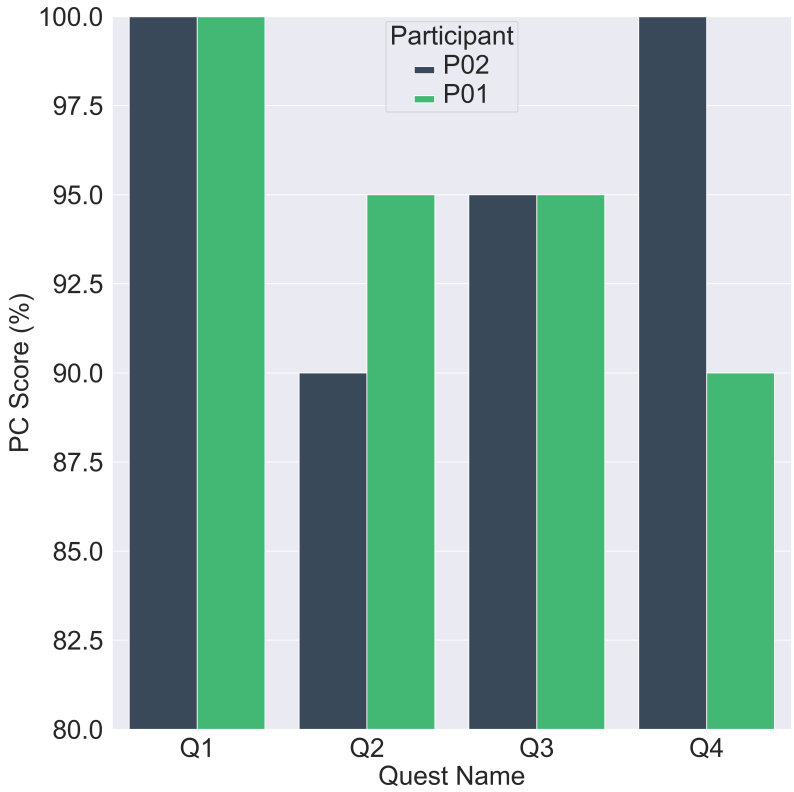


Fig. 4.10.: Performance of participants in the first version of Haptos (chance level = 33.33%)

the cumulative training times on the equivalent number of words correctly recognized proved to be somewhat significant for both P01 ($p = 0.042$) and P02 ($p = 0.00739$). It's worth noting that both participants were acquiring words through TAPS at rates higher than 1 English word per minute, the rate that was reported earlier on the best participants in [2]. Nevertheless, the results of Fig. 4.11 are derived from tests with a chance level of 33.3% since tests presented 3 alternatives per trial. In [2], word identification tests presented the full set of words as alternatives on every trial.

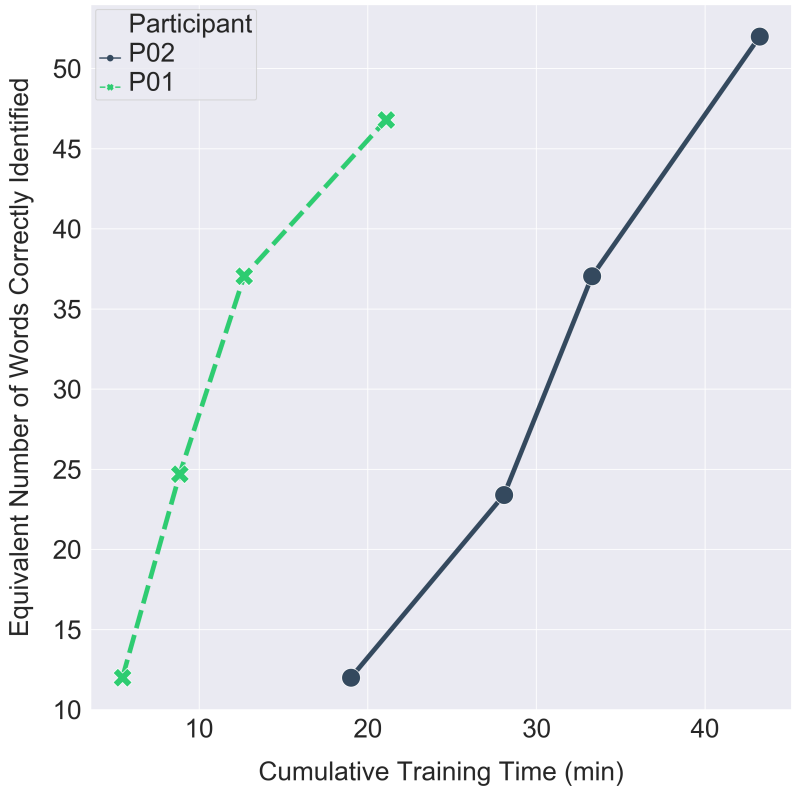


Fig. 4.11.: Equivalent number of words correctly identified in the tests of the first version of Haptos. Tests had a chance level of 33.3%

4.6 Conclusion

According to the performance of the participants in the first exploratory study, it was observed that the designed framework is able to support the learning process of the haptic language through the use of a role playing game. Participants demonstrated that the training quest and complementary quest were effective means of training, as they were able to achieve high performance levels on word identification experiments with testing words (different from training words) and mastered a set of 52 words and 20 phonemes within one hour of game play. These limited data points to the potential of a role-playing game such as the Haptos game at being more effective at supporting language acquisition at a faster pace.

The equivalent number of words correctly identified from both participants resemble the results obtained from Fig.7 of [2], where the top performers had achieved a level of close to 40 words in 40 minutes. The results from Fig. 4.11 show that P01 and P02 had already passed this 40-word threshold after 40 minutes of accumulated training. Furthermore, P01 proved to be a potential top performer as he was able to correctly identify close to 45 words with near to 20 minutes of training. It is important to note that this is not a strict comparison, since tests in [2] used the full set of words as alternatives but the tests in game quests only used 3 alternatives.

The self-paced training paradigm also helped participants learn at their own pace and allowed them to make the best out of their preferred strategies. P01 was able to master all the quests focusing on the perception of individual phonemes learned at the school and further reinforced in the complementary quest. P02 focused on extracting phonemes out of words from the word play mode at the school and learned the association between phoneme sequences and English words using this Freeplay mode.

Even though the first version of the game excluded difficulty parameters and didn't incorporate redundant information with visual and/or auditory cues in word identification tests, the game proved to be a successful proof of concept on how Haptos can be used as a second language learning tool to learn the proposed haptic language. The designed framework and its first implementation proved to be potentially beneficial in the task of teaching the new language in an entertaining and effective way.

5. IMPROVED HAPTIC CODE DESIGN FOR FASTER TRANSMISSION RATES WITH TAPS

5.1 Introduction

Previous work on the phonemic-based approach for tactile speech communication has led to several insights. High percent-correct scores on phoneme identification experiments were achieved with the 24-tactor TAPS array worn on the forearm and the encoding of the 39 English phonemes guided by rules related to the manner of articulation of phonemes [3]. The use of distributed training paradigms that rely on memory consolidation, and strategies to train phonemes before words showed that participants are able to reach word identification rates of up to 40 words per minute [1, 2]. Furthermore, a pilot study on the use of a custom-designed role-playing game that supports language acquisition showed that the game-based design is effective at teaching the haptic symbols in a more engaging way.

Based on all these results, the TAPS device and the coding strategy proved to be a promising communication mechanism in real life situations that could eventually approach normal conversation rates. Towards this ultimate goal, one of the main improvements needed in the system is the enhancement of word recognition rates. Considering the communication rates of 60 to 80 words per minute of the Tadoma method as a benchmark [11, 12], the transmission rates achievable with the current encoding of the 39 English phonemes may be further improved.

For example, one possibility is to incorporate the statistics of conversational English into the design and mapping between phonemes and tactile symbols. This information can benefit the presentation of simpler and faster codes for the most frequent phonemes or combinations of phonemes in regular conversations. Hence, the

average presentation rate of messages could be increased and the word recognition rates could potentially be improved.

This chapter presents work that led to the design and evaluation of an improved coding of the 39 English phonemes based on the statistics of spoken English. A review of two important statistical studies on conversational English is presented, as well as the decision rules derived from such statistics to modify the haptic symbols. The design of the improved codes is described, as well as the design of an extra set of codes that abbreviate the most frequently co-occurring pairs of phonemes. Furthermore, a second version of the Haptos game is presented, which was used to train and test participants on the new haptic symbols. The results of the experiments conducted with eight participants are presented. A summary of the work with data from six participants is available in the following draft paper:

- **J. S. Martinez**, H. Z. Tan, and C. M. Reed, "Improving haptic codes for increased speech communication rates on a phonemic-based tactile display," Draft paper. [50]

5.2 Statistical Studies of Spoken English

Two studies on the statistics of spoken English were reviewed. One study is presented in [51] and reports data on the frequency of occurrence of phonemes, and pairs of co-occurring phonemes in spoken English. A second study presented in [52] reports data on the frequency of occurrence of phonemes in a larger data set of conversational material.

First, the work presented in [51] studied the frequency of occurrence of English phonemes in a set of 72,210 phonemes extracted from conversational material by "phonetic readers". These readers provided a phoneme transcription of the conversational material collected in the study. Furthermore, the study reported the distribution of "digram" (one phoneme following another) frequencies in a set of 49,159 pairs of phonemes.

Table 5.1.: The top ten most frequent phonemes out of 72,210 phonemes (derived from Table II of [51])

Phoneme Label	**Percentage of Occurrence (%)**
UH	9.04
T	8.40
IH	8.25
N	7.08
S	5.09
D	4.18
L	3.69
M	3.29
DH	2.99
K	2.89

From the first analysis of the study, the ten most frequent phonemes were derived. Table. 5.1 shows the selected phonemes using the notation from Table. A.1 and Table. A.2. Although the ranking of phonemes in the table does not show gaps larger than 2% (≈1444 phoneme occurrences), the change in the percentage of occurrence below phoneme K is no larger than 1% (≈722 phonemes). Therefore, the top ten most frequent phonemes are considered to be the most critical phonemes to be encoded with the shortest durations.

As a second analysis performed on the same study, the frequencies of pairs of phonemes (named digrams in [51]) were analyzed to derive the most frequent pairs of phonemes. The data reported in the study were first visualized as a heat map to characterize the distribution among all possible pairs of phonemes in the data set. Fig. 5.1 shows a normalized heat map of the pairs using the label convention defined in Table I of [51] and divided into four quadrants. In the map, the color of a cell in position (i, j) reflects the relative frequency of occurrence of the pair (i, j) (where

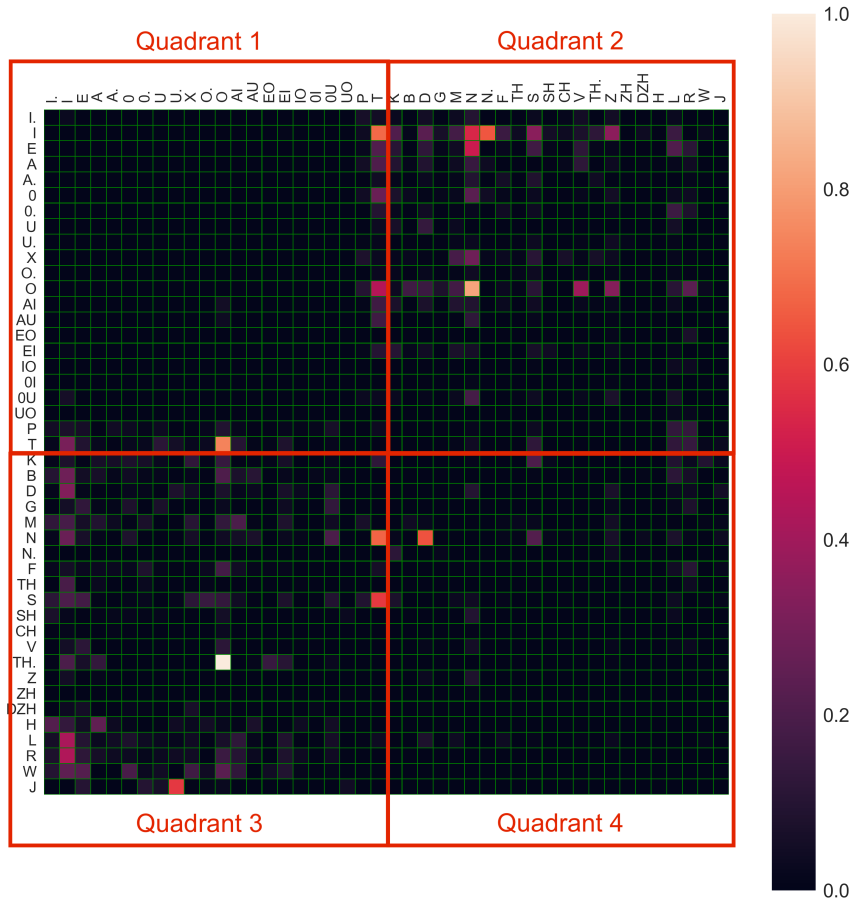


Fig. 5.1.: Normalized heat map of frequencies of occurrence of pairs of phonemes (derived from Table IV of [51])

phoneme i is followed by phoneme j) relative to the maximum frequency in the data set. This maximum frequency was found to be 1192 occurrences and corresponds to the pair DH-UH, which translates to the word "the", the most common word in English. The map shows how the combinations of phonemes create an approximately clustered matrix with occurrences in the second and third quadrants and almost no occurrences in the first and fourth quadrants. Although this is a consequence of the arbitrary ordering of the phonemes, it shows how occurrences are not evenly distributed across all possible combinations. Nevertheless, the occurrences appeared to be evenly scattered among the non-zero (darkest areas) values of the map except

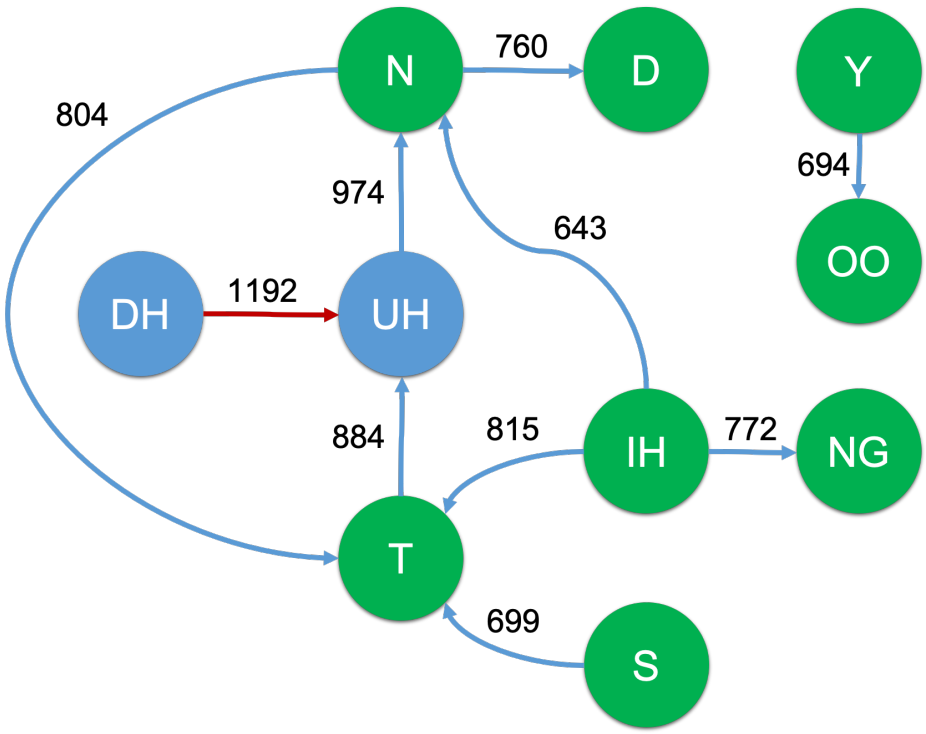


Fig. 5.2.: Top ten most frequent pairs of co-occurring phonemes in [51]

for approximately ten pairs of phonemes detected as the most frequently occurring pairs.

