VIBROTACTILE SPEECH COMMUNICATION:

PERCEPTUAL STUDIES WITH A PHONEME-BASED DISPLAY

A Thesis

Submitted to the Faculty

of

Purdue University

by

Jaehong Jung

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science in Mechanical Engineering

August 2020

Purdue University

West Lafayette, Indiana

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF THESIS APPROVAL

Dr. Hong Z. Tan, Co-Chair

School of Electrical and Computer Engineering, Purdue University

Dr. Patricia Davies, Co-Chair

School of Mechanical Engineering, Purdue University

Dr. George T. C. Chiu

School of Mechanical Engineering, Purdue University

Dr. Charlotte M. Reed

Research laboratory of Electronics, Massachusetts Institute of Technology

Approved by:

Dr. Nicole Key

Head of the School Graduate Program

ACKNOWLEDGMENTS

I would like to thank Professor Hong Z. Tan for her advice and guidance throughout my graduate study and for being a friendly mentor. I sincerely appreciate Dr. Reed's willingness to spend time giving advice to me and all the contributions that she has made to the work presented in this thesis. I would also like to thank Professor Patricia Davies and Professor George T. Chiu for serving on my thesis committee and providing help for my graduate study. Special thanks to Juan Sebastian Martinez for developing the TCP protocol and TTS code adaption for the 4-phoneme word acquisition study presented in Chapter 4, and for being a mentor regarding graduate study. I truly appreciate all the advice and help towards my thesis offered by Gina Marie Clepper. It was very helpful to have colleagues in the lab be involved as participants in the studies presented in this thesis.

I would also like to thank Facebook Inc. and Google LLC for supporting the research presented in this thesis.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

ABSTRACT

Jung, Jaehong , Purdue University, August 2020. Vibrotactile Speech Communication: Perceptual Studies with a Phoneme-Based Display. Major Professor: Hong Z. Tan.

Tactile communication systems provide an alternative channel of communication for people with all levels of sensory capabilities and can help those with sensory impairments to receive information through another sensory modality. Recently, a TActile Phonemic Sleeve (TAPS) has been developed with the objective of enabling people to "hear through the skin." This thesis presents three studies that evaluate the feasibility of the TAPS system for phoneme and word acquisition and for two-way tactual communication. The TAPS system is based on a phonemic-based coding scheme that uses an array of 24 (6-by-4) tactors to convey haptic stimuli on the forearm. In Study 1, an effective mechanism for learning phonemes and words with TAPS based on the theory of memory consolidation was explored. Four naive participants learned to recognize 51 words made up of 10 phonemes within 60 min of experimental time. A fifth naive participant demonstrated the ability to learn all 39 phonemes of the English language as haptic codes after a total of 80 min with a phoneme identification score of 93.8%. We found that with the distinctive set of haptic symbols that had been developed prior to this thesis, participants were able to learn phonemes and words in a short amount of time. We also validated the memory consolidation theory by showing an improvement in phoneme recognition score when the fifth naive participant was re-assessed the day after he had learned the phonemes. In Study 2, we evaluated the learning performance of longer (four-phoneme) words. A total of three experienced participants spent 20 min per day for 3 days to review 39 phonemes and 500 words (with most of the words containing two or three phonemes) that they had learned in an earlier study. They then spent 10 more days to practice and test with a word list consisting of 100 four-phoneme words (List #1). A generalization study was conducted

by testing the same participants with a different set of 100 four-phoneme words (List #2) during the last 2 days of the experiment. After the 15-day experimental period, the average word percent-correct (PC) scores of the three participants for List #1 and List #2 were 80.2% and 72.3%, respectively. Both results were well above the corresponding chance levels (1% for the closed set of words in List #1 and near 0 for the open set of words in List #2) which demonstrated that the participants were able to learn longer words with the TAPS system within a reasonable amount of time. In Study 3, the feasibility of TAPS for tactile communication of spontaneous speech was evaluated. Two of the three experienced participants from Study 2 sent text messages to each other through two identical TAPS systems with an open set of words. The average percent-correct (PC) scores for the two participants for messages (PC_{msg}) and words (PC_{word}) were 73.4% and 81.7%, respectively. These results are impressive considering that the participants had to recognize words and phrases using an open vocabulary. Overall, the three studies demonstrate that the users of the TAPS system can successively receive phonemes, isolated words up to 4 phonemes in length, phrases, and sentences in a two-way exchange that simulates daily communication scenarios. Future work will explore the design of additional haptic symbols for conveying punctuation marks and investigate the efficacy of the TAPS system in helping people with sensory impairments to communicate via the sense of touch.

1. INTRDODUCTION

The studies conducted in this thesis were motivated by the need for tactile devices that serve as communication when the normal channels of audition and/or vision are compromised. A number of studies have tried to develop tactile devices to deliver speech to the skin and to enable successful tactual communication. To achieve this goal, several approaches have been used by extracting certain properties of the speech, including spectral-based, letter-based, and phoneme-based approaches. The spectral-based approach is based on encoding the acoustic signals of speech to actuators located on the skin. It has been used in several tactile aids like the Tactaid VII [1], tactile vocoder [2], and VEST [3]. Last but not least, the phonemic-based approach is a new way that takes the phonemes and encodes them into haptic symbols that are transmitted to the skin. It was used in several devices worn on the forearm or upper-arm such as the display in Zhao et al. (2018) [4], the MISSIVE device in Dunkelberger et al. (2018) [5], and the TActile Phonemic Sleeve (TAPS) system in Reed et al. (2019) [6] and Tan et al. (2020) [7].

Although the above-mentioned tactile aids were able to transmit speech information to the skin, some limitations of each approach still exist. For the spectral-based approach, a considerable time of learning is required. For example, it took as much as 80.5 hours for one participant to acquire 250 words in Brooks et al.'s (1985) study [8]. In addition to a long period of learning, the spectral-based approach confronts a limitation when dealing with spectral properties and token variations. For example, it was found with Tactaid VII that the frequency of the first formant of a vowel (/i/) could have a similar frequency to that of another vowel (/ʊ/) due to an on-glide, a transitional sound generated by articulators when moving to another position to create a new speech sound, which could make word identification more difficult [9]. Across different speakers, and even within the same speaker, there will be variations in the spectral properties of a given speech sound across productions. Hence, the receiver faces the difficulty of learning different tactile signals for the same words [10].

The letter-based approach is a method that encodes the letters of a language into haptic symbols and conveys them to the skin. This approach was used in a 6-Channel haptic display worn on the hand and forearm [11]. The letter-based approach has a relatively slower communication rate as compared to the phonemic-based approach. This is because that given any English word, the number of phonemes is always equal to or less than the number of letters in the word. Therefore, to the extent that users can learn 39 phonemes as opposed to 26 letters, something that has already been proven in several earlier studies using the TAPS system (e.g., [12], [13], [6], [7]), it is preferable to use phonemes as the smallest "building blocks" for a tactile speech communication system, as explained below.

The author of this thesis is part of a research team that developed the TAPS system based on a phonemic-based coding scheme and conducted a series of studies to assess its efficacy in tactile communication of speech. The TAPS system is worn on the forearm and uses 24 (a 6×4 array) tactors that deliver tactile stimuli on the skin. It encodes the 39 phonemes of English (24 consonants and 15 vowels) into 39 haptic symbols with different durations, frequencies, locations of stimuli, patterns of movement, and with the presence of amplitude modulation [6]. An initial study of the TAPS system confirmed the feasibility of the haptic symbols and developed a learning strategy based on memory consolidation [12]. Another study compared a phoneme-based learning approach to a word-based learning approach in the acquisition of 100 words, and concluded that the phoneme-based learning curriculum led to a more consistent learning progress among the participants [13]. It was also found that the inter-phoneme interval in words and the inter-word interval in two-word phrases can be reduced to 75 ms and 500 ms, respectively, without a significant impact on word recognition performance [14]. Recently, it has been demonstrated that the best users of the TAPS system can acquire up to 500 English words at a learning rate of one word per minute [7]. Compared to many other studies, the studies on TAPS have demonstrated the ability to transmit any English word to the skin of the forearm within a reasonable learning period.