From this observation, Fig. 5.2 shows a directed graph that visualizes the top ten most frequently co-occurring pairs of phonemes using the labeling defined in Table. A.1 and Table. A.2. In the graph, nodes represent individual phonemes and arrows between nodes represent how phonemes are paired. The tail of the arrow corresponds to the proceeding phoneme, and the head of the arrow points to the succeeding phoneme. The number of occurrences of every pair is displayed to the side of every arrow. As depicted in the graph, the most common co-occurring pair is (as expected) the combination DH-UH or the word "**the**" with 1192 occurrences (highlighted with a red arrow). The rest of the pairs in the graph are: UH-N as in "g**un**", T-UH as in "**tou**ch", N-T as in "a**nt**", N-D as in "wa**nd**", IH-N as in "coff**in**", S-T as in "gho**st**", IH-T as in "rabb**it**", IH-NG as in "k**ing**", and Y-OO as in "**you**".

Table 5.2.: The top ten most frequent phonemes out of 103,887 phonemes (derived from Table 3 of [52])

Phoneme Label	**Percentage of Occurrence (%)**
UH	7.30
N	6.72
T	5.78
IH	5.15
S	4.61
R	3.87
EE	3.69
L	3.64
D	3.33
EH	3.21

The phonemes identified in Table. 5.1 and Fig. 5.2 served as the first source of information on the most frequently occurring phonemes and phoneme pairs in spoken English. As a second source of information, the studies from [52] were also analyzed. In this work, researchers performed a statistical study of the frequency of occurrence of phonemes in conversational English. The materials studied were collected from a database of 26 recorded interviews manually transcribed into phonemic representations. The resulting data set consisted of a total of 103,887 phonemes. As in the analysis performed on the data reported by [51], the top ten most frequent phonemes were obtained from the results of [52]. Table. 5.2 shows the identified phonemes using the notation from Table. A.1 and Table. A.2. Unlike the case with the data of [51], gaps between phonemes in the ranking of [52] are not bigger than 1% ($\approx$1039 phonemes). Nevertheless, due to the increased size of the data set, this percentage corresponds to a larger number of phonemes.

Considering both Table. 5.2 and Table. 5.1, it is interesting to note that the two statistical studies agree on seven out of the ten phonemes when combining the two lists of the most frequent phonemes. The resulting thirteen phonemes after forming the union of the two tables are: **UH**, **N**, **T**, **IH**, **S**, **R**, **EE**, **L**, **D**, **EH**, **M**, **DH**, and **K**. This list of phonemes also include all of the phonemes in the pairs of Fig. 5.2 except for Y, OO and NG.

The resulting list of thirteen phonemes and the ten most frequent pairs of co-occurring phonemes represent the general outcomes of the analyzed statistical studies. These results were used as the basis of all the decision rules that guided the design of an improved set of haptic codes.

5.3 Design of Improved Haptic Codes

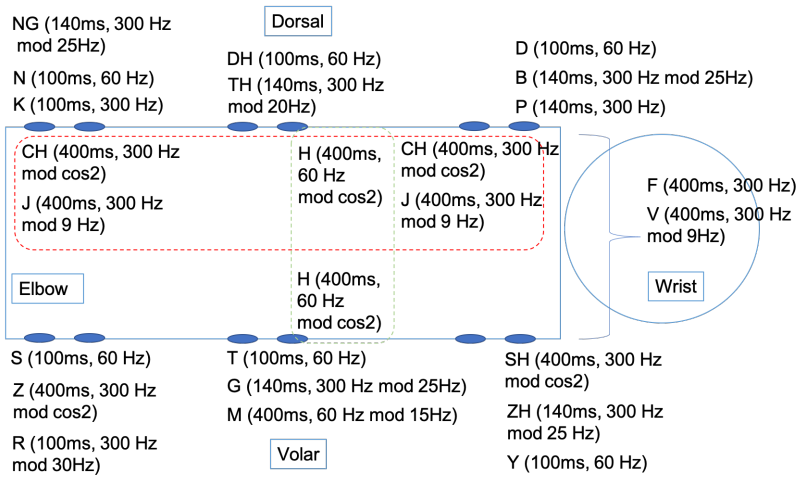
The design of the 39 improved haptic codes was guided by decision rules based on the results of the statistical studies of phonemes in spoken English. The two major rules defined in the design process were:

1) The most frequent phonemes should be designed as short and simple signals.

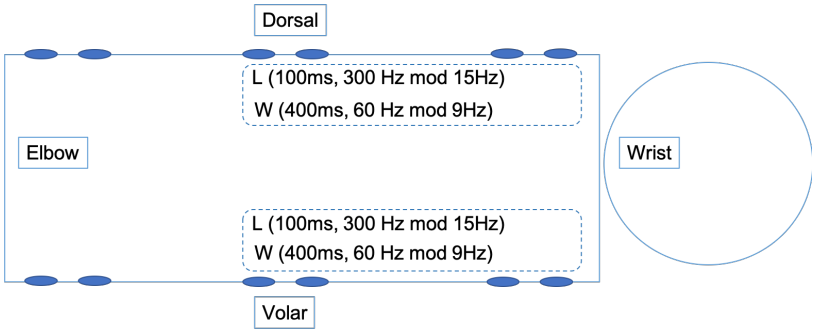
2) The most frequent pairs of phonemes should be coded in adjacent spatial locations as an abbreviated "chunk."

Additionally, due to the high distinctiveness of the previous codes used in [1–3], the new codes also tried to preserve the modulation and movement characteristics of the 39 symbols designed for these studies.

The design of the new haptic symbols followed a sequence of steps that favored the design of consonants before vowels. Consonants were designed and located based on rules 1) and 2) preserving some characteristics of the previous codes used in [1–3]. A visual representation of the location of the new consonant codes and some of their parameters is summarized in Fig. 5.3, and a detailed description of the signal parameters can be found in Table. D.1.



(a) Set of 22 new consonants



(b) Set of 2 new consonants

Fig. 5.3.: Visual representation of the improved set of consonants

In Fig. 5.3, the phonemes are located relative to their designated locations in one or more of the six areas of vibration in the 24-tactor array (three on the dorsal side and three on the volar side). As depicted in the diagrams, no more than three phonemes are coded at the same location. Phonemes are labeled according to their corresponding name and their duration, frequency, and possible modulation parameters in parenthesis. For example, NG (located at the dorsal side near the elbow) is a 140-ms long, 300-Hz signal modulated by a 25-Hz signal. All modulated signals were amplitude modulations with a modulation index of 1 (the ratio between the amplitudes of the signal used to modulate and the modulated signal is 1), or with

a cosine-squared window (shown in the diagram as the "cos2" modulation). Some phonemes are coded with multiple tactors simultaneously. For example, F and V use tactors near the wrist on both the dorsal and volar sides, and L and W use tactors from the middle of the arm and near the wrist on both the dorsal and volar sides. The steps followed to design the consonant signals that resulted in the codes visualized in Fig. 5.3 can be summarized as follows:

i) The seven consonants from the graph of the most frequent co-occurring pairs of phonemes (Fig. 5.2) were coded first. These were: DH, N, D, T, S, Y, and NG. The phonemes were coded at locations distributed among the following six distinct locations on the forearm: the volar and dorsal areas closest to the elbow, the volar and dorsal areas closest to the wrist, and the volar and dorsal areas at the middle of the arm.

ii) The distribution of these phonemes was defined so that it favored the adjacency of the phonemes in the pairs. For example, S and T were coded in adjacent locations as the pair S-T is a common co-occurring pair. This distribution favored rule 2) as the stimuli of the two phonemes would be felt close together so that the overall sensation resembles a "chunk".

iii) A short, 100-ms pulse at 60 Hz was assigned to these seven initially coded phonemes. The short duration of this pulse characterized the critical consonants identified as the most frequent ones, or the consonants that take part in the graph of Fig. 5.2.

iv) The remaining four consonants of the thirteen most frequent phonemes that were not part of the most frequent co-occurring pairs were coded next. These were: L, R, M, and K. These phonemes were designed again as 100-ms pulses at 60 Hz and were distributed among the six areas of the arm so that they kept the same locations and modulation characteristics as in the previous encoding. Every location was associated with a maximum of three phonemes. If a loca-

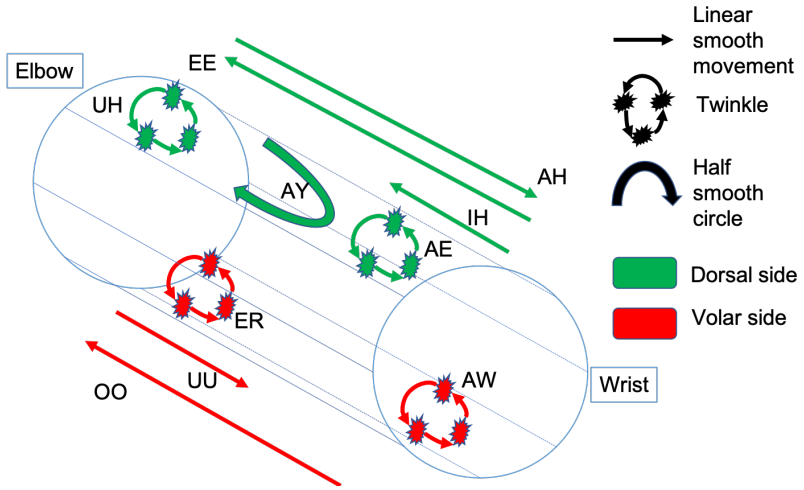
tion reached the maximum, the next phoneme was coded at another available location.

v) For phonemes that shared the same location, the phonemes added last were coded with characteristics to differentiate them from previously coded consonants. The changes included a change of frequency to 300 Hz or the addition of amplitude modulation. The duration of 100 ms was kept constant except for the special case of NG and M which required a duration of 140 ms and 400 ms, respectively.

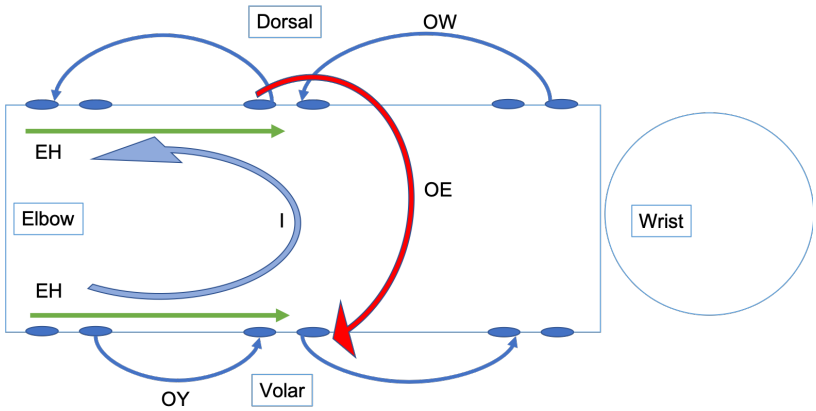
vi) The remaining consonants of the 39 phonemes were distributed along the arm so that they kept the same locations as in the previous codes. The duration, frequency and modulation characteristics of these consonants remained consistent and were minimally changed in specific cases to differentiate them from other phonemes already present at the same location. Modifications included frequency and modulation changes.

For the case of vowels, the general design decisions preserved more characteristics of the previous codes than in the case of the consonants. Movements and illusions were minimally changed to favor rule 2). A visual representation of the vowels is shown in Fig. 5.4 that represented the movement directions and the associated sensations. A detailed description of the sensations and the signal parameters can be found in Table. D.2.

In Fig. 5.4(a), the "twinkle"-like sensation is represented as a circular motion that produces localized "tickles" during the movement. On the other hand, smooth sensations produced by apparent motion illusions are represented as straight arrows. These movements vary in length depending on the vowel. For example, AH is a long movement that runs from the elbow to the wrist on the dorsal side, but IH is a shorter movement that starts at the wrist and finishes at the middle of the arm. Other types of movements are shown in Fig. 5.4(b), and can be described as: a circular ring motion around the middle point of the arm for OE, a "grabbing sensation" from the volar



(a) Set of 10 new vowels



(b) Set of 5 new vowels

Fig. 5.4.: Visual representation of the improved set of vowels

and dorsal sides close to the elbow to the volar and dorsal sides at the middle point for EH, a semi-circular motion from the volar to the dorsal side of the elbow passing through the middle point for I, and hoping sensations along the longitude of the arm produced by cutaneous rabbit effects for OW (from wrist to elbow) and OY (from elbow to wrist). The steps followed to design the vowel signals visualized in Fig. 5.4 can be summarized as follows:

i) All vowels were kept as consistent as possible with the previous codes. The haptic illusions of every vowel (including those on the thirteen most frequent phonemes) remained unchanged except for the phonemes that belonged to the most frequent co-occurring pairs.

ii) When a vowel was involved in the most frequent phoneme pairs, the movement sensations were reallocated so that the start (or end) of the vowel connected with the end (or start) of the other phoneme in a pair. One example is the case of DH-UH. In the previous codes, DH was located at the dorsal middle area of the arm and UH was a "grabbing" sensation moving from near the wrist towards the middle of the arm. In the new codes, DH was designed at the dorsal middle side and UH was modified to be a "twinkle"-like moving sensation on the dorsal side that started on the same tactors used by DH (at the middle section), and finished off at the tactors closest to the elbow.

In addition to the improved 39 codes, an extra set of ten symbols was designed, one for each of the pairs in the list of most frequent co-occurring pairs of phonemes. These symbols were denoted as "chunks" and corresponded to abbreviated versions (in duration) of the two phonemes in the pair that preserved the individual "feel" of each phoneme. Using these symbols, a pair of phonemes presented in a word could be replaced by the single abbreviated signal, which can improve word presentation rates even further. A diagram that shows a visual representation of the signals used for each pair is provided in Fig. 5.5 and a detailed description of the signals can be found in Table. D.3. Every chunk signal corresponded to a 100 ms or 200 ms stimulus, where 50% of the duration was designated to a pulse that represented the first phoneme of the pair (blue pulse in the pair in Fig. 5.5), and the other 50% was designated to a pulse that represented the second phoneme (yellow pulse in the pair in Fig. 5.5). A 60 Hz pulse was used to represent a consonant, and a 300 Hz pulse modulated by a cosine-squared window was used for a vowel. A special signal of a 300-Hz carrier

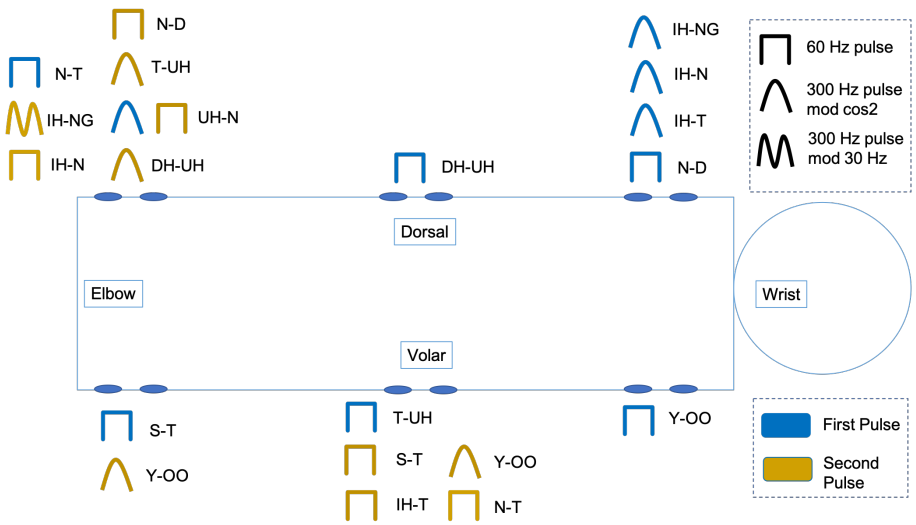


Fig. 5.5.: Visual representation of abbreviated pairs of phonemes

modulated by a 30-Hz signal was used to differentiate the second portion of IH-NG from the second portion of IH-N.

As depicted in Fig. 5.5, the pulses in every pair preserve the location of the original phonemes being abbreviated. For example, a 60 Hz pulse at the dorsal side of the arm closest to the elbow was used to represent the N portion of the chunk N-T. This location is the same as the original code for N. Similarly, the T portion of the signal was located at the middle volar side of the arm, as the T phoneme is coded at this location. It is worth noting that the pulses used to represent vowels are no longer dynamic and don't resemble the movements from the original codes. Nevertheless, the modulated 300 Hz pulses were chosen as the signals that capture the main "essence" of the sensations, given the fact that all signals of the original vowels have a base frequency of 300 Hz (see Table. D.2).

With all the newly designed symbols, the learnability of the signals was tested under the same game environment used in previous pilot studies that showed how the "Haptos" game could be used as an effective training mechanism for learning the haptic language.

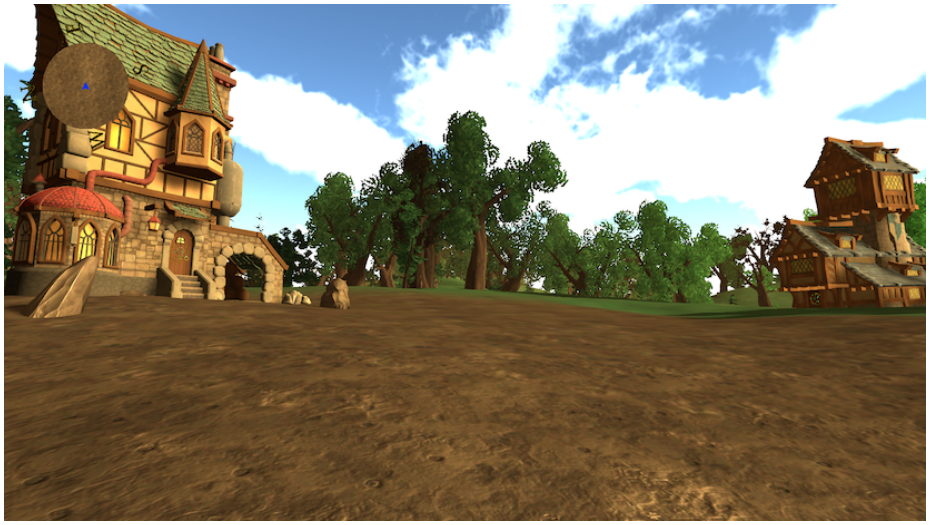


Fig. 5.6.: The village in the second version of Haptos

5.4 Evaluation of Improved Haptic Codes

5.4.1 A new version of the Haptos game

A new version of the Haptos game was designed and implemented to apply a game-based approach to training and testing participants with the improved haptic codes. The objectives were to evaluate the distinctiveness of the new codes and the learning curve with the new haptic symbols. Because of this, the new game design corresponds to a smaller version of the Haptos game, the training school is only focused on phoneme learning, and a learning mechanism based on progressive quests is not needed. In addition to the training school, the new version of the game included a testing area denoted as "The Lab" where cumulative phoneme-identification experiments were conducted with a forced-choice, no-feedback paradigm. Fig. 5.6 shows a screen capture of the new village, which only includes the lab (house to the left) and the training school (house to the right).

Even though the new version of the game only included the necessary assets to train and test phonemes, the game has a generalized structure that can be expanded to include quests and allow for the learning and testing of words.

Table 5.3.: Haptic symbols by training zones

Training Zone	Haptic Symbols Included
Z1	T, S, DH, N, NG, D
Z2	P, K, B, M, G, TH, Z
Z3	F, V, CH, J, SH, ZH
Z4	H, L, R, W, Y
Z5	UH, IH, OO, UU, EE, AH
Z6	AW, ER, AY, AE
Z7	I, OE, OW, OY, EH
Z8	DH-UH, UH-N, T-UH, IH-T, N-T IH-NG, N-D, S-T, Y-OO, IH-N

5.4.2 Learning Materials

The training school in the new version of the game divided the 49 new symbols (39 English phonemes and 10 abbreviated chunks) into eight training zones as shown in Table. 5.3. The first four zones (Z1 to Z4) included all 24 consonants, zones Z5 to Z7 included all 15 vowels, and zone Z8 included all the signals designed for the most frequently co-occurring pairs of phonemes.

In the school, all zones were distributed throughout a closed room that participants could freely explore by approaching a specific area. Fig. 5.7 shows screen captures of the school and one zone. To enter a zone, a participant would approach a lighted area under the zone number and press the F key on the keyboard. As show in Fig. 5.7(b), a participant was able to click on individual phonemes (represented as spheres with the phoneme labels at the center) to bring them into focus and interact with them. When clicked, the phoneme was brought to the center of the screen along with an object that represented the usage of the phoneme in a word. The example in Fig. 5.7(b) shows how a participant clicked on the phoneme SH after entering zone

(a) Partial view of the training school in the second version of Haptos. A user can interact with any zone by approaching the lighted area under the zone number. All zones were available at all time

(b) Zone 3 of the training school in the second version of Haptos. The zone includes phonemes J, F, ZH, V, SH and CH. The participant has clicked on the sphere "SH" to bring the phoneme into focus

Fig. 5.7.: Training school in the second version of Haptos

Z3 in the school. The phoneme is presented with a sea shell as an example of the usage of the phoneme in the word **sh**ell. Users could feel the haptic symbol associated with the phoneme by pressing "Play Phoneme" or return it to its original position by pressing "Return the phoneme". Moreover, a practice test with feedback was available

within every zone by pressing the "Test" button on the screen. A randomly selected phoneme was presented and the participant was instructed to click on the perceived phoneme. Correct-answer feedback was provided after clicking the perceived stimulus. Additionally, participants were able to quickly navigate between zones by specifying a zone number in the text input at the top right corner of Fig. 5.7(b) and pressing "Go".

5.4.3 Procedure

Eight participants (4 male and 4 female) recruited from Purdue University participated in an experiment to evaluate the newly designed codes. All participants provided informed-consent through an approved IRB protocol and were paid for their participation. The participants (P001 through P008) ranged in age from 19 to 26 years with a mean of 22.13 and standard deviation of 2.23 years. Four participants were native English speakers, two were native Spanish speakers, and two were native Chinese speakers. Non-native English speakers reported that they acquired English as a second language at 6, 7 or 9 years of age.

The experiment was designed to last for 11 days with daily sessions of approximately one hour. During the experiment, participants learned increasing sets of the learning material based on the learning levels presented in Table. 5.4. The levels distinguish sets of consonants (L1 through L4), vowels (L5 through L7), the chunks group (L8) and the full set of 49 symbols (L10). In the table, an additional level L9 is omitted. This level included zones Z1 to Z7 and was tested in the experiments described in [50], which describes procedures with four participants (P001 through P004) and served as a precedent to the experiments described in this chapter. In [50], P001 and P003 were trained and tested on all levels up to L9 and excluding L8, and P002 and P004 were trained and tested on all levels up to L10, excluding L9. To include P001 and P003 in the experiments described in this chapter, the participants' training was extended to include L8 and L10.

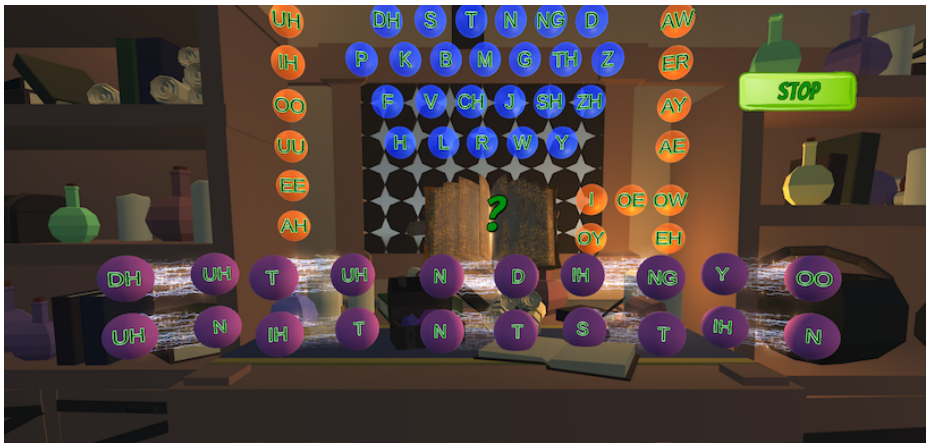
Table 5.4.: Learning levels for the 49-symbol experiment

Learning Level	**Zones Included**	**Number of Haptic Symbols Included**
L1	Z1	6
L2	Z1 to Z2	13
L3	Z1 to Z3	19
L4	Z1 to Z4	24
L5	Z5	6
L6	Z5 to Z6	10
L7	Z5 to Z7	15
L8	Z8	10
L10	Z1 to Z8	49

In the 11-day curriculum, the first eight days of the experiment were devoted to learn and test levels L1 through L8. On the remaining three days, participants were tested on level L10 three times (one test for L10 per day). Every day, a participant was informed about which level was to be trained and tested, and the zones of phonemes included in the test. The participant was then left to train for the corresponding zones in the training school with no time restrictions. Participants could also use their time freely, which allowed them to spent their time learning or reviewing other zones not necessarily included in the test for a specific day.

Whenever ready, participants went to the "Lab" to be tested on the corresponding level of the day using a forced-choice, symbol-identification test. The lab was a special testing area in the game that included an interface to perform the test and go through a simple tutorial. Fig. 5.8 shows screen captures of the interface (Fig. 5.8(a)) and a trial of the test for level L10 (Fig. 5.8(b)). The tutorial contained an explanation of the procedure and offered a mock trial of the test. Participants were advised on the first day of the experiment to go through the tutorial before attempting the test for

(a) Interface to operate the tutorial and to initiate the test



(b) Trial of a test for level L10. A participant feels a random haptic stimulus and is asked to click on the percieved stimulus

Fig. 5.8.: The lab: Testing zone in the second version of Haptos

level L1. To perform a test, participants indicated the level number to be tested in the text input of Fig. 5.8(a) and then clicked on the button "Make a Crystal". The tutorial explained the meaning of a "crystal" in the context of the videogame and how solving a test involved creating a crystal. On every trial of a symbol-identification test, one of the haptic symbols from the set of zones contained in the level tested was selected at random (with replacement) and delivered to TAPS. As shown in Fig. 5.8(b), participants were asked to click on the perceived haptic symbol from the set of possible answers (49 symbols in the example shown in the figure).

In every test, randomization was performed with replacement and the total number of trials was level-dependent. It was set by the software so that the probability of each symbol appearing at least once during the test was at least 70%. The number of trials ranged from 20 to 30 from L1 to L4, and remained 20 for L5 to L8. The participants had to pass the test of a certain level with a score of $\geq 80\%$ before they were allowed to progress to a higher level (see Table. 5.4). The test was repeated the following day if the participant failed. For L10, 60 trials were performed for each of the last three days of the experiment. On the last three days, there was no passing score and participants completed exactly 3 tests on level L10.

5.4.4 Data Analysis

The training behaviour of participants was examined through the average time spent on the different zones of the training school, as well as the distribution of data among the zones. The average response time of participants on every test as well as its distribution for all learning levels was also analyzed. This was recorded as the time elapsed between the offset of a haptic symbol played on TAPS, and the onset of a participant's response. Additionally, performance metrics of percent-correct scores for level tests that met the criterion of $> 80\%$ and the equivalent number of symbols correctly identified were logged. The latter was a metric calculated from multiplying the score of a test and the number of haptic symbols in the test, all for levels L1

through L10 keeping the distinction between consonants, vowels, chunks, and the full set of symbols.

A separate analysis was made for the tests conducted on the last three days of the experiment for level L10. In this case, all the three blocks of 60 trials from all participants were pooled into a data set used to build a 49×49 confusion matrix. The matrix contained a total of 1,440 trials (8 participants × 3 blocks × 60 trials per block). This matrix was used to determine: an overall percent-correct score, an information transfer estimate, and a visualization of confusions between symbols.

The overall percent correct score was determined from:

$$pc = \frac{\sum_{i=1}^{49} m_{ii}}{n} \tag{5.1}$$

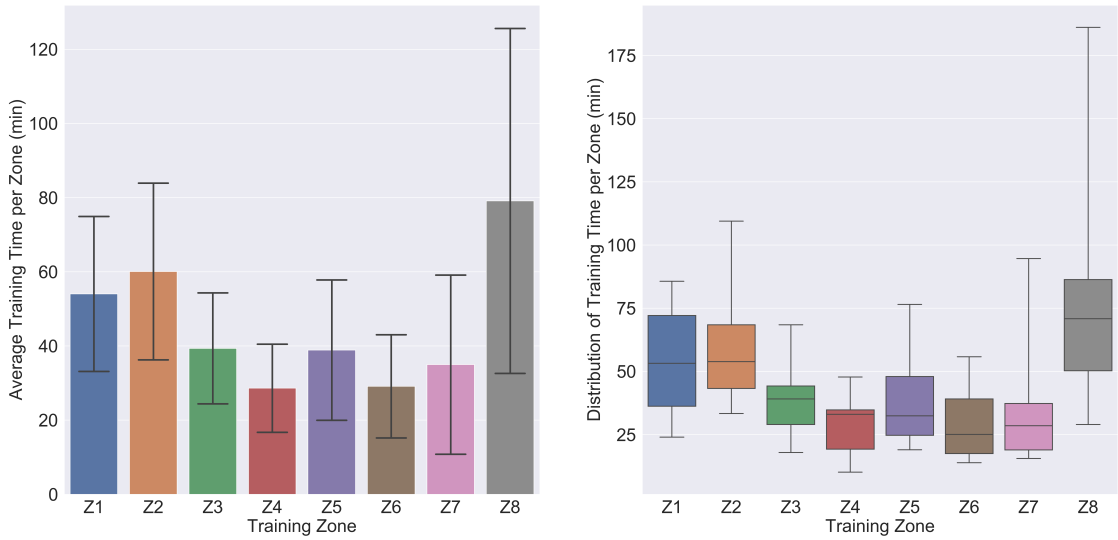
where n is the number of trials in the matrix ($n = 1,440$) and m_{ij} is the number of times phoneme i was confused as phoneme j. Hence, m_{ii} are the diagonal elements of the matrix which correspond to correct responses.