This thesis reports three studies of which the author led the efforts. Initially, an effective mechanism for learning phonemes and words with TAPS was developed based on memory consolidation. It was found that an efficient learning curriculum was to break up learning over a long period of time (days to weeks) with a relatively short (10-20 min) daily practice. This strategy was utilized in subsequent studies involving the TAPS system. This study has since been published [12] and appears in Chapter 3 of this thesis. The second study expanded upon previous studies involving TAPS by using longer (four-phoneme) words, and it appears in Chapter 4 of this thesis. The third and last study was conducted under a more ecologically valid setting, where two participants sent text messages composed on-the-fly to each other and received the messages using TAPS. This study appears in Chapter 5. The rest of this thesis is structured as follows. Chapter 2 introduces related background of this thesis. Chapter 6 concludes this thesis with an outlook for future work.

2. BACKGROUND

This chapter provides a review of the background literature that is relevant to this thesis research and to the design of the phonemic-based tactile speech communication system called TActile Phonemic Sleeve (TAPS) (see Jung et al., 2018 [12], Jiao et al., 2018 [13], Reed et al., 2019 [6], Reed et al., 2020 [14], and Tan et al., 2020 [7]). It starts with an overview of the tactile sensory system responsible for receiving speech information on the skin, followed by a description of psychophysical methodology for evaluating human performance with haptic displays. The literature review continues with a discussion of tactile movement illusions that are effective at conveying information through the skin, and ends with a summary of natural and device-mediated tactile speech communication methods and systems.

2.1 Tactile Sensory System

Haptics refers to sensing and manipulation through touch. When a person touches an object, a force and displacement can be felt on the skin and this sensory information is delivered to the brain through the nervous system. Tactual perception can be classified into two components: tactile perception and kinesthetic perception. Tactile perception denotes the sensation arising through the contact with an object mediated by four types of mechanoreceptors responding to stimulation on the skin. Kinesthetic perception refers to the sense of limb position, motion and muscle tension mediated by sensory receptors in muscles, tendons and around joints [15]. In this chapter, we focuse on tactile perception because the tactile device used in the present study, the TAPS system, stimulates the skin of the forearm with a 4-by-6 tactor array. This section covers two topics of tactile perception: mechanoreceptors and their detection thresholds for vibrations, and two-point spatial discrimination thresholds on the skin.

2.1.1 Mechanoreceptors and Detection Thresholds

Tactile receptors exist in the skin of human's body, including the outermost layer called "epidermis" and the layer located below epidermis named "dermis". These tactile receptors can be classified into four types of receptors known as "mechanoreceptors" which react to mechanical stimulation such as pressure and vibration. The four receptors are called Meissner corpuscles (FA I), Merkell cell neurite complexes (SA I), Pacinican corpuscles (FA II), and Ruffini endings (SA II), categorized based on their adaptation rates (slow or fast) and sizes of their receptive fields (small or large) (Chapter 12, [16]). Figure 2.1 displays the cross section of the skin along with the relative locations of the four mechanoreceptors. A fast-adapting (FA) receptor usually reacts during the moment when the stimulus is either applied or removed. A slow-adapting (SA) receptor reacts during the period when the stimulus on the skin is slowly or barely changing [16].

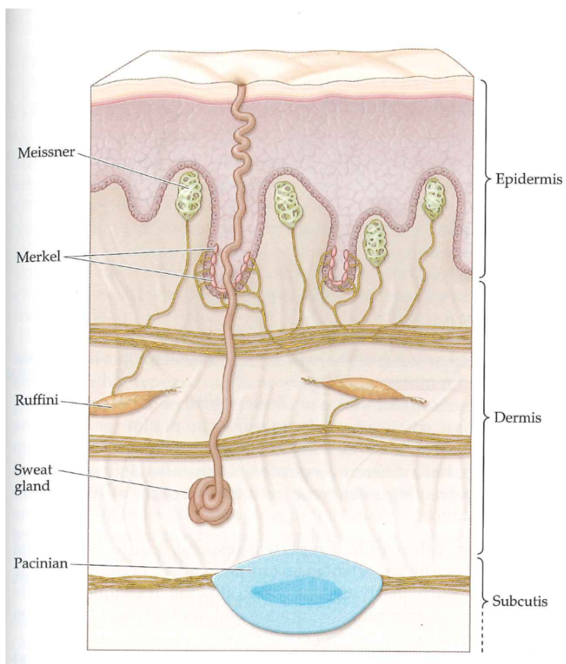


Fig. 2.1.: Cross section of the skin (from [16] with source data from [17]).

The Meissner corpuscle is one of the fast-adapting mechanoreceptors with a small receptive field that detects the rate of skin deformation. Especially, they are responsible for minute skin motion and provides information for grip control [18, 19]. The Pacinian is the other fast-adapting mechanoreceptors with a large receptive field that responds to vibrations with high frequency [19]. It provides information regarding perception of a distant event by detecting the vibration delivered through an object [20]. The Merkel cell neurite complexes are slow-adapting mechanoreceptors with a small receptive field which is responsible for detecting spatial features such as form and texture perception [18, 19]. The Ruffini endings are also slow-adaptive mechanoreceptors but with a large receptive field that is responsible for sensing the stretch of the skin. It also provides information for perception of the direction of the force exerted by an object [19, 20]. A summary of the mechanoreceptors is provided in Table 2.1.

Table 2.1.: Summary of the four mechanoreceptors regarding adaption rate, size of receptive field, and primary functions (adapted from [16], [18]).

Adaptation Rate	Size of Receptive Field	Type of mechanoreceptor	Primary Functions
Fast	Small	FA I (Meissner)	Detection of slipping between the skin and an object
	Large	FA II (Pacinian)	Perception of a conveyed vibration when grasping an object with the hand
Slow	Small	SA I (Merkel)	Texture perception and form detection
	Large	SA II (Ruffini)	Perception of motion of an object

The four mechanoreceptors have different frequency sensitivity ranges for responding to mechanical stimulators. Figure 2.2 illustrates the frequency sensitivity range of each mechanoreceptor. The ordinate axis represents the threshold, a minimally detectable displacement, in decibels relative to 1 μm, while the abscissa axis is the frequency of a single-frequency vibration. In the figure, there are four colored lines and black dotted marks. The green line represents the detection thresholds for SA I (Merkel) and it shows that the thresholds range between 18 dB and 27 dB over a frequency range from 0.4 to 100 Hz. It is observed that the threshold becomes lower as the frequency of vibration increases, and the lowest threshold point is around 18 dB at 100 Hz. Data for FA I (Meissner) is presented in the purple-colored line. The threshold gradually decreases as the frequency increases from 4 to 30 Hz, and the threshold increases once the frequency is increased beyond 30 Hz. Hence, the minimum threshold of FA I is approximately 13 dB when the frequency reaches 30 Hz. The red and blue lines display the data of SA II (Ruffini) and FA II (Pacinian), respectively. Regarding FA II, the threshold decreases almost linearly as the frequency increases from 10 to approximately 200 Hz, reaches a minimum at 300 Hz, and increases again as the frequency rises above 300 Hz. The threshold of SA II decreases as the stimulating frequency increases from 20 to 250 Hz and rises again above 250 Hz. In addition, the minimum threshold values for FA II (blue line) and SA II (red line) are equal to −17 dB and 8 dB at around 250-300 Hz, respectively. The black dotted line presents detection thresholds measured psychophysically and are determined by the mechanoreceptors that are most sensitive at each frequency [21]. Based on Figure 2.2, humans are most sensitive (i.e., detection threshold is the lowest) at −17 dB in the frequency range of 250 to 300 Hz.