The information transfer estimate was calculated based on the lower bound discussed in [47]. This bound is a conservative estimate that avoids over estimating the information transfer due to the limited number of trials, i.e, the condition $n > 5k^2 = 5 \cdot (49)^2 = 12005$ is not met. With a low error rate, the estimate was calculated as:

$$IT_{pc} = (1 - 2e) \cdot \log_2 k \quad ; \quad e = 1 - pc. \tag{5.2}$$

Finally, a visualization tool designed to view confusion patterns was used to represent the confusion matrix as a directed graph. The tool permits the viewing of confusions within a range of errors per haptic symbol. For a symbol i, this error is calculated as:

$$e_i = 100\% \times \frac{\sum_{j=1, i \neq j}^{49} m_{ij}}{\sum_{j=1}^{49} m_{ij}} \tag{5.3}$$



(a) Average training time per training zone. Error bars display a ±1 standard deviation

(b) Distribution of training time per training zone. Every box plot shows data within the 25th and 75th percentile with a horizontal bar representing the median. Whiskers in the plots extend to show maximum and minimum data points

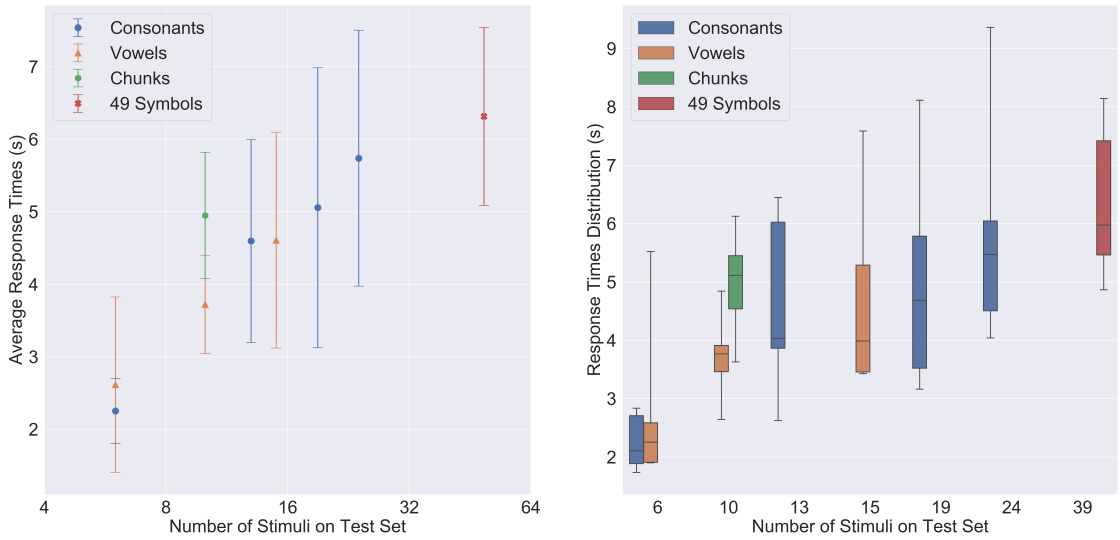
Fig. 5.9.: Training behaviour of participants in experiments with the new haptic symbols

5.4.5 Results

The training behaviour of participants is summarized in Fig. 5.9. The figure shows a plot with the average time spent by participants in each of the zones (see Fig. 5.9(a)) and the distribution of the training time data per zone as a box plot (see Fig. 5.9(b)). From the plots, it can be observed that participants spent a significant amount of time in zone Z8 with a high degree of variability among participants. This can be concluded from the size of the error bars in Fig. 5.9(a) and the extension of the whiskers in Fig. 5.9(b) for this zone. This conclusion can be explained by the fact that zone Z8 was tested and trained independently from the other symbols, which may

have encouraged participants to spend the full day session training for this specific zone. Moreover, participants also reported that the zone was one of the hardest to learn, meaning that more time was required to learn to distinguish the chunk symbols from each other. In terms of the other zones, it can be observed that participants also spent a great amount of time training for zones Z1 and Z2. This can be explained by the fact that the zones were the first two zones of the experiment, which meant they were included in all the tests up to level L4, and on the last three days where L10 was tested. This encouraged participants to review these zones extensively, which was an expected behaviour that the experimental design meant to achieve. Since the two zones contained consonants that could be easily confused due to their similarity in location (see Table. D.1 for details on spatial parameters), participants would benefit from constantly reviewing the zones every day, at least until the test for L4 was passed. In terms of the data distribution, only data for zone Z4 appeared to be significantly skewed towards the 75th percentile (top quartile) of its distribution.

The average response time was also calculated for tests passed at every level from L1 to L8, and for the pooled results of L10. Fig. 5.10 shows the behaviour of participants during tests in terms of their response time based on the number of haptic symbols included in each test (see Table. 5.4 for the number of symbols in every level). As shown, the response time appears to increase linearly as the number of symbols in the set tested is doubled. This can be observed from the patterns in Fig. 5.10(a) for each type of symbol (consonants, vowels, chunks, and 49 symbols). This reflects an increase in processing time with the increasing complexity of the tests with respect to the number of symbols. As this number increases, participants have to choose a response from a larger set of options every time. The results are also similar to those presented in [3], where tests with consonants only, vowels only, and 39 phonemes yield average response times with the same increasing behaviour. A high degree of variability was observed among participants, as shown by the error bars in Fig. 5.10(a). During the tests, it was observed that some of the participants hesitated during trials where the stimulus was either forgotten after presentation, or possibly



(a) Average response time as a function of the number of symbols in the tests. Error bars display a ±1 standard deviation

(b) Distribution of response time as a function of the number of symbols in the tests. Every box plot shows data within the 25th and 75th percentiles with a horizontal bar representing the median. Whiskers in the plots extend to show maximum and minimum data points

Fig. 5.10.: Response time behaviour of participants in level tests with the new haptic symbols

confused with another symbol. These participants tended to pause for a significant amount of time before choosing a response. This can be observed in the distribution of the data in the box plot of Fig. 5.10(b), where the whiskers for minimum and maximum response time extend far from the interquartile range.

In terms of performance, Fig. 5.11 shows the average percent correct scores across participants from all successful level tests from L1 to L8, i.e., tests where participants obtained a percent correct score $\geq 80\%$. For L10, the data was averaged from pooled results of the three tests with 60 trials. From the plot, it can be observed that on average, the performance was always higher than the 80% criterion. The highest scores were recorded for levels L1, L5 and L6, which suggest that these were the "easiest"

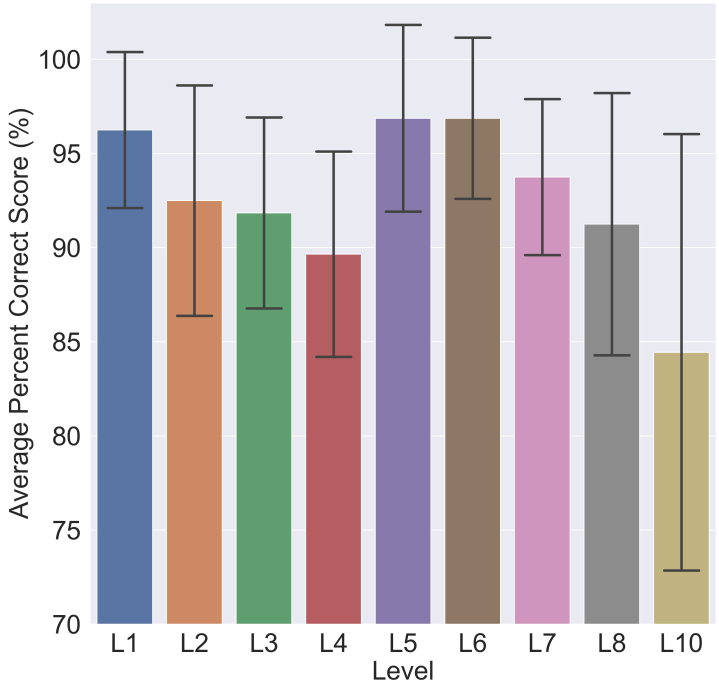
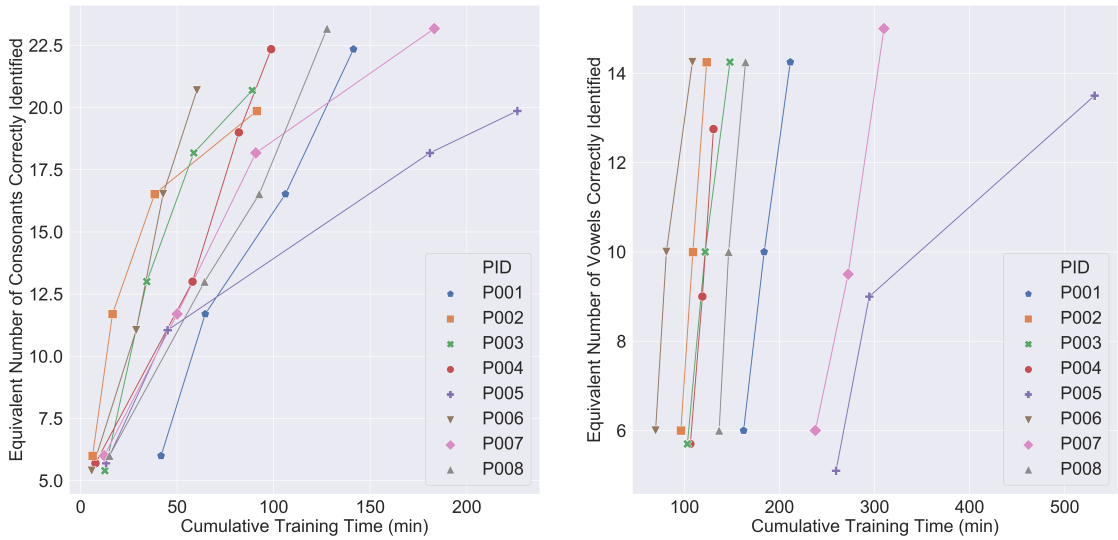


Fig. 5.11.: Performance of participants in successful level tests with the new haptic symbols. Error bars display a ± 1 standard deviation

levels since the number of haptic symbols per level is only 6 for L1 and L5, and 10 for L6. It is interesting to note that although L8 also contains only 10 symbols, the performance of participants in this test was lower than the performance for L6. This suggests that the chunk signals may be more difficult to learn. Their short duration is presumably the main cause of this issue, which results in a lower performance. Nevertheless, the average performance level was still over 90% for this test. The plot also shows that levels L4 and L10 were difficult tests, with L10 showing the worst average performance among all. This was due to the highest number of haptic symbols involved in this test. Nonetheless, the average performance is quite high for L4 (close to 90%), and above the 80% criterion for L10. The lowest performance level on the last test reflects the increased difficulty of the task, where 49 symbols had



(a) Equivalent number of symbols correctly identified for consonants

(b) Equivalent number of symbols correctly identified for vowels

Fig. 5.12.: Equivalent number of symbols correctly identified for groups of phonemes in level tests with the new haptic symbols

to be distinguished, as opposed to 24 constants or 15 vowels in levels L4 and L7, respectively.

To analyze how participants progressed throughout the experiment, Fig. 5.12 shows the equivalent number of symbols correctly identified as a function of the cumulative training time for the groups of consonants (Fig. 5.12(a)) and the groups of vowels (Fig. 5.12(b)). In the figures, the cumulative training time was calculated as the total duration spent in the training school at the time of passing a level test. Since participants tested all the consonant levels before vowels, the cumulative time shown in Fig. 5.12(b) includes the time already shown in Fig. 5.12(a). As shown in the plots, participants showed a similar trend in their learning curves for both consonants and vowels except for participant P005, who had some difficulty with a few tests. During the sessions, this participant reported that the level tests were very difficult compared to the mock tests available in the training school. This was

Table 5.5.: Statistics for the learning pace of participants for consonants and vowels

	Min	**Max**	**Average**
Consonants learned per Minute	0.062 (from P005)	0.288 (from P006)	0.175
Vowels learned per Minute	0.026 (from P005)	0.304 (from P002)	0.165
Time to learn a consonant (min)	3.472	16.129	5.714
Time to learn a vowel (min)	3.289	38.462	6.061

because the training school tests allowed participants to play signals randomly within individual training zones, but not with a cumulative set of symbols from several zones which was the case of the level tests. None of the other participants expressed that this design represented a major difficulty in the tests. The trends from both plots also show how vowels were learned faster than consonants in the sense that the slopes of the lines in Fig. 5.12(a) were less steep than those in Fig. 5.12(b). To further evidence this fact, the statistics of the learning pace of participants for the groups of vowels and consonants are shown in Table. 5.5. These statistics are derived from linear regressions applied to the plots in Fig. 5.12. On average, it is shown that consonants were learned faster, mainly due to the outlying data of participant P005. The time required to learn a new consonant was less than that required to learn a new vowel. Nevertheless, this difference is only on the order of 21 seconds.

For an analysis of the participants' performance with all 49 haptic symbols, the results from the eight participants on the last three days of tests on level L10 were pooled into one confusion matrix. Table. 5.6 shows the percent correct score calculated from the matrix, the lower bound information transfer estimate (IT_{pc}), the number of correctly-recognized symbols without any error ($2^{IT_{pc}}$), and the total training time

Table 5.6.: Results from the pooled tests on level L10 with the new haptic symbols

Percent Correct Score (%)	$\mathbf{IT_{pc}}$**(bits)**	$\lfloor \mathbf{2^{IT_{pc}}} \rfloor$	**Total Time (hr)**
84.44	3.87	14	5.51 ± 1.80

averaged across the eight participants at the last day of the experiment. According to the table, the participants achieved a high percent correct score. These results can be compared with those in [3] where similar phoneme identification tests with 10 participants and 1,560 trials resulted in scores of 85.77% with 39 haptic symbols. As demonstrated in Table. 5.6, the addition of 10 symbols did not represent a significant decrease in performance. Additionally, the number of symbols that participants could correctly identify without any error was 14, and the average training time required to learn the new set of symbols was less than 6 hours. Moreover, the equivalent number of symbols correctly identified was averaged across the eight participants from the pooled results of the last three days of experiments for L10. This averaged metric was 44.24 ± 2.64.

A custom visualization tool was developed to show the confusion patterns from the pooled confusion matrix as a directed graph. The full graph is shown in Fig. 5.13. The tool represents all haptic symbols as colored circles of different sizes, and the confusions between them are represented by blue lines. The yellow, green and blue circles represent consonants, vowels, and chunks, respectively. The size of the circles is proportional to the number of times the haptic symbol was presented in L10 tests. For each pair of connected circles, the red semicircle at one end of the line indicates the erroneous phoneme that was used as a response. For example, the pair OW and EE is a confusion at the top of the graph in Fig. 5.13. In this case, OW was confused as EE since the red semicircle is attached to the circle that encloses the EE label. The tool also displays the minimum and maximum error per haptic symbol found in the graph. For the data collected with the eight participants, the error rate per haptic

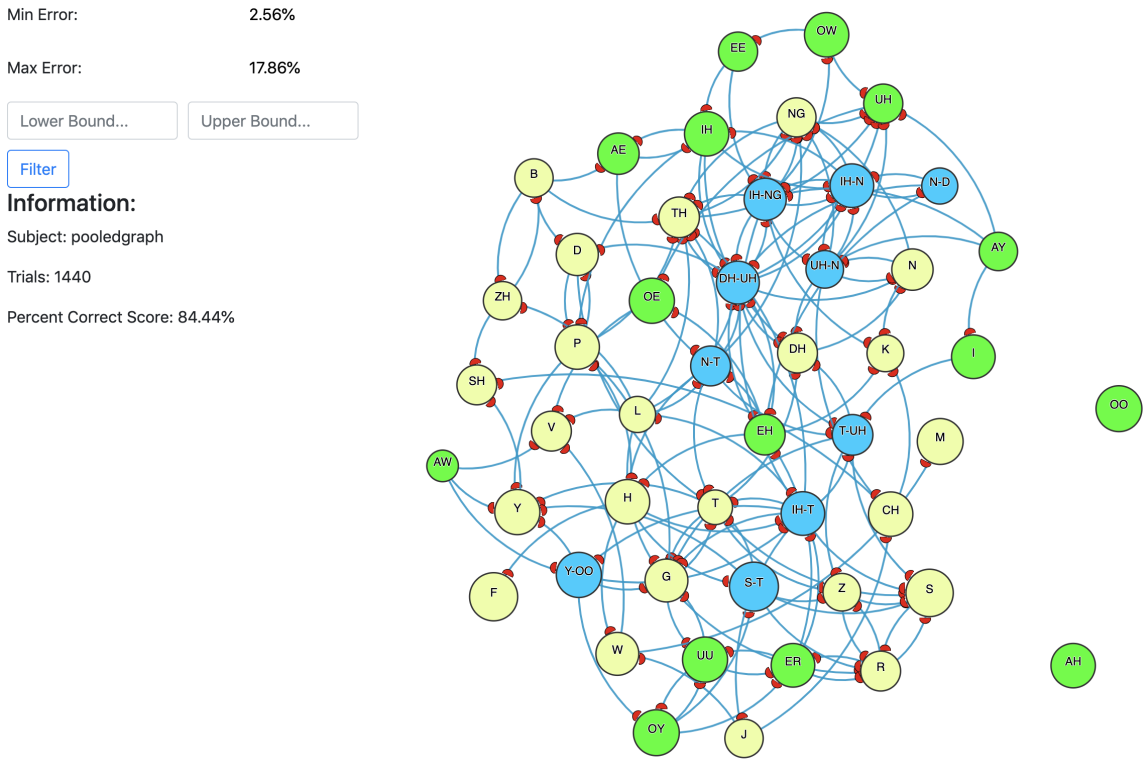


Fig. 5.13.: Visualization tool used to analyze the pooled confusion matrix in the L10 tests with the new haptic symbols

symbol ranged between 2.56% and 17.86%. From the pooled matrix, the minimum and maximum number of trials a symbol was presented was 16 (for AW) and 39 (for S-T), respectively. It can be observed that the vowels AH and OO were never confused with any other phoneme.

The visualization tool allows a user to filter the confusions in the graph by specifying the lower and upper bounds to be visualized in the "Lower Bound" and "Upper Bound" text fields. The confusions outside the given error range were then removed from the directed graph. This feature allows removal of the clutter due to occasional errors in order to focus on the major confusion patterns among the haptic symbols. For example, the confusions between consonants and chunks after filtering the graph with a minimum error of 8% are shown in Fig. 5.14 which is much less cluttered. In this graph, the minimum error was 8% (2 out of 25 trials) between J→W. The

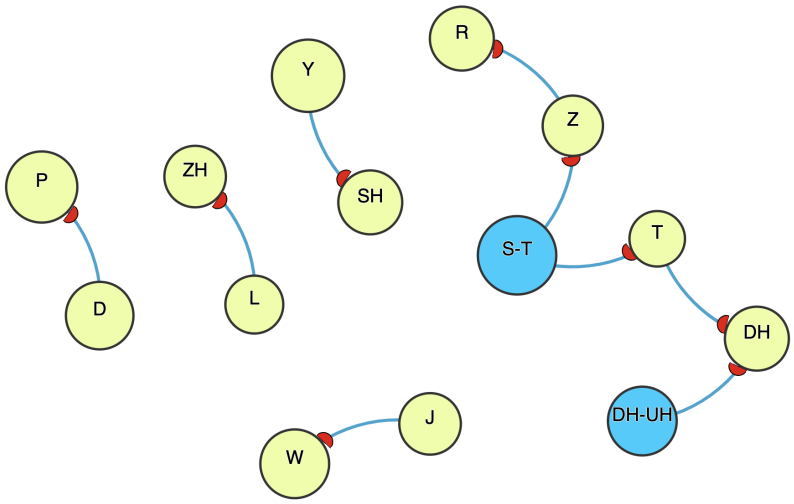


Fig. 5.14.: Confusions between consonants and between consonants and chunks in the pooled results of the L10 tests with the new haptic codes

maximum error was 14.71% (5 out of 34 trials) for the pair Y→SH. For confusions between consonants, the majority of the errors corresponded to pairs of phonemes that share the same location in the TAPS' tactor array. Other confusions involve attention errors, where participants may have missed part of the stimulus or only paid attention to a subset of vibrations from the symbol presented. This is the case for L→ZH and J→W. Other special confusion was the pair T→DH with an error of 10% (2 out of 10 trials). Even though T and DH are coded as the same 60-Hz pulse of 100 ms in duration, they are located at opposite sides of the arm and this confusion is rather rare. An interesting pattern of confusions was evidenced between some chunks and consonants. The confusions show how participants confused the chunks with a phoneme that feels like either the first pulse or the second pulse of the chunk. For example, the confusion DH-UH→DH shows that the participant may have only perceived the first pulse (DH portion) of the chunk in the trials when the confusion occurred. Similarly, the confusion S-T→T shows that the participant missed the first pulse (S portion) and only noticed the second pulse (T portion) of the chunk. Confusions of this type combined with location errors were also observed. For example, the

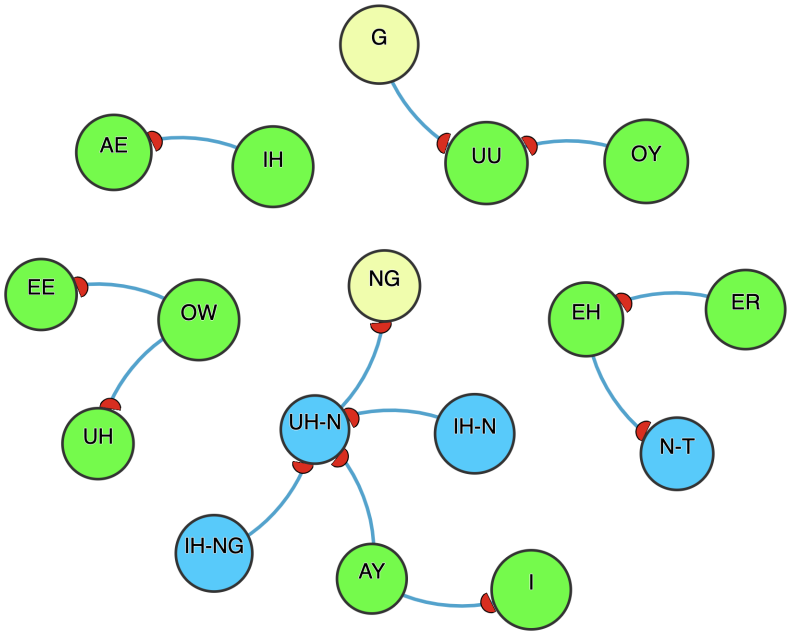


Fig. 5.15.: Confusions between vowels, consonants and chunks in the pooled results of the L10 tests with the new haptic codes

confused pair S-T→Z suggests that the participants only noticed the first half of the chunk and then misrecognized S as Z, which is coded at the same location as S. The aforementioned confusions between chunks and individual phonemes suggest that the duration of the signals may be too short. Nevertheless, the error rates are generally quite low.

In a similar fashion, Fig. 5.15 shows the confusions between vowels, chunks, a vowel and a consonant, a chunk and a vowel, and a chunk and a consonant. The graph was again generated after setting a minimum error per haptic symbol of 8%. The minimum error in the cluster of confusions was 8% (2 out of 25 trials) for both pairs AY→I and AY→UH-N. The maximum error was 17.86% (5 out of 28 trials) and corresponded to the confused pair EH→N-T. Two of the confusions among vowels correspond to perceived sensations that share the same direction in movements and the same locations on the arm, regardless of the type of haptic illusion used to code the phonemes. For example, OW and EE are movements from the wrist to the elbow

on the dorsal side of the arm. Even though EE was coded as a smooth movement and OW corresponds to a discrete movement using the cutaneous rabbit effect, the symbols may have been confused due to their direction and location. This is also the case for the pair OY→UU. As in the case of Fig. 5.14, other type of confusions among vowels may have involved momentary attention lapse, in which participants only focused on a fraction of a vowel, resulting in a confusion with a vowel whose movement felt like that of a fraction of the presented phoneme. This might be the case for the pair AY→I, in which the second portion of the movement evoked by I is similar to the sensation produced by AY. Additionally, the confusion IH→AE may have occurred due to the location of the phonemes on the arm. Even though the vowels correspond to different movements, they use the same tactors in the array and this might have been the source of the confusion. Other errors among vowels could be described as rare and random, since the pairs do not share any characteristics regarding location, direction, or type of movement. Such is the case for the two pairs OW→UH and EH→ER. A special confusion between a consonant and a vowel was discovered for G→UU which occurred in 5 out of 31 trials. In this case, the tactors at the end of the movement of UU correspond to the tactors at the middle volar side of the arm where G is coded. Moreover, both the consonant and the vowel were coded as 300-Hz signals with an amplitude modulation of either 25 Hz (in the case of G) or 30 Hz (in the case of UU). Since all vowels share a base frequency of 300 Hz, this similarity may have caused the confusion.

Confusions that involve chunks show different characteristics. First, as in Fig. 5.14, the pair UH-N→NG is another example of how participants might have focused on the second pulse of the chunk (the N portion) and mistakenly picked a phoneme located at the same location as the pulse. In this case, NG was the selected response. Another pattern was noticed in the confusions IH-N→UH-N and IH-NG→UH-N, where the IH portion of the chunks was not taken into account and the second pulse of the chunk ruled the decision. For these cases, the N and NG portions of the chunks both occur at the dorsal side close to the elbow, which is the only location used by the

two pulses of the UH-N chunk. A common error was to respond with either N or NG after focusing on only the second pulse of the presented chunk. Nevertheless, this occurred infrequently and the participants interpreted the last portion of the signal as the sensation produced by UH-N in 3 out of 30 trials for IH-NG→UH-N, and 5 out of 32 trials for IH-N→UH-N. Another type of confusion was evidenced in the pair AY→UH-N. AY was coded as a semicircle movement that starts and ends at the dorsal side close to the elbow. It may be the case that participants felt the start and end of the vowel to be the same sensation as the pulses produced by UH-N which was also coded at the dorsal side close to the elbow. Finally, the confusion EH→N-T represents the biggest error per haptic symbol from the whole data set with 5 out of 28 trials. The confusion is mainly due to the grabbing sensation of EH which involves two subsequent pulses, one at the dorsal and volar sides near the elbow, and another at the dorsal and volar sides at the middle of the arm. This two-pulse sensation is similar to the two-part sensation of all the chunks. Since the chunk N-T produces a first pulse at the dorsal side near the elbow (N portion) and a second pulse at the volar side of the middle of the arm (T portion), the phoneme EH was easily confused with this chunk.

6. CONCLUSIONS AND FUTURE WORK

Using a phonemic-based approach to tactile speech communication and the TAPS speech display, this thesis research made two important contributions. First, a role-playing game named Haptos designed for second language learning proved to be an effective mechanism to cumulatively train participants on the haptic codes and progressively identify words in an engaging way. Two participants in a pilot study using a first implementation of the game were able to master a list of 20 phonemes and 52 testing words within 45 minutes of game play. Using this game-based paradigm, a second version of the game was implemented to train and test eight naive participants on a set of improved haptic codes. Second, revisions to a first-generation of haptic symbols were made and new codes were designed to increase word presentation rates based on the statistics of phonemes in spoken English. The two design guidelines were to assign short duration codes to the most frequent phonemes and to use nearby locations for the most frequently co-occurring pairs of phonemes. In addition to improving the first-generation haptic symbols for the 39 phonemes of English, ten additional haptic symbols called "chunks" were created to abbreviate the top ten most frequent pairs of phonemes in terms of duration. The results indicate faster learning of more haptic symbols and a theoretical increase in speech transmission rates.

The symbol identification tests conducted in the second version of the game proved that the new set of codes was learnable within a reasonable amount of training time. The results showed that the full set of 49 symbols (39 improved phonemes and 10 new chunks) could be learned within 6 hours of training distributed over several days. An average percent correct score of 84.44% from the identification of 49 haptic symbols over 1,440 trials collected from eight participants showed that the game-based paradigm was an effective learning mechanism. Furthermore, the information transfer

estimate calculated from the tests with 49 symbols was 3.87 bits, which means that participants were able to recognize 14 symbols without error. Compared to previous codes studied in [1–3], the introduction of an extra set of 10 symbols and the redesign of the codes did not result in a significant drop in performance, compared to the phoneme identification tests in [3] that showed a performance level of 85.77% with ten participants and 1,560 trials.

The confusions derived from the final tests showed the error rates per haptic symbol to be below 17.86% for all 49 haptic symbols. These confusions were attributed to different design aspects of the signals, but the majority corresponded to confusion of symbols that shared the same location on the arm or used the same tactors in similar sequences. Confusions between the chunk signals and other haptic symbols indicated how participants sometimes focused only on one of the two pulses in a chunk, leading to an incorrect response with a phoneme that felt similar to the perceived pulse. The short duration of the chunk signals could have been the main cause of the confusions, as the time devoted to individual pulses of each chunk was only 50 or 100 ms. The special case of the phoneme EH confused with the chunk N-T represented the maximum error per haptic symbol in the data set. The confusion was attributed to the nature of the vowel that highly resembles the double-pulse design of the chunks. Regardless of the confusions observed, no significant confusions between vowels and consonants were observed for errors per haptic symbol above 8%. Overall, the error rates per haptic symbol for all trials and symbols were quite low.

With the high recognition rates and low error rates, the set of 39 improved phonemes are deemed highly distinct to be used for conveying speech information at a faster rate. Furthermore, the set of chunks can also potentially be used to replace common pairs of phonemes with a single code that shortens the duration of presenting the two phonemes. The potential of the new set of codes for increasing communication rates can be demonstrated by calculating the duration and word presentation rates of the following sentence: "**I have a year to be in the top ten of the chart and win that prize**." The phrase contains some of the most common

words in English. By transcribing the sentence into a phoneme sequence, the duration of the sentence presented through TAPS can be calculated as the sum of the durations of all the phonemes in the sentence. Using the first-generation haptic codes studied in [1–3], the sentence has a duration of 13.26 s. Using the new 39 codes developed in this thesis, the phrase has a reduced length of 10.68 s. Furthermore, by replacing the pairs of phonemes that appear in the list of the 10 most common co-occurring phonemes with their corresponding chunks, the sentence can be further reduced to 9.1 s. Word presentation rates for the sentence using these durations are calculated to be 76.9, 95.5, and 112.1 words per minute using the first-generation codes, the improved codes for 39 phonemes, and all 49 haptic symbols, respectively. Thus, the new set of codes represent a 24 to 46% increase in word presentation rates.