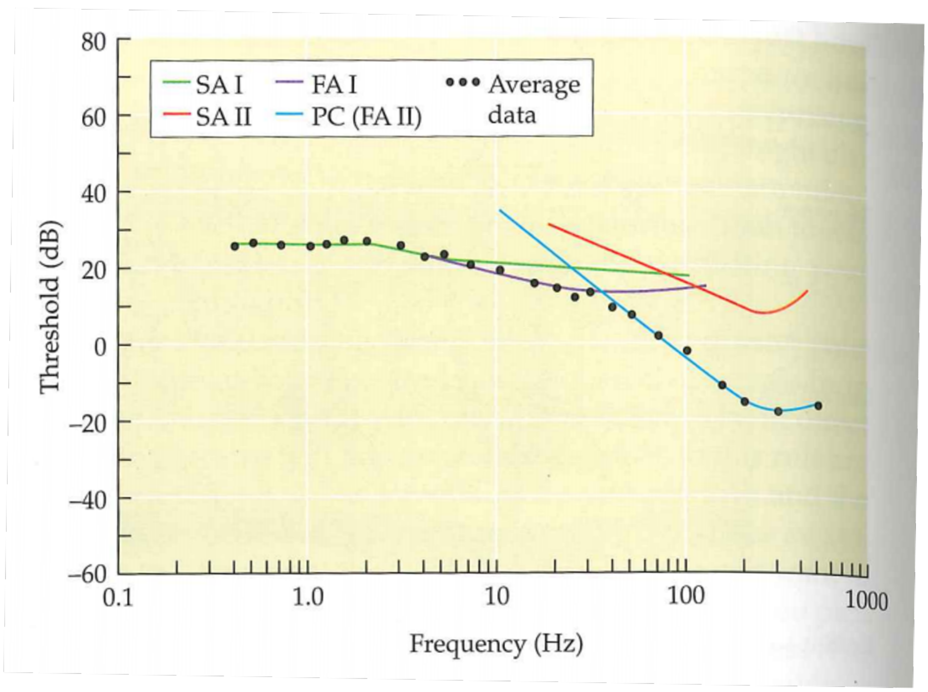


Fig. 2.2.: Sensitivity ranges of the mechanoreceptors (from [16] with source data from [21]).

The size of a stimulator has an effect on the detection thresholds. Figure 2.3 shows how thresholds vary with contact area as measured on the thenar eminence of the right hand as a function of vibrating frequency. The ordinate represents the detection (threshold in decibels relative to 1 μm. The figure shows data for different contact areas of stimulators ranging from 0.005 to 5.1 cm^2. The lowest threshold for each contact area appears near

250 Hz. Among the lowest threshold values of different sizes of stimulators, the maximum sensitivity (i.e., lowest threshold) is achieved for an optimal contactor size around 3.0 cm^2 or higher [22], [23].

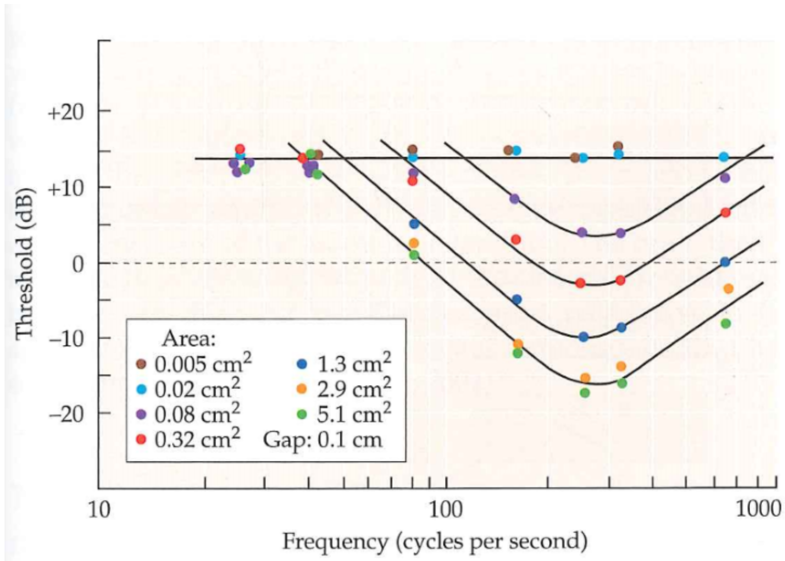


Fig. 2.3.: Detection threshold (dB re 1 μm) of a vibrating stimulus applied to the palm as a function of frequency and contact area (from [16] with source data from [22]).

Figure 2.4 shows the threshold shift for shorter stimuli compared to the threshold measured for a stimulus with a duration of 1000 ms at the same frequency. The figure clearly shows that detection thresholds become higher as the duration of the stimulus becomes shorter, meaning that a short-term stimulus is harder to detect than a longer one [24]. In addition, the threshold barely decreases as the duration of the stimulus exceeds 700 ms and the rate of decline of the threshold becomes zero after the duration goes beyond 1000 ms. In other words, detection thresholds no longer decrease further once the stimulus duration is longer than a second.

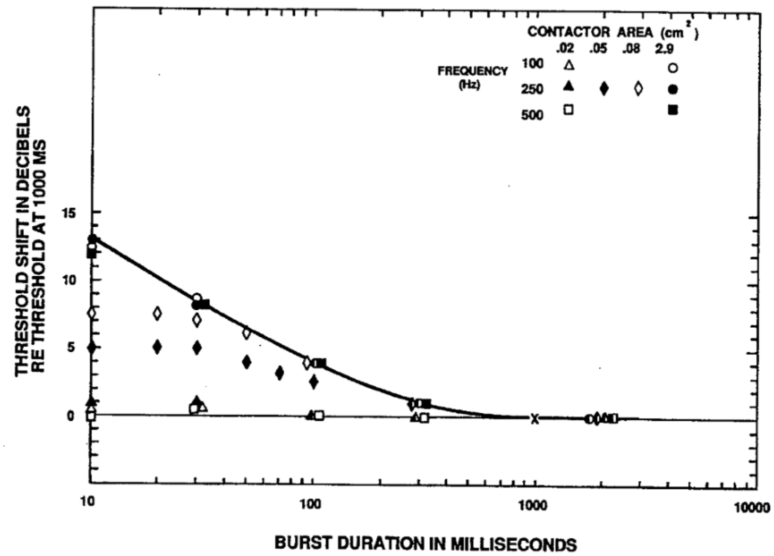


Fig. 2.4.: Detection thresholds of vibrotactile stimuli presented at the thenar eminence as a function of frequency (from Figure 4 in [24]).

In summary, the role of mechanoreceptors and information about tactile spatial acuity, and dependency of threshold on contact area and duration of a stimulus were studied. These concepts provides a brief perspective of how to design a tactile aid that could interact with people through speech communication delivered by tactile cues.

2.1.2 Tactile Spatial Acuity

Exposed to various types of mechanical stimulation, humans have different tactile sensitivity that varies with different parts of their bodies. The two-point touch threshold refers to the minimum separation between two points that can be reliably judged as two separate stimuli rather than a single stimulus. Figure 2.5 shows the mean of the two-point touch threshold values on different areas of the body from the Hallux up to the Forehead (with data from [25]). In Figure 2.5, the two-point thresholds for the left (red) and right (blue) sides of the body are presented. It can be observed that the minimum two-point threshold is located on the fingers of the hand for both the left and right hands. The maximum two-point thresholds are at the shoulder, thigh, and calf on both the left and right sides.