With these encouraging results, future work with the new set of 49 symbols will be developed in several directions. First, the ability of the set of 39 improved haptic symbols for phonemes to convey words needs to be tested. A new experiment using the original design of the Haptos game can evaluate how participants can learn English words using the improved codes. As discussed in Chapter 4, the design will include a modular implementation of the game based on storyline quests, and a separation between training and testing materials in order to train and test increasing sets of phonemes and words in a generalized way. Elements omitted in the first implementation of the game can now be incorporated. They include: presentation of redundant information in visual and/or auditory modalities, virtual characters that communicate in the tactile language, and an inter-code interval that separates phonemes delivered in a sequence to represent words through TAPS. These elements will support the training process by providing more contextual information while learning, enriching the game contents, and introducing other challenges in the process.

In terms of signal design, it was determined that the temporal characteristics of the pulses in chunks are the main cause of confusions between chunks and other phonemes. The duration of pulses within chunks and the temporal gap between the pulses may be too short for participants to identify the presence of a chunk and

individually identify each pulse. Thus, a study of the physical parameters of these signals is necessary, where the appropriate duration of the pulses and gap between them will be determined. It has been shown that temporal gap detection thresholds are as low as 10 to 5 ms using mechanical clicks as stimuli [53]. In TAPS, it would be worth studying these thresholds with the vibrotactile pulses produced by chunks, in order to determine their appropriate characteristics.

Furthermore, the use of chunks in words would also require experiments to determine word recognition accuracy when pairs of phonemes are replaced with their corresponding chunks. Despite the different confusions evidenced when presenting chunks, it is hypothesized that the additional contextual information provided by other phonemes in a word may allow participants to identify full words without the need of correctly identifying every single phoneme and/or pulse in a chunk. Testing this hypothesis extends to determining how contextual cues help restore missing information in a complex stimulus (e.g., a word as a sequence of phonemes). This restoration of information has been evidenced using the auditory sense, where participants are able to restore missing or masked phonemes from a recorded sentence in the so called "phoneme restoration effect" [54]. Testing whether there exists such an effect when speech information is presented through the skin can help determine if participants are able to understand words even if some of the phonemes or fractions of the phonemes are not fully perceived. This will then support the claim that chunks can be used in words even though they may not be as well perceived as two phonemes presented separately.

Experiments with single words in the Haptos game will also determine performance of participants in terms of single-word identification rates. As a second step towards the ultimate goal of using the device in real life conversations, experiments with multi-phoneme (> 3) words, two-word phrases, and sentences can be conducted. Ultimately, we want to achieve effective speech communication on the skin within a reasonable amount of training time, and at performance levels that improve with time and experience.

REFERENCES

REFERENCES

[1] J. Jung, Y. Jiao, F. M. Severgnini, H. Z. Tan, C. M. Reed, A. Israr, F. Lau, and F. Abnousi, "Speech communication through the skin: design of learning protocols and initial findings," in *Proceedings of HCI International*, 2018, pp. 15–20.

[2] Y. Jiao, F. M. Severgnini, J. S. Martinez, J. Jung, H. Z. Tan, C. M. Reed, E. C. Wilson, F. Lau, A. Israr, R. Turcott, K. Klumb, and F. Abnousi, "A comparative study of phoneme-and word-based learning of english words presented to the skin," in *Proceedings of EuroHaptics 2018*. Springer, 2018, pp. 623–635.

[3] C. M. Reed, H. Z. Tan, Z. D. Perez, E. C. Wilson, F. M. Severgnini, J. Jung, J. S. Martinez, Y. Jiao, A. Israr, F. Lau, K. Klumb, R. Turcott, and F. Abnousi, "A phonemic-based tactile display for speech communication," *IEEE Transactions on Haptics*, vol. 12, pp. 2–17, 2019.

[4] J. M. Wolfe, K. R. Kluender, D. M. Levi, L. M. Bartoshuk, R. S. Herz, R. L. Klatzky, and S. J. Lederman, *Sensation and perception*. Sinauer Associates, 2006.

[5] S. J. Bolanowski Jr, G. A. Gescheider, R. T. Verrillo, and C. M. Checkosky, "Four channels mediate the mechanical aspects of touch," *The Journal of the Acoustical society of America*, vol. 84, no. 5, pp. 1680–1694, 1988.

[6] R. S. Johansson and Å. B. Vallbo, "Tactile sensory coding in the glabrous skin of the human hand," *Trends in neurosciences*, vol. 6, pp. 27–32, 1983.

[7] S. Weinstein, "Intensive and extensive aspects of tactile sensitivity as a function of body part, sex and laterality," *The skin senses*, 1968.

[8] J. C. Craig and K. O. Johnson, "The two-point threshold: Not a measure of tactile spatial resolution," *Current Directions in Psychological Science*, vol. 9, no. 1, pp. 29–32, 2000.

[9] J. M. Loomis, "An investigation of tactile hyperacuity," *Sensory processes*, vol. 3, no. 4, pp. 289–302, 1979.

[10] R. T. Verrillo, "Effect of contactor area on the vibrotactile threshold," *The Journal of the Acoustical Society of America*, vol. 35, no. 12, pp. 1962–1966, 1963.

[11] C. M. Reed, W. M. Rabinowitz, N. I. Durlach, L. D. Braida, S. Conway-Fithian, and M. C. Schultz, "Research on the tadoma method of speech communication," *The Journal of the Acoustical society of America*, vol. 77, no. 1, pp. 247–257, 1985.

[12] C. M. Reed, "Tadoma: An overview of research," in *Profound deafness and speech communication.*, G. Plant and K. Spens, Eds. London: Whurr Publishers, 1995, pp. 40–55.

[13] S. Engelmann and R. Rosov, "Tactual hearing experiment with deaf and hearing subjects," *Exceptional Children*, vol. 41, no. 4, pp. 243–253, 1975.

[14] E. T. Auer Jr, L. E. Bernstein, and D. C. Coulter, "Temporal and spatio-temporal vibrotactile displays for voice fundamental frequency: An initial evaluation of a new vibrotactile speech perception aid with normal-hearing and hearing-impaired individuals," *The Journal of the Acoustical Society of America*, vol. 104, no. 4, pp. 2477–2489, 1998.

[15] F. A. Geldard, "Adventures in tactile literacy." *American Psychologist*, vol. 12, no. 3, p. 115, 1957.

[16] Y. Gaffary, F. Argelaguet, M. Marchal, A. Girard, F. Gosselin, M. Emily, and A. Lécuyer, "Toward haptic communication: Tactile alphabets based on fingertip skin stretch," *IEEE transactions on haptics*, vol. 11, no. 4, pp. 636–645, 2018.

[17] H. Z. Tan, N. I. Durlach, W. M. Rabinowitz, C. M. Reed, and J. R. Santos, "Reception of morse code through motional, vibrotactile, and auditory stimulation," *Perception & psychophysics*, vol. 59, no. 7, pp. 1004–1017, 1997.

[18] J. M. Weisenberger and M. E. Percy, "Use of the tactaid ii and tactaid vii with children." *Volta Review*, vol. 96, no. 5, pp. 41–57, 1994.

[19] C. M. Reed and L. A. Delhorne, "Current results of a field study of adult users of tactile aids," in *Seminars in Hearing*, vol. 16, no. 04. Copyright© 1995 by Thieme Medical Publishers, Inc., 1995, pp. 305–315.

[20] ——, "The reception of environmental sounds through wearable tactual aids," *Ear and hearing*, vol. 24, no. 6, pp. 528–538, 2003.

[21] L. A. Delhorne, J. M. Besing, C. Reed, and N. Durlach, "Tactual cued speech as a supplement to speechreading," *Cue Speech Journal*, vol. 6, pp. 71–87, 1998.

[22] M. S. Bratakos, C. M. Reed, L. A. Delhorne, and G. Denesvich, "A single-band envelope cue as a supplement to speechreading of segmentals: A comparison of auditory versus tactual presentation," *Ear and hearing*, vol. 22, no. 3, pp. 225–235, 2001.

[23] P. Ranjbar, E. C. Wilson, C. M. Reed, and L. D. Braida, "Auditory-tactile integration: Effects of phase of sinusoidal stimulation at 50 and 250 hz," *International journal of engineering technology and scientific innovation*, vol. 1, no. 2, p. 209, 2016.

[24] E. C. Wilson, C. M. Reed, and L. D. Braida, "Integration of auditory and vibrotactile stimuli: Effects of frequency," *The Journal of the Acoustical Society of America*, vol. 127, no. 5, pp. 3044–3059, 2010.

[25] G. A. Gescheider, *Psychophysics: method, theory, and application*, 2nd ed. Lawrence Erlbaum Associates, Inc., 1985.

[26] L. A. Jones and H. Z. Tan, "Application of psychophysical techniques to haptic research," *IEEE transactions on haptics*, vol. 6, no. 3, pp. 268–284, 2013.

[27] H. Levitt, "Transformed up-down methods in psychoacoustics," *The Journal of the Acoustical society of America*, vol. 49, no. 2B, pp. 467–477, 1971.

[28] N. A. Macmillan and C. D. Creelman, *Detection theory: A user's guide*. Psychology press, 2004.

[29] S. J. Lederman and L. A. Jones, "Tactile and haptic illusions," *IEEE Transactions on Haptics*, vol. 4, no. 4, pp. 273–294, 2011.

[30] R. Flach and P. Haggard, "The cutaneous rabbit revisited." *Journal of Experimental Psychology: Human Perception and Performance*, vol. 32, no. 3, p. 717, 2006.

[31] A. Barghout, J. Cha, A. El Saddik, J. Kammerl, and E. Steinbach, "Spatial resolution of vibrotactile perception on the human forearm when exploiting funneling illusion," in *Haptic Audio visual Environments and Games, 2009. HAVE 2009. IEEE International Workshop on*. IEEE, 2009, pp. 19–23.

[32] M. Niwa, R. W. Lindeman, Y. Itoh, and F. Kishino, "Determining appropriate parameters to elicit linear and circular apparent motion using vibrotactile cues," in *EuroHaptics conference, 2009 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics 2009. Third Joint*. IEEE, 2009, pp. 75–78.

[33] A. Israr and I. Poupyrev, "Tactile brush: drawing on skin with a tactile grid display," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2011, pp. 2019–2028.

[34] C. S. Green and D. Bavelier, "Action-video-game experience alters the spatial resolution of vision," *Psychological science*, vol. 18, no. 1, pp. 88–94, 2007.

[35] R. Li, U. Polat, W. Makous, and D. Bavelier, "Enhancing the contrast sensitivity function through action video game training," *Nature neuroscience*, vol. 12, no. 5, p. 549, 2009.

[36] T. Wade and L. L. Holt, "Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task," *The Journal of the Acoustical Society of America*, vol. 118, no. 4, pp. 2618–2633, 2005.

[37] G. Kimball, R. Cano, J. Feng, L. Feng, E. Hampson, E. Li, M. G. Christel, L. L. Holt, S.-j. Lim, R. Liu *et al.*, "Supporting research into sound and speech learning through a configurable computer game," in *Games Innovation Conference (IGIC), 2013 IEEE International*. IEEE, 2013, pp. 110–113.

[38] S.-j. Lim and L. L. Holt, "Learning foreign sounds in an alien world: Videogame training improves non-native speech categorization," *Cognitive science*, vol. 35, no. 7, pp. 1390–1405, 2011.

[39] S. D. Krashen, "Principles and practice in second language acquisition," 1982.

[40] J. J. Shepherd, R. J. Doe, M. Arnold, N. Cheek, Y. Zhu, and J. Tang, "Lost in the middle kingdom: a second language acquisition video game," in *Proceedings of the 49th Annual Southeast Regional Conference*. ACM, 2011, pp. 290–294.

[41] K. P. Jantke and T. Hume, "Effective learning through meaning construction in digital role playing games," in *Consumer Electronics (ICCE), 2015 IEEE International Conference on*. IEEE, 2015, pp. 653–656.

[42] N. Lane and N. R. Prestopnik, "Diegetic connectivity: Blending work and play with storytelling in serious games," in *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM, 2017, pp. 229–240.

[43] P. Wouters, H. van Oostendorp, and E. D. Van der Spek, "Game design: The mapping of cognitive task analysis and game discourse analysis in creating effective and entertaining serious games," in *Proceedings of the 28th Annual European Conference on Cognitive Ergonomics*. ACM, 2010, pp. 287–293.

[44] L. D. Grace, "Polyglot cubed: the design of a multi-language learning game," in *Proceedings of the International Conference on Advances in Computer Enterntainment Technology*. ACM, 2009, pp. 421–422.

[45] G. A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information." *Psychological review*, vol. 63, no. 2, p. 81, 1956.

[46] C. E. Shannon, "A mathematical theory of communication," *Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[47] H. Z. Tan, N. I. Durlach, C. M. Reed, and W. M. Rabinowitz, "Information transmission with a multifinger tactual display," *Perception & Psychophysics*, vol. 61, no. 6, pp. 993–1008, 1999.

[48] Y. Dudai, A. Karni, and J. Born, "The consolidation and transformation of memory," *Neuron*, vol. 88, no. 1, pp. 20–32, 2015.

[49] X. Cai, "Principles of human-computer interaction in game design," in *2009 Second International Symposium on Computational Intelligence and Design*, vol. 2. IEEE, 2009, pp. 92–95.

[50] J. S. Martinez, H. Z. Tan, and C. M. Reed, "Improving haptic codes for increased speech communication rates on a phonemic-based tactile display," 2019, draft paper.

[51] P. B. Denes, "On the statistics of spoken english," *The journal of the Acoustical Society of America*, vol. 35, no. 6, pp. 892–904, 1963.

[52] M. A. Mines, B. F. Hanson, and J. E. Shoup, "Frequency of occurrence of phonemes in conversational english," *Language and speech*, vol. 21, no. 3, pp. 221–241, 1978.

[53] R. T. Verrillo and G. A. Gescheider, "Perception via the sense of touch," in *Tactile aids for the hearing impaired*, I. R. Summers, Ed. London: Whurr Publishers, 1992, ch. 1, pp. 1–36.

[54] R. M. Warren, "Perceptual restoration of missing speech sounds," *Science*, vol. 167, no. 3917, pp. 392–393, 1970.