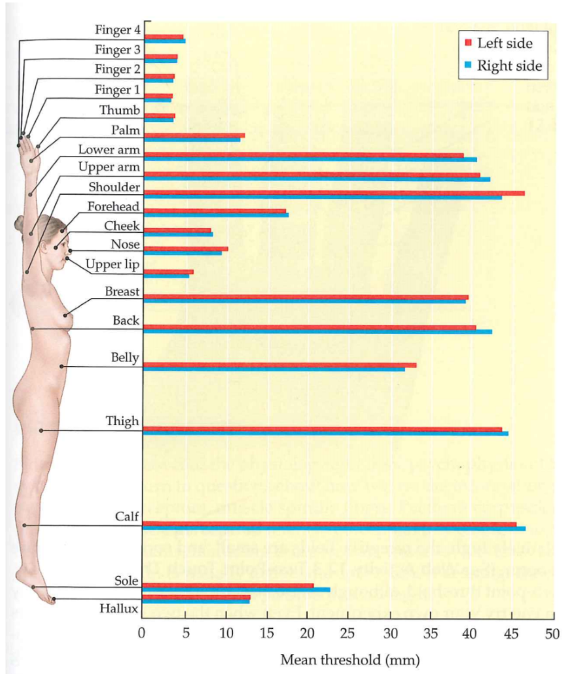


Fig. 2.5.: Two-point limen, the minimal separation between two points to feel them as separate points, as a function of body sites (from [16] with source data from [25]).

Weber discovered that the perceived distance between two tactile stimuli is larger for the regions of the body having higher spatial acuity, denoted as *Weber's Illusion* [26], [27]. In other words, the two-point threshold is low when a stimulated skin region of the body consists of a high density of receptors and a small receptive area [16]. In addition, an object is perceived to be larger when it is placed across the forearm than along the length of the forearm [28]. When touched with the dorsal surface of the hand, an object is perceived to be 30% - 40% larger when placed across the hand than along the hand [29].

2.2 Tactile Movement Illusions

This subsection provides an overview of the various tactile movement illusions that have played an important role in the encoding of English phonemes into tactile symbols [6]. Tactile movement illusions include the mislocalization of tactile stimulation, and perceived movement that is characterized by smooth motions or discrete, hopping sensations. They are distinct, easy to learn and memorize, and learning occurs effortlessly without special attention or cognitive load on the user's part. We discuss four illusions where the temporal and spatial properties of stimuli are misinterpreted by the somatosensory system: tau and kappa effects, apparent motion, funneling, and sensory saltation.

2.2.1 Tau and Kappa Effects

Tau effect refers to an illusion that the distance between stimulated points are perceived to be either closer or farther away than the actual distance between them. If three vibrotactile stimuli are presented consecutively where the physical distance between the first and second stimuli is twice the distance between the second and third stimuli, but the time interval between the first and second stimuli is half of that between the second and third stimuli, the perceived distance between the second and third stimuli would be twice as large as the perceived distance between the first and second stimuli. Among the consecutively aligned stimulated points, it was also demonstrated that the perceived distance between two tactors felt shorter when the stimulus moved at a faster rate (250 cm/s) compared to that with a velocity of 1 cm/s [30]. This states that the temporal parameters of a stimulus affects the perceived distance. However, the velocity does not necessarily form a linear relationship with the perceived distance between stimulated points. For example, the velocity in the range from 5 to 20 cm/s hardly had any effect on the perceived distance [30].

Kappa effect addresses the distortion of the judgement of temporal intervals based on different spatial separations of tactile stimuli [31]. Several experiments have shown that the spatial parameter affected the estimation of time interval between stimulated locations. For example, the perceived time interval between the first two stimulated points was longer when the actual distance between the first two points was longer than that between the second and third ones [32].

Further research has been done by using the concepts of tau and kappa effects. While the tau effect describes the effect of temporal parameters on the perceived spatial separation between consecutively stimulated points, the kappa effect illustrates the opposite effect of perceived time interval between successive stimuli being affected by the spatial parameters. Using the tactile mislocalization caused by either spatial or temporal parameters, research has discovered novel mislocalization effects. As a result, three additional tactile illusions were found on the skin: Apparent motion, funneling, and sensory saltation.

2.2.2 Apparent Motion

Apparent motion describes the illusion of a stimulus moving smoothly on the skin. In order to create an apparent motion, two parameters need to be controlled: *Duration of Stimulus* (DOS) and *Inter-Stimulus Onset Asynchrony* (SOA). Duration of Stimulus is defined as the time period of a vibration on one tactor whereas the Stimlulus-Onset Asynchrony is the time interval between the onset of two successively activated tactors. Figure 2.6 is a diagram for both DOS and SOA for two successively activated stimulators A and B [33]. As shown in the figure, two signals are delivered to two stimulators, A and B, with periods of ON/OFF of the stimuluators marked. DOS for both signals are indicated. The SOA is shown as the interval between the onset of stimulator A and the onset of stimulator B. To create an apparent motion, the DOS is controlled in order to govern the speed of the tactile movement by activating successive stimuluators. Although the velocity of the tactile movement can be increased by decreasing the DOS, the SOA has to be adjusted to provide an optimal apparent motion. Once the SOA becomes too large, participants will perceive the successive stimuli as discrete events. On the contrary, the participants will feel the successive stimuli as a single stimulus if the SOA is too small because the two stimuli will be played nearly stimultaneously [34]. Israr and Poupyrev (2011) derived an equation for determining SOA as a function of durations of two consecutive stimuli that creates an optimal apparent tactile motion [34].

Multiple studies have tried to create apparent motion by varying numbers of stimulators. A study has shown that participants who were supposed to identify the direction of the tactile motion could achieve a percent-correct score of greater than 95% for a linear apparent motion on the bicep if the circuit-completion time (=SOA$\times$Number of tactors)

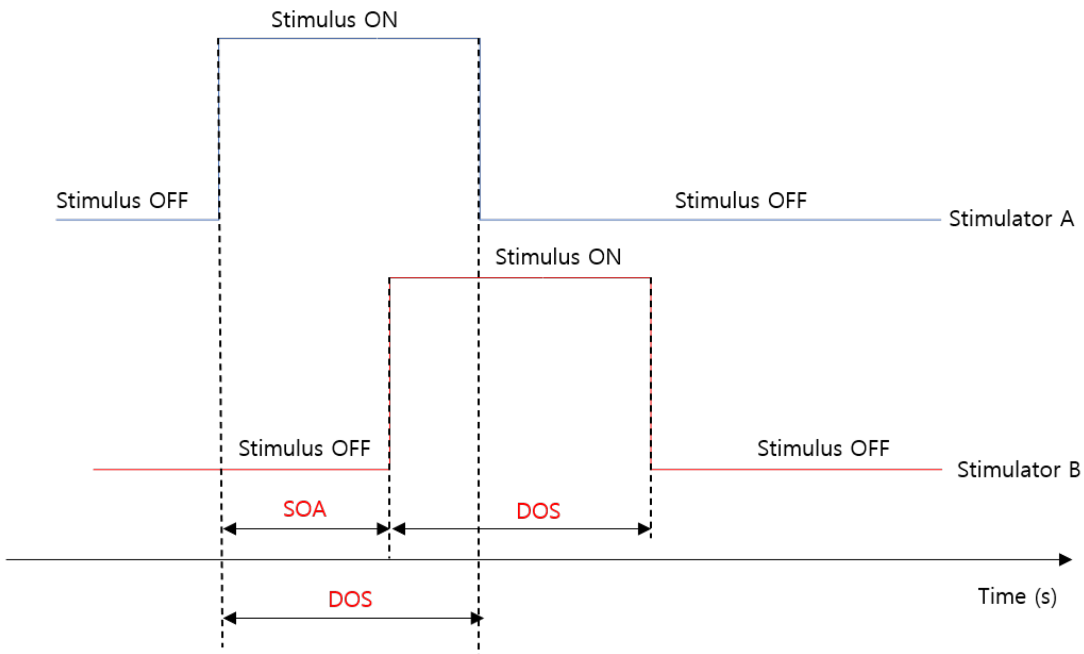


Fig. 2.6.: Illustration of Stimulus-Onset Asynchrony (SOA) and Duration of Stimulus (DOS) (modified from Figure 1 in [33]).

was greater than 400 ms [33]. To create a circular apparent motion on the upper arm with an improved direction identification rate close to 100%, participants needed at least four tactors around their upper arm [33]. It was also shown that an increase in the number of tactors stimulated consecutively on the skin leads to a decrease in the minimum SOA required to provide a clear apparent motion, resulting in a 75% correct discrimination of movement directions [33]. The critical SOA is approximately 120 ms for four tactors and decreases to 20 ms for 12 tactors [35]. Showing that the use of four stimulators led to a higher percent-correct score of movement judgment than that of two stimulators, Kirman demonstrated that the optimal interstimulus onset interval (ISOI), same as SOA, decreased as the number of stimulators operated in a row increased [36].

2.2.3 Funneling

Funneling effect describes a phantom perception that occurs at the location between multiple simultaneously presented stimuli [37]. When two stimulators are activated simultaneously with equal perceived intensity, the phantom sensation is perceived at the midpoint between the stimulators [31]. The location of the phantom sensation can be controlled by varying the relative intensities of the stimulators. The perceived location of the phantom sensation is closer to the stimulator with the stronger intensity [38]. For two actuators closely located, the intensity of a virtual actuator A_v induced by the funneling effect can be controlled by the intensities of the two actuators (A_1 and A_2) based on the energy summation model as shown below [34].

$$A_v^2 = A_1^2 + A_2^2 \tag{2.1}$$

Barought (2009) further described that the funneling effect could induce both an illusory static mislocalization and a vibrotactile movement. However, he suggested that the spacing between acutators needed to be approximately 40 to 80 mm to create a dynamic vibrotactile movement [37].

The funneling effect is useful since it can create multiple virtual actuators with a small number of real actuators. This is especially helpful for a tactile device such as a wearable haptic device or tactile speech communication device that has a limited number of stimulators in a confined space but need to provide sufficient amount of information such as the TActile Phonemic Sleeve (TAPS) (see [7]).

2.2.4 Sensory Saltation

One of the most frequently used tactile illusions is *sensory saltation*. With a sequence of short successive stimuli at different locations on the skin, a series of discrete stimuli moving in one direction is perceived. It is also called the "*cutaneous rabbit*" effect since it felt as if a tiny rabbit was hopping on the skin [31]. It has been shown that sensory saltation can be perceived when the interstimulus interval (ISI), the temporal separation between the offset of a stimulus and the onset of the upcoming stimulus, ranges from 40 to 200 ms and the number of taps is between 2 and 12 [39]. Geldard further suggested that sesnsory saltation could be perceived when the spatial separation between actuators ranges from 2 cm to 35 cm. He furthermore demonstrated that the most vivid sensory saltation effect occurs when the ISI reaches between 40 and 60 ms and the number taps is between 4 and 6 [39]. Further studies showed that the cutaneous rabbit effect provides distinct directional information. With a sensory saltation display on the back, participants were able to identify the directions of a series of tactile movement cues with an average percent-correct score of 83% [40].

The sensory saltation effect provides another illusory mislocalization on the skin which is useful for delivering infromation through a haptic deviceFor example, it is critical to have abundant haptic cues to achieve tactile speech communication. Unlike the apparent motion or phantom effect, the cutaneous rabbit is a novel illusory effect because it creates a 'discrete' movement rather than a 'continuous' movement. Hence, it provides another form of haptic information and this sensory illusion can be used to improve tactile speech communication [40], [31].

2.3 Psychophysics for Haptics Research

In this section, four methods for measuring perceptual performance are explained with their procedures and data analyses: Method of Constant Stimuli, Method of Limits, Method of Adjustment, and Adaptive Procedures. In addition, the concept of threshold is introduced since it quantifies sensitivity of the sensory system [41]. It is important to know the methods for measuring sensitivity (i.e., reciprocal of detection threshold) because the perceived intensity of vibrotactile stimuli is quantified relative to detection thresholds (Chapter 1, [41]). A stimulus that is below the detection threshold cannot be felt, and one that is more than 50 dB above the detection threshold starts to feel uncomfortable. Quantifying detection thresholds using psychophysical methods is the first step towards providing a stimulus intensity that is neither too strong nor too weak with tactile communication devices such as TAPS introduced in [7] or wearable haptic systems designed for rehabilitation and prosthetics introduced in [42] and [43].

Interested in the human sensory system, Ernst H. Weber and Gustav F. Fechner, natural scientists and physiologists, concentrated on finding a method of measuring the limitation of human sensory systems. The field of psychophysics studies the relationship between sensations in the perceptual domain and stimuli in the physical domain, and proposes models characterizing the relationship. Two sensory limits that are frequently studied and quantified through psychophysical methods are *absolute threshold* (AL) and *difference threshold* (DL). The AL is obtained from a detection experiment and refers to the lowest stimulus energy required to form a sensation that can be reliably detected. The absolute threshold or AL is also called a detection threshold. The DL or difference threshold is also called a discrimination threshold or *just noticeable difference* (jnd). It can be derived from a discrimination experiment and is defined as the smallest difference between two stimuli that can be reliably detected [41].

Classical psychophysics or Fechnerian psychophysics incorporates the statistical properties of thresholds in estimating thresholds using one of the following three methods: Method of Constant Stimuli, Method of Limits, and Method of Adjustment. Each method can be used for estimating both detection and discrimination thresholds, and the procedures and data analyses are different. They are discussed separately in Sections 2.3.1 to 2.3.3, respectively. Adaptive procedures are more efficient at deriving detection and discrimination thresholds and enjoy widespread use in the current psychophsycial literature. This method will be presented in Section 2.3.4.

2.3.1 Method of Constant Stimuli

Detection Threshold (AL)

Within a detection experiment, *method of constant stimuli* often consists of a set of five to nine different stimuli varying in intensity in a detection experiment. A participant receives a stimulus and the task is to respond "Yes" if the stimulus is detected and "No" otherwise. A psychometric function is plotted based on the proportion of "Yes" responses as a function of stimulus intensity. A fitted curve, typically a cumulative Gaussian distribution, is then estimated and drawn. From this fitted curve, the detection threshold is derived as the stimulus intensity corresponding to 50% of "Yes" responses, as presented in Figure

2.7. During the experiment, catch trials are often used where there the stimulus intensity is 0 (i.e., no stimulus). This is to reduce the response bias on the part of the participants. The responses to the catch trials can be used to gauge response bias. If the proportion of responses as "Yes" on catch trials exceeds 30%, then there is too much bias.