APPENDICES

A. LABELS FOR PHONEME IPA SYMBOLS

Table A.1.: Phoneme labels for consonants

Consonants					
IPA Symbol	**Label**	**Word Example**	**IPA Symbol**	**Label**	**Word Example**
/p/	P	**P**age	/b/	B	**B**ell
/t/	T	**T**ea	/d/	D	**D**ish
/k/	K	**C**andle	/g/	G	**G**old
/f/	F	**F**ish	/v/	V	Lo**v**e
/θ/	TH	**Th**orn	/ð/	DH	**Th**e
/s/	S	Che**s**t	/z/	Z	Ro**se**
/ʃ/	SH	**Sh**ell	/ʒ/	ZH	Trea**su**re
/ʧ/	CH	**Ch**ess	/ʤ/	J	Ima**ge**
/m/	M	**M**ap	/n/	N	**Kn**ob
/ŋ/	NG	Wi**ng**	/h/	H	**H**at
/w/	W	**W**ine	/r/	R	**R**ock
/l/	L	**L**eaf	/j/	Y	**Y**ellow

Table A.2.: Phoneme labels for vowels

Vowels					
IPA Symbol	**Label**	**Word Example**	**IPA Symbol**	**Label**	**Word Example**
/i/	EE	K**ey**	/ɑ/	AH	**A**rch
/u/	OO	F**oo**d	/æ/	AE	P**a**n
/ɔ/	AW	S**aw**	/ɝ/,/ɚ/	ER	**Ea**rth
/ʌ/	UH	Th**e**	/ɪ/	IH	Sh**i**p
/ɛ/	EH	Br**ea**d	/ʊ/	UU	W**oo**d
/eɪ/	AY	**A**ce	/aɪ/	I	S**i**gn
/aʊ/	OW	H**ou**se	/ɔɪ/	OY	T**oy**
/ɔʊ/	OE	B**o**ne			

B. LEARNING MATERIAL IN THE FIRST VERSION OF HAPTOS

Table B.1.: Learning material for the first version of Haptos

Group	**Phonemes**		**Training Words**	**Testing Words**
	Consonants	**Vowels**		
1	D, M, S	AY, EE, OO	day, doom, do, seam, may, mace, say, maim, moo, made, dame, deed	aid, moose, me, mood, sue, seed, see, deem, ace, dude, same, mead
2	W, DH, K	I	die, weed, way, wide, sigh, meek, woo, came, they, make, key, wake, dime	my, wade, we, womb, why, seek, thy, cake, thee, sake, kay, side, coup, mime
3	SH, V, R	AH, UH	she, rave, shoe, shame, shade, rock, vase, mush, sock, duck, some, come	shy shave, ray, shove, us, dock, mock, shock, rush, rum, wash, raid, read

Continued on next page

Table B.1 – continued from previous page

Group	Phonemes		Training Words	Testing Words
	Consonants	Vowels		
4	B, L, TH	OE, IH	low, rub, show, thumb, though, wish, so, dish, lime, like, bike, bill	bow, limb, row, will, oak, lake, oath, base, the, rim, dim, woke, dome
5	CH, Z, H	AE, OW	vow, hose, wow, him, home, chose, choose, cows, choke, sad, mad, maze	how, hatch, cow, has, chow, chum, char, bath, loud, vowed, dad, bad, cheese
6	P, N, J	OY, EH	jay, pen, pay, pan, knee, nap, keep, noise, hen, men, bed, shed	now, pin, no, pool, joy, join, chin, gym, them, den, then, when, check
7	T, G, ZH	ER, UU	too, put, tie, could, guy, would, gay, azure, what, look, book, learn, turn	toy, tug, toe, gut, go, gun, nurse, burn, should, but, dirt, shirt
8	F, Y, NG	AW	fee, fool, foe, foil, you, fowl, tall, fought, sing, sang, pawn, your	off, fun, on, fan, or, young, all, yawn, ought, ring, rang, thing, wing

C. SOFTWARE LIBRARY ARCHITECTURE

The software developed to perform phoneme identification experiments required a custom and specialized library that could play phonemes in the TAPS device outside the MATLAB environment.

For this purpose, a general library was built on the C and C++ languages with a C# wrapper as an extensible plugin for C# applications (such as the Unity application used in the symbol identification experiments). The library offers different services to Windows applications that are not only restricted to playing individual phonemes. Services also include: playing sequences of phonemes, playing the sequences of phonemes from a string sentence written in English (using an underlying text-to-speech software and a phoneme transcription of the sentence), and playing arbitrary 24-channel audio signals.

The library is composed of 3 main components compiled as `.dll` files offered to C# applications. The files also make use of a pre-installed application named CMU Flite (http://www.festvox.org/flite/) that offers third-party text-to-speech services. This application is used by the main component to obtain the phoneme transcription of a sentence written in regular English. Fig. C.1 shows a simple deployment diagram that shows how the components are loaded into a Windows C# application and how the services are exposed and used in the Windows environment.

In the diagram, all services are packed into the "MOTU services" interface which are exposed to client applications via a singleton instance of the "Motu" class component. Using this singleton, an application can access all the available services using thread-safe operations and callback functions from a single and global instance of the Motu component. The Motu singleton uses underlying calls to static functions exposed by a "MotuCore" component in order to offer playback services. This component depends on a compiled version of the PortAudio I/O library (http:

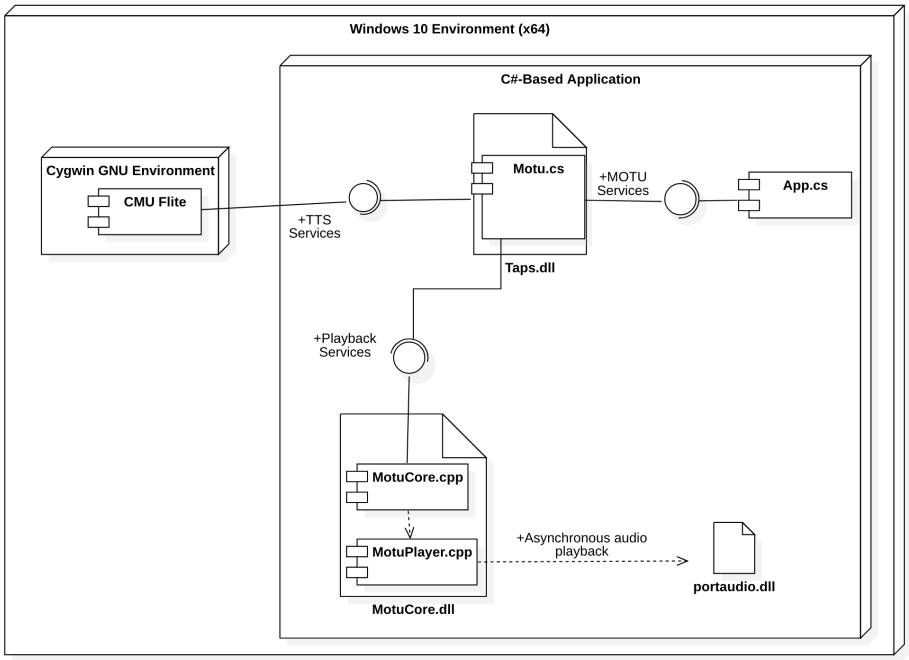


Fig. C.1.: Deployment diagram of the software library

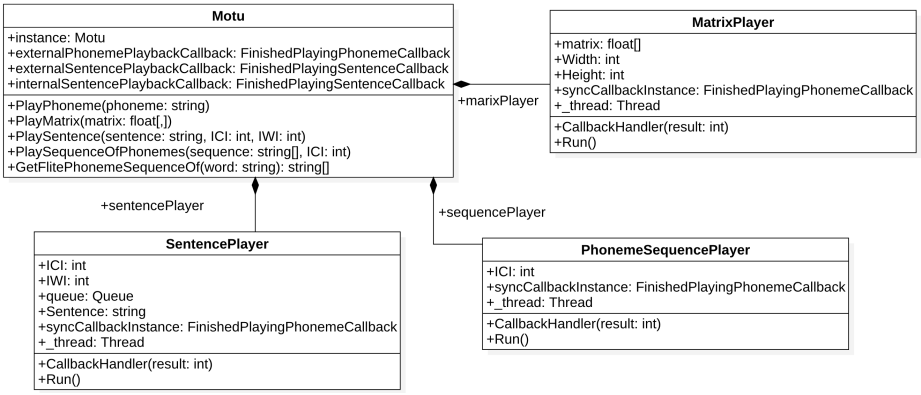


Fig. C.2.: Simplified class diagram of the Motu singleton component

//portaudio.com/) named as “portaudio.dll” to play audio data into the TAPS interface (see Fig. C.1).

In order to play arbitrary 24-channel audio data, sequences of phonemes, and regular English sentences, the Motu singleton uses inner components that use thread-safe operations and callback functions to control timing between playback of words,

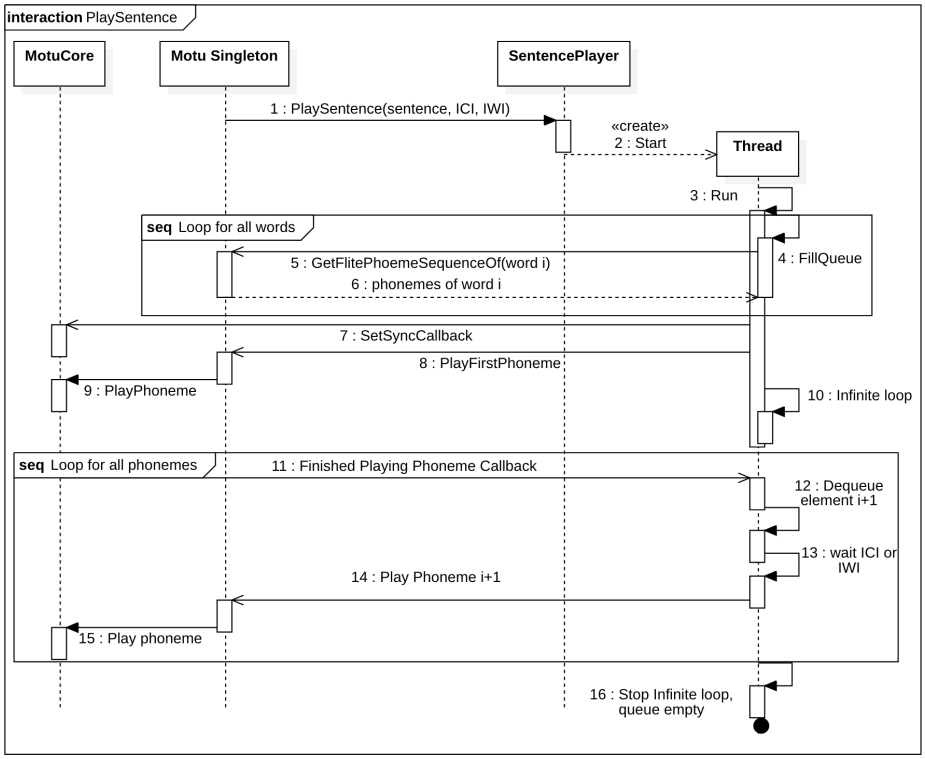


Fig. C.3.: Sequence diagram for playing a sentence in the software library

and playback of phonemes within words or within sequences of phonemes. A simplified class diagram of the singleton with these inner components is shown in Fig. C.2.

The first component is named the "SentencePlayer" and is activated by calling the operation `PlaySentence(string sentence, int ICI, int IWI)` from the Motu singleton. This component activates and fills a queue of phonemes and pause events to be played in the MOTU device using a separate thread. The phonemes are obtained from the phoneme transcription of the sentence using the CMU Flite software (deployed on a GNU environment through the Cygwin software: https://www.cygwin.com/). Fig. C.3 shows a sequence diagram of all the interactions that take place while playing a sentence. First, with the given string sentence, the component uses the `GetFlitePhonemeSequenceOf(string word)` to obtain a phoneme transcription of all the words in the given sentence. This is done by calling the Flite software for every word in the sentence. The phonemes of every word are stored in a

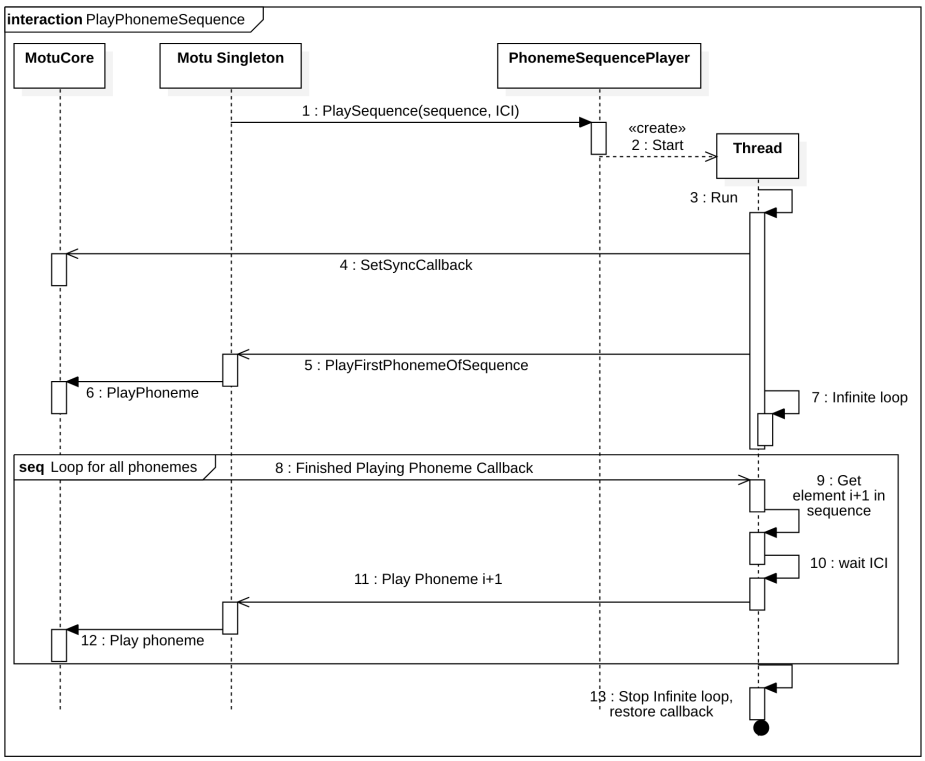


Fig. C.4.: Sequence diagram for playing a sequence of phonemes in the software library

queue for them to be played after all words have been transcribed. The thread also adds pause events to the queue between words. After the queue is full, the thread sets a synchronization callback into the MotuCore component. This callback is a function that gets called every time the MotuCore component finishes playing a phoneme in the device. The sentence playing thread uses this function as a timing mechanism to control time intervals between phonemes and between words. Once the callback is set, the thread plays the first element in the queue and enters an infinite loop. Once the first phoneme is played, the synchronization callback will be called, this operation triggers a routine in the thread that dequeues the next element to be played. If the element is a phoneme, the thread waits for the specified inter-code interval (ICI) time and plays the phoneme in the device, if the element is a word, a pause corresponding to the inter-word interval (IWI) is made. This process repeats for every element in

the queue. Once the queue is empty, the infinite loop in the thread breaks and the operation finishes.

A similar routine is performed when a sequence of phonemes is to be played. The sequence diagram that shows this operation using the "PhonemeSequencePlayer" component in shown in Fig. C.4. As opposed to Fig. C.3, the phoneme sequence player does not require the transcription of sentences to phoneme sequences. Thus, the operation starts by creating a thread from the phoneme sequence player with the phoneme sequence and an ICI interval used to pause between phonemes. The thread then sets the same synchronization callback as in the previous operation (playing a sentence), plays the first phoneme in the sequence, and enters an infinite loop to wait for the end of the operation. As in Fig. C.3, the synchronization callback gets called every time a phoneme finishes playing on TAPS. This triggers a routine that dequeues the next phoneme in the sequence, and plays it on the device after a pause of exactly the specified ICI time. Once the queue is empty, the infinite loop breaks and the operation finishes.

An additional service added to the library is the functionality of playing any arbitrary 24-channel signal in the device as a matrix of `float` values. Fig. C.5 shows the sequence diagram that the "MatrixPlayer" component uses to perform this operation. The operation begins by calling the `PlayMatrix` function of the MatrixPlayer component. The function receives the matrix to be played, the width of the matrix (24 columns), and the height (the number of rows). As in the previous operations, the MatrixPlayer component initializes a thread that sets a synchronization callback. This callback will be used by the MotuCore component to alert the thread that the matrix finished playing. The thread then enters an infinite loop to wait until the callback is triggered. When the signal (matrix) finishes playing on the device, the loop breaks and the operation finishes.