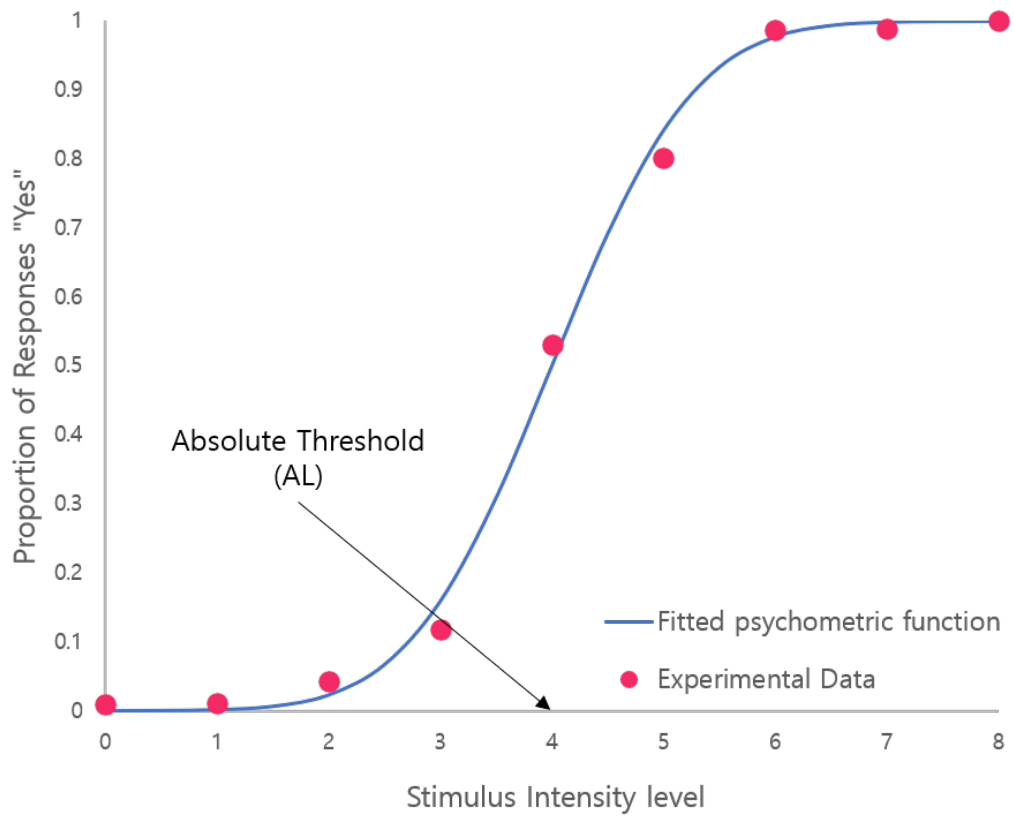


Fig. 2.7.: A representative psychometric function based on a detection experiment.

Discrimination Threshold (DL)

During a discrimination experiment a pair of stimuli is presented on each trial (a 2-interval experiment) or one of the two is presented (a 1-interval experiment). Within this pair of stimuli, one is a fixed stimulus called the reference or standard stimulus and the other is named the test or comparison stimulus that changes from trial to trial in a random order. The participant is asked to judge whether the stimulus is the "Greater" (e.g., longer, heavier, louder, etc.) one in a 1-interval experiment, or if the first (or second) stimulus is the "Greater" one in a 2-interval expeirment. Results are plotted as the proportion of "Greater" responses as a function of stimulus intensity of the comparison stimulus, and a psychometric function is fitted to the data. Figure 2.8 presents a representative model of such a psychometric function.

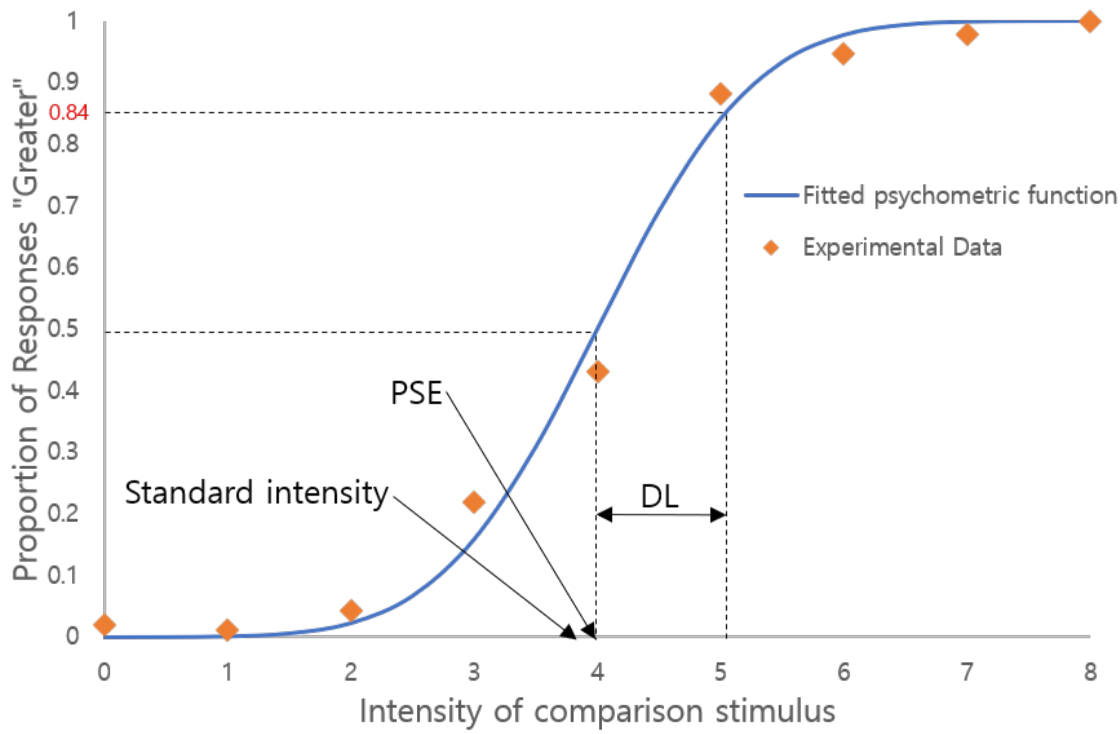


Fig. 2.8.: A representative psychometric function based on a discrimination experiment.

In this psychometric function acquired from the discrimination experiment, the intensity of the comparison stimulus at the 50-percentile point is defined as the "Point of Subjective Equality" (PSE). After estimating the psychometric function, the DL is measured as the difference between the 84-percentile point and PSE.

2.3.2 Method of Limits

Method of Limits is one of the most popular procedures to estimate sensory thresholds such as absolute or difference thresholds. One of the advantages of this method is that it takes less time to estimate the sensory thresholds compared to the method of constant stimuli, reducing participants' fatigue during the experiment.

Detection Threshold (AL)

In a detection experiment, the absolute threshold (AL) can be estimated by conducting an equal number of the following two types of experiments: ascending and descending series. An *ascending series* starts with a stimulus intensity that is well below the threshold and increases the intensity with a small increment until the participants can perceive the stimulus with a "Yes" response. The AL for the ascending series is then taken as the midpoint between the intensities of the last stimulus with a "No" response and the first stimulus with a "Yes" response. A *descending series* begins with a stimulus well above the threshold. The stimulus intensity is gradually decreased with a small step size until the participants cannot perceive the stimulus any more. Then the AL is computed by taking the mid-point between the intensities of the last stimulus with a "Yes" response and the first stimulus with a "No" response. The ascending and descending series should be randomized in the order to have a more reliable estimation of AL. In addition, it is preferred to have an equal number of ascending and descending series (Chapter 2, [41]).

Discrimination Threshold (DL)

The Difference threshold (DL) is derived from a discrimination experiment based on the method of limits, using an equal number of ascending or descending series. During the experiment, two types of stimuli, standard and comparison stimuli, are given. Whereas the intensity level of the standard stimulus is fixed, the intensity of the comparison stimulus changes by a small amount, either toward or away from that of the standard stimulus per trial. The participant's task is to respond whether the comparison stimulus is "greater","equal", or "less" than the standard stimulus. In either ascending or descending series, there are two transition points denoted as the upper limen (L_u) and the lower limen (L_l). The L_u is defined as the point where the participants' response changes between "greater" and "equal" while the L_l denotes the transition between "less" and "equal". For ascending series, the intensity of the comparison stimulus starts below that of the standard stimulus, leading participants to a "less" response. Once the intensity of the comparison stimulus is perceived to be the same as that of the standard stimulus, the response changes from "less" to "equal", the lower limen. The intensity of the comparison stimulus is gradually increased until the point where the response turns into "greater" from "equal" which is the upper limen. For descending series, the response shifts from "greater" to "equal", and then to "less". By running an equal number of ascending and descending series, the difference threshold is estimated to be the mid-point between the average upper limens and the average lower limens (Chapter 2, [41])

2.3.3 Method of Adjustment

Method of Adjustment can be used to estimate both difference threshold and absolute threshold where participants can adjust the intensity of a stimulus. When it is used in discrimination experiment, participants are told to adjust a comparison stimulus till the intensity of the stimulus becomes equal to that of a standard stimulus. This is important when using a device that uses several actuators that should generate equal intensity at a given stimulus.

Detection Threshold (AL)

Absolute threshold in a detection experiment is measured by letting participants gradually increase the stimulus intensity until it is just perceivable or decrease till it is barely detectable, starting from an initial stimulus intensity below or above the threshold. The adjusted intensity of the stimulus is regarded as the absolute threshold [44]. Unlike the method of limits, the participants can adjust the intensity of the stimulus multiple times by themselves until they are satisfied.

Discrimination Threshold (DL)

In a discrimination experiment, the difference threshold is estimated by instructing participants to adjust a comparison stimulus intensity until it is perceived as the same as the standard stimulus. Then, the difference threshold is estimated by calculating the difference between the average of adjusted intensities and the intensity at 84-percentile point [44].

2.3.4 Adaptive Procedures

Adaptive procedure is an efficient way to estimate thresholds by adjusting stimulus intensities based on the participant's responses to prior stimuli. Whereas the method of limits terminates an experiment once the first response reversal occurs, adaptive procedure is composed of multiple reversals. The advantage of using the adaptive procedure is that it provides a more efficient measurement of threshold by placing most of the stimulus levels near the threshold to be estimated. Adaptive procedures contain various methods of measuring thresholds: simple up-down method and transformed up-down procedure. The *simple up-down* method, also known as the staircase method, often starts with a high initial stimulus intensity in a detection experiment. The stimulus intensity decreases if the participant responded "Yes" and increases if "NO" response is made. In a discrimination experiment, participants are asked to judge which stimulus intensity is higher between two stimuli and responses are scored as "correct" or "incorrect".

Before the experiment, an *initial value* or initial stimulus level is determined. Based on the participant's response, the stimulus level is either increased or decreased by an amount called *step size* [45]. The step size is often computed by using an equation c/n, where c is a constant and n refers to the trial number [46]. Beginning with the initial stimulus level, the direction of the ascending or descending series reverses if the current response is opposite to the previous response. After 6 to 8 reversals the experiment is terminated. Computing the average of peak-valley pairs at reversal location, the 50 percentile point (X_{50}) of the psychometric function can be estimated that is regarded as the threshold. This measured threshold value at X_{50} can be treated as the absolute threshold for a detection experiment or the Point of Subjective Equality (PSE) in a discrimination experiment.

While the simple up-down method cannot estimate other levels besides the 50-percentile point, the *transformed up-down method* allows estimation of other points on the psychometric function. For example, the one-up, two-down method measures the threshold at the point $X_{70.7}$ on the psychometric function [31]. In this method, the stimulus intensity decreases if two "Yes" responses are made in a row in a detection experiment. In a discrimination experiment, the comparison stimulus intensity decreases if two "correct" responses are made. Table 2.2 illustrates the rules for a simple up-down method and a one-up two-down method. For each method, the table describes how stimulus level should change based on response sequences. For the simple up-down method, the stimulus intensity increases (hence "UP") following a "No" response (represented as a minus sign) in a detection experiment and "incorrect" response in a discrimination experiment. On the other hand, the "DOWN group" occurs if the response is "Yes" or "correct". In a one-up two-down method, the DOWN group occurs only when two successive "+" responses occur. Based on the probability of a sequence from DOWN group, the threshold could be derived as explained above. Figure 2.9 further describes a sequence of trials for a one-up two-down procedure. In both Table 2.2 and Figure 2.9, the plus sign '+' denotes a "Yes" response in the detection experiment or a 'correct' response in the discrimination experiment whereas the minus sign '−' stands for a "NO" response and 'incorrect' response for detection and discrimination, respectively. The term *peaks* refer to the local maxima of stimulus intensity levels whereas *valleys* represent the local minima. Each *run* refers to the trials between an adjacent peak and valley pair. In the simple staircase method, the 50 percentile point is measured by calculating the average of stimulus intensities at peaks and valleys. The $X_{70.7}$ can be estimated at the 70.7-percentile point on a psychometric function.

Overall, 4 types of measurements for estimating AL and DL were studied. In this this thesis, we used the transformed up-down method (one-up two-down) to measure abso-

Table 2.2.: Rules of increasing or decreasing stimulus levels for a simple up-down (1U1D) procedure and a transformed up-down (1U2D) procedure (modified from Table I in [45]).

	Response Sequences		Probability of positive response at convergence
Type of method	UP group increase level after:	DOWN group decrease level after:	
Simple up-down method (One-up one-down	−	+	P(X)=0.5 (=50%)
Transformed up-down method (One-up two-down)	+− or −	+ +	P(X)=0.707 (=70.7%)

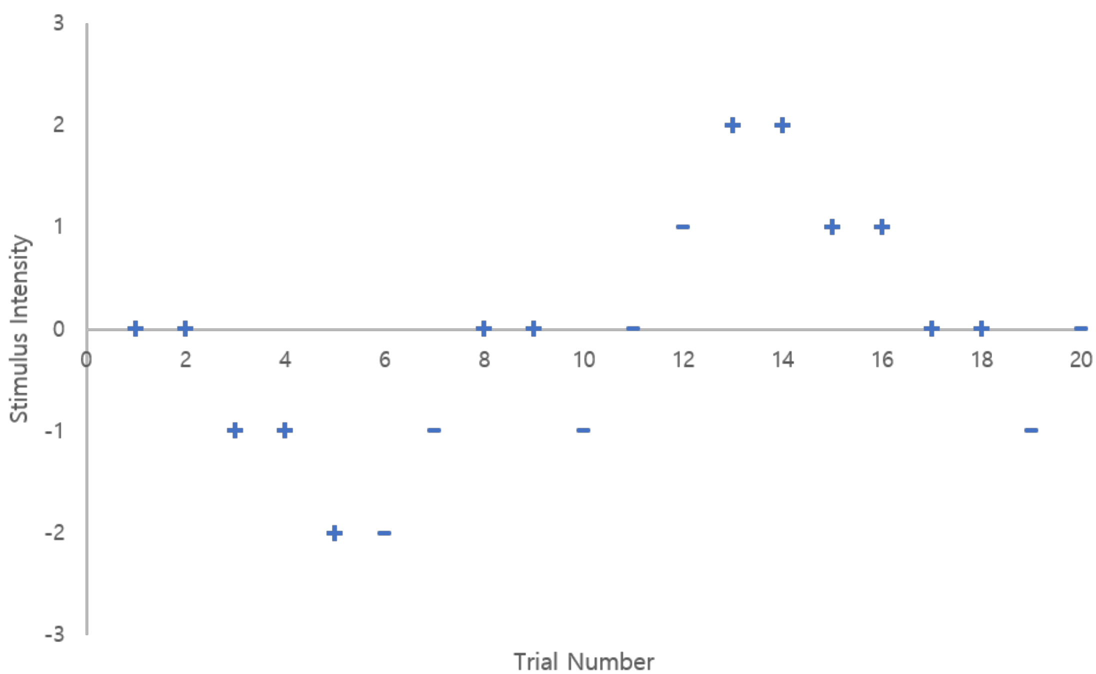


Fig. 2.9.: Hypothetical data of a transformed up-down method using the one-up two-down procedure (modified from Figure 5 in [45]).

lute threshold of participants and method of adjustment to equalize the perceived intensity among 24 tactors used in the TAPS system.