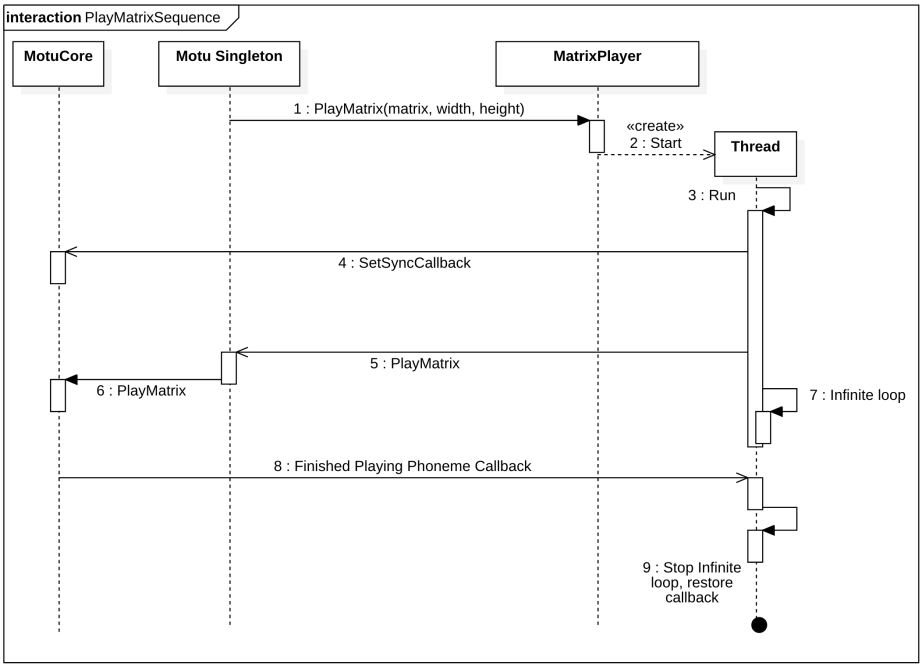


Fig. C.5.: Sequence diagram for playing an arbitrary matrix in the software library

D. IMPROVED HAPTIC CODES DESCRIPTION

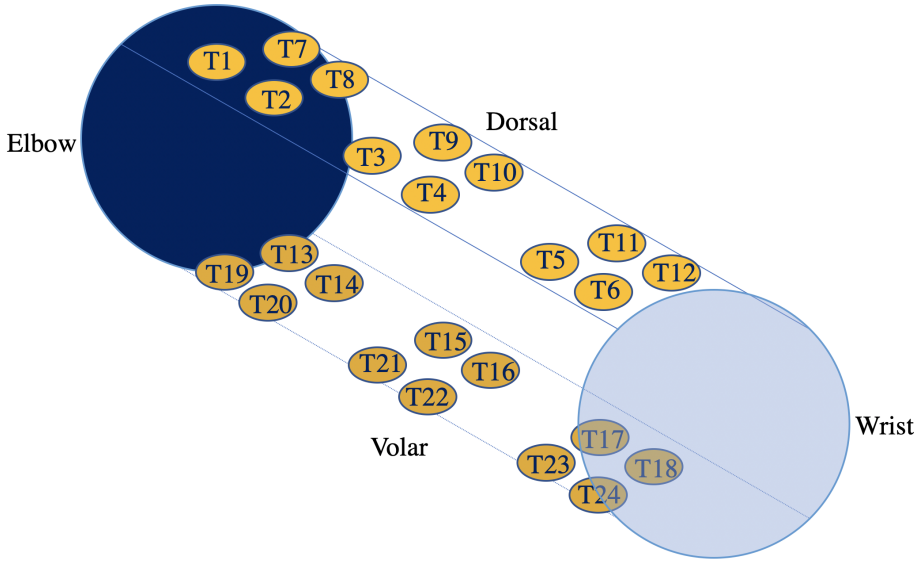


Fig. D.1.: Tactor layout of the phonemic-based tactile display

The newly designed haptic codes based on the decision rules derived from the statistics of spoken English are defined on Table. D.1, Table. D.2 and Table. D.3. All the parameters are based on the tactor layout presented in Fig. D.1. In the tables, all modulations with sinusoidal tones have a modulation index of 1 and "Cos^2" stands for a cosine squared modulation.

Table D.1.: Description of improved consonant codes

Phoneme Code	**Waveform: Amplitude, Frequency, Modulation**	**Location**		**Duration (ms)**	**Tactors Involved**
		Transverse	**Longitudinal**		
P	25 dBSL, 300 Hz	Dorsal	Wrist	140	T5, T6 T11, T12
T	30 dBSL, 60 Hz	Volar	Mid-forearm	100	T15, T16 T21, T22
K	25 dBSL, 300 Hz	Dorsal	Elbow	100	T1, T2 T7, T8
B	25 dBSL, 300 Hz, 25 Hz	Dorsal	Wrist	140	T5, T6 T11, T12
D	30 dBSL, 60 Hz	Dorsal	Wrist	100	T5, T6 T11, T12

Continued on next page

Table D.1 – continued from previous page

Phoneme Code	Waveform: Amplitude, Frequency, Modulation	Location		Duration (ms)	Tactors Involved
		Transverse	**Longitudinal**		
G	25 dBSL, 300 Hz, 25 Hz	Volar	Mid-forearm	140	T15, T16 T21, T22
CH	25 dBSL, 300 Hz, Cos^2	Dorsal	Wrist and elbow	400	T1, T6 T7, T12
J	25 dBSL, 300 Hz, 9 Hz	Dorsal	Wrist and elbow	400	T1, T6 T7, T12
F	25 dBSL, 300 Hz	Dorsal and volar	Wrist	400	T6, T12 T18, T24

Continued on next page

Table D.1 – continued from previous page

Phoneme Code	Waveform: Amplitude, Frequency, Modulation	Location		Duration (ms)	Tactors Involved
		Transverse	**Longitudinal**		
V	25 dBSL, 300 Hz, 9 Hz	Dorsal and volar	Wrist	400	T6, T12 T18, T24
TH	25 dBSL, 300 Hz, 20 Hz	Dorsal	Mid-forearm	140	T3, T4 T9, T10
DH	30 dBSL, 60 Hz	Dorsal	Mid-forearm	100	T3, T4 T9, T10
S	30 dBSL, 60 Hz	Volar	Elbow	100	T13, T14 T19, T20

Continued on next page

Table D.1 – continued from previous page

Phoneme Code	Waveform: Amplitude, Frequency, Modulation	Location		Duration (ms)	Tactors Involved
		Transverse	**Longitudinal**		
Z	25 dBSL, 300 Hz, Cos^2	Volar	Elbow	400	T13, T14 T19, T20
SH	25 dBSL, 300 Hz, Cos^2	Volar	Wrist	400	T17, T18 T23, T24
ZH	25 dBSL, 300 Hz, 25 Hz	Volar	Wrist	140	T17, T18 T23, T24

Continued on next page

Table D.1 – continued from previous page

Phoneme Code	Waveform: Amplitude, Frequency, Modulation	Location		Duration (ms)	Tactors Involved
		Transverse	**Longitudinal**		
H	25 dBSL, 60 Hz, Cos^2	Dorsal and volar	Mid-forearm	400	T4, T5 T10, T11 T16, T17 T22, T23
M	25 dBSL, 60 Hz, 15 Hz	Volar	Mid-forearm	400	T15, T16 T21, T22
N	30 dBSL, 60 Hz	Dorsal	Elbow	100	T1, T2 T7, T8
NG	25 dBSL, 300 Hz, 25 Hz	Dorsal	Elbow	140	T1, T2 T7, T8

Continued on next page

Table D.1 – continued from previous page

Phoneme Code	Waveform: Amplitude, Frequency, Modulation	Location		Duration (ms)	Tactors Involved
		Transverse	**Longitudinal**		
L	25 dBSL, 300 Hz, 15 Hz	Dorsal and volar	Mid-forearm and wrist	100	T9, T10 T11, T12 T21, T22 T23, T24
R	25 dBSL, 300 Hz, 30 Hz	Volar	Elbow	100	T13, T14 T19, T20
W	25 dBSL, 60 Hz, 9 Hz	Dorsal and volar	Mid-forearm and wrist	400	T9, T10 T11, T12 T21, T22 T23, T24

Continued on next page

Table D.1 – continued from previous page

Phoneme Code	Waveform: Amplitude, Frequency, Modulation	Location		Duration (ms)	Tactors Involved
		Transverse	**Longitudinal**		
Y	30 dBSL, 60 Hz	Volar	Wrist	100	T17, T18 T23, T24

Table D.2.: Description of improved vowel codes

Phoneme Code	**Waveform: Amplitude, Frequency, Modulation**	**Location**	**Movement**	**Tactors Involved**	**Duration (ms)**	**Subjective Impression**
AE	25 dBSL, 300 Hz	Dorsal	Pulses that create a circular motion at the dorsal side of the wrist	T4, T5, T6 T10, T11, T12	400	"Twinkle" sensation
AH	25 dBSL, 60 Hz	Dorsal	Longitudinal, from the elbow to the wrist	T1, T2, T3 T4, T5, T6 T7, T8, T9 T10, T11, T12	400	Long movement
OE	25 dBSL, 300 Hz, Cos^2	Dorsal and volar	Circular movement around the mid-forearm	T3, T4, T9 T10, T15, T16 T21, T22	400	Circular ring

Continued on next page

Table D.2 – continued from previous page

Phoneme Code	Waveform: Amplitude, Frequency, Modulation	Location	Movement	Tactors Involved	Duration (ms)	subjective Impression
EH	25 dBSL, 300 Hz	Dorsal and volar	From the elbow to the mid-forearm on the volar and dorsal sides	T1, T4, T7 T10, T13, T16 T19, T22	240	Grabbing sensation near the elbow and the mid-forearm
ER	25 dBSL, 300 Hz	Volar	Pulses that create a circular motion at the volar side of the elbow	T13, T14, T15 T19, T20, T21	400	"Twinkle" sensation
IH	25 dBSL, 300 Hz	Dorsal	Longitudinal, from the wrist to the mid-forearm	T3, T4, T5 T6, T9, T10 T11, T12	240	Short movement

Continued on next page

Table D.2 – continued from previous page

Phoneme Code	Waveform: Amplitude, Frequency, Modulation	Location	Movement	Tactors Involved	Duration (ms)	subjective Impression
EE	25 dBSL, 300 Hz	Dorsal	Longitudinal, from the wrist to the elbow	T1, T2, T3 T4, T5, T6 T7, T8, T9 T10, T11, T12	400	Long movement
UH	25 dBSL, 300 Hz	Dorsal	Pulses that create a circular motion at the dorsal side close to the elbow	T1, T2, T3 T7, T8, T9	400	"Twinkle" sensation
OO	25 dBSL, 300 Hz, 30 Hz	Volar	Longitudinal, from the wrist to the elbow	T13, T14, T15 T16, T17, T18 T19, T20, T21 T22, T23, T24	400	Long movement

Continued on next page

Table D.2 – continued from previous page

Phoneme Code	Waveform: Amplitude, Frequency, Modulation	Location	Movement	Tactors Involved	Duration (ms)	subjective Impression
UU	25 dBSL, 300 Hz, 30 Hz	Volar	Longitudinal, from elbow to the wrist	T13, T14, T15 T16, T19, T20 T21, T22	240	Short movement
AW	25 dBSL, 300 Hz	Volar	Pulses that create a circular motion at the volar side of the wrist	T16, T17, T18 T22, T23, T24	400	"Twinkle" sensation

Continued on next page

Table D.2 – continued from previous page

Phoneme Code	Waveform: Amplitude, Frequency, Modulation	Location	Movement	Tactors Involved	Duration (ms)	subjective Impression
AY	25 dBSL, 300 Hz, 30 Hz	Dorsal	Parabolic shape drawn starting on the mid-forearm, passing through the wrist and ending at the mid-forearm	T3, T4, T5 T6, T9, T10 T11, T12	400	Rumbling sensation

Continued on next page

Table D.2 – continued from previous page

Phoneme Code	Waveform: Amplitude, Frequency, Modulation	Location	Movement	Tactors Involved	Duration (ms)	subjective Impression
I	25 dBSL, 300 Hz, 30 Hz	Dorsal and volar	Movement from the volar side close to the elbow to the mid-forearm and crossing to the dorsal mid-forearm, ending at the dorsal side close to the elbow	T1, T2, T3 T4, T7, T8 T9, T10, T13 T14, T15, T16 T19, T20, T21 T22	400	Sweeping motion and rumbling sensation

Continued on next page

Table D.2 – continued from previous page

Phoneme Code	Waveform: Amplitude, Frequency, Modulation	Location	Movement	Tactors Involved	Duration (ms)	subjective Impression
OW	25 dBSL, 300 Hz	Dorsal	Cutaneous rabbit from wrist to the elbow; 3 taps on each tactor	T8, T10, T11	400	Tapping sensation
OY	25 dBSL, 300 Hz	Volar	Cutaneous rabbit from elbow to the wrist; 3 taps on each tactor	T20, T22, T24	400	Tapping sensation

Table D.3.: Description of chunk codes for the most frequent co-occurring pairs of phonemes

Phoneme Pair	**First Stimulus Parameters**		**Second Stimulus Parameters**		**Location**	**Duration (ms)**
	Amplitude, Frequency, Modulation	**Tactors Involved**	**Amplitude, Frequency, Modulation**	**Tactors Involved**		
DH-UH	30 dBSL, 60 Hz	T4, T10	25 dBSL 300 Hz, Cos^2	T1, T2 T7, T8	Dorsal	200
UH-N	25 dBSL, 300 Hz, Cos^2	T1, T2 T7, T8	30 dBSL 60 Hz	T1, T2 T7, T8	Dorsal	200
T-UH	30 dBSL, 60 Hz	T15, T16 T21, T22	25 dBSL 300 Hz, Cos^2	T1, T2 T7, T8	Dorsal and volar	200
N-D	30 dBSL, 60 Hz	T1, T2 T7, T8	30 dBSL 60 Hz	T5, T6 T11, T12	Dorsal	100

Continued on next page

Table D.3 – continued from previous page

Phoneme Pair	First Stimulus Parameters		Second Stimulus Parameters		Location	Duration (ms)
	Amplitude, Frequency, Modulation	Tactors Involved	Amplitude, Frequency, Modulation	Tactors Involved		
S-T	30 dBSL, 60 Hz	T13, T19	30 dBSL 60 Hz	T16, T22	Volar	100
IH-T	25 dBSL, 300 Hz, Cos^2	T5, T6 T11, T12	30 dBSL 60 Hz	T15, T16 T21, T22	Dorsal and volar	200
IH-N	25 dBSL, 300 Hz, Cos^2	T5, T6 T11, T12	30 dBSL 60 Hz	T1, T2 T7, T8	Dorsal	200
IH-NG	25 dBSL, 300 Hz, Cos^2	T5, T6 T11, T12	25 dBSL 300 Hz, 30 Hz	T1, T2 T7, T8	Dorsal	200

Continued on next page

Table D.3 – continued from previous page

Phoneme Pair	First Stimulus Parameters		Second Stimulus Parameters		Location	Duration (ms)
	Amplitude, Frequency, Modulation	**Tactors Involved**	**Amplitude, Frequency, Modulation**	**Tactors Involved**		
N-T	30 dBSL, 60 Hz	T1, T2 T7, T8	30 dBSL 60 Hz	T15, T16 T21, T22	Dorsal and volar	100
Y-OO	30 dBSL, 60 Hz	T17, T18 T23, T24	25 dBSL 300 Hz, Cos^2	T13, T14 T15, T19 T20, T21	Volar	200