2.4 Tactile Speech Communication

Speech communication can be accomplished through visual, auditory, or tactile perception. Tactile speech communication methods allow people with either hearing or visual impairments to converse with others. However, it has been a challenge to develop tactile speech communication devices, and many past and ongoing research efforts have attempted to address this issue.

One of the natural tactile speech communication methods used by people who are both deaf and blind is called the *Tadoma* method that monitors the articulatory actions of a talking face. The person who is listening places a hand on the face of the speaker, and perceives speech information through cues such as mouth opening, muscle contraction, or any vibration involved. It demonstrates that speech communication through the tactile sense is attainable and research has estimated its information transmission (IT) rate to be up to 12 bits/s [47]. Furthermore, the two-way communication rate ranged between 60 words per minute (wpm) and 80 wpm [48]. Other natural tactile speech communication methods include *fingerspelling* where the individual English alphabet is encoded into gestures of fingers and the sender spells out a word by showing a sequence of finger gestures. Based on this approach, it was found that proficient users could spell out words at a delivery rate of 40 to 45 wpm [49], or at an IT rate of 7.5 bits/s which is lower than that of the Tadoma method [50].

Efforts to develop tactile devices to enable people with severe hearing impairments to receive speech communication through the skin have largely focused on encoding the spectral properties of speech. Such devices transmit the characteristics of acoustic signals to actuators that deliver tactile cues to the skin. An example is a 7-channel tactile aid called *Tactaid VII* [1]. Tactaid VII normally worn around the waist or on the forearm uses seven channels to present spectral properties of speech. The device indicates the first and second formants of a sound, and it delivers the first formant to channels from 1 to 4 and the second formant to channels from 4 to 7 [1]. Using a hand-based configuration, an experiment on phoneme discrimination was conducted by one participant though tactile aid alone where the 5 tactors were worn on each of 5 fingers of a hand and the remaining 2 tactors were worn on the palm of the hand. The experiment was based on an ABX paradigm where the participant received a pair of words (stimulus A and B) followed by either A or B (stimulus X). The participant's task was to identify whether stimulus X was A or B. Within the pair of words, only one phoneme was different between the words. The percent-correct scores for consonant and vowel identification using Tactaid VII alone were 50% and 93%, respectively [9].

A new tactile aid that uses phonemic-based coding scheme (referred to as TAPS) has shown promising results in several recent studies [6], [7], [12], [13]. The TAPS system uses 24 tactors in a a 6×4 array worn on the forearm to convey tactile stimuli. The TAPS system encodes the 39 English phonemes into 39 haptic symbols with different parameters such as durations, frequencies, location of stimuli, and patterns of movement [6]. An initial study of the TAPS system presented the feasibility of the haptic symbols and developed a learning protocol based on memory consolidation [12]. Jiao et al. (2018) demonstrated that 8 participants were able to learn 39 haptic symbols and 100 words within short amount of time with an average percent-correct (PC) score of 86.6% [13]. Furthermore, Tan et al.(2020) [7] used 500 words which consisted primarily of two or three phonemes. Fewer than 10% of the 500 words were composed of four and five phonemes. The average word PC score of 11 experienced participants was 80.7% [7]. These results demonstrated the feasibility of delivering English words to the skin within a reasonable learning period compared to other studies using a spectral-based or letter-based approach. The TAPS system also demonstrated its efficacy compared to Multi-sensory haptic device (MISSIVE) [5] that also used phonemic-based approach. The word PC score of these participants for 50 word identification task based on a subset of 23 phonemes after 100 minutes of training was 86%. Compared to this result, Jiao et al. (2018) [13] showed a similar word PC score with a larger word list (100 words) that demonstrated the efficacy of delivering English words within a reasonable amount of time.

Overall, the TAPS system seems to be highly effective in transmitting English words within a short amount of time. However, most of words tested on previous studies (e.g.,

[13], [7]) were relatively short by containing 2 or 3 phonemes per word. Hence, one of the goals of this thesis was to test and analyze the learning of longer, namely 4-phoneme, words (Chapter 4). Then, this thesis tested the feasibility of the TAPS systems for two-way communication (Chapter 5). In the next chapter (Chapter 3), a previously-published paper is introduced that developed a learning protocol for the TAPS system.

3. SPEECH COMMUNICATION THROUGH THE SKIN: DESIGN OF LEARNING PROTOCOLS AND INITIAL FINDINGS [1]

3.1 Introduction

For a very long time, individuals with hearing impairments have utilized various visual and touch-based methods for speech communication such as *sign language* and *finger-spelling*. Individuals who are both deaf and blind, however, use a natural (i.e., not assisted by man-made devices) method called *Tadoma* that relies solely on the sense of touch. In this method, the listener places the hand on the speakers face, and in the absence of any visual or auditory information, feels the articulatory processes (e.g., mouth opening, air flow, muscle tension and presence/absence of laryngeal vibration) associated with speech production (Figure 3.1). The Tadoma method is a living proof that speech communication through the skin alone is entirely achievable, as established by past research on Tadoma and other natural speech communication methods [47]. Inspired by the Tadoma method, electromechanical devices have been developed to study and replicate the cues produced by a talking face that are crucial for speech reception by Tadoma users [51]. Experimental results on one such device, the Tactutor, demonstrate an information transmission rate of 12 bits/s, the same rate that has been established for speech communication by Tadoma users [47], [52]. Despite such promising results in the laboratory, however, there are still no commercially-available communication devices that can be used by people who are deaf or deaf-and-blind for speech reception at a level that is comparable to that shown by Tadoma users. The challenges are manifold. First, a talking face is a rich display and require actuators that can deliver a wide range of sensations including movement, vibration, airflow, etc.

[1] THIS CHAPTER IS A REPRINT OF JAEHONG JUNG, YANG JIAO, FREDERICO M. SEVERGNINI, HONG Z. TAN, CHARLOTTE M. REED, ALI ISRAR, FRANCES LAU, AND FREDDY ABNOUSI, "SPEECH COMMUNICATION THROUGH THE SKIN: DESIGN OF LEARNING PROTOCOLS AND INITIAL FINDINGS," *PROCEEDINGS OF HCI INTERNATIONAL 2018*, LAS VEGAS, NE, P. 447-460, JULY 15-20, 2018.

Second, such communication devices should be wearable, or at least portable. Third, the input to such devices should ideally be processed, as opposed to raw, acoustic signals for consistent conversion of sound to touch. Fourth, the mapping between speech sounds and haptic symbols needs to be designed such that it is easy to learn and retain. Last but not the least, a significant effort is expected of any individual who wishes to learn to use such a communication device.

Fig. 3.1.: The Tadoma method of speech communication. Shown are two individuals who are both deaf and blind (on the left and right, respectively) conversing with a researcher (center) who has normal vision and hearing (Photo courtesy of Hansi Durlach).

With the recent development in haptic display and speech recognition technologies, now is the time to give it another try to develop a wearable system for speech communication through the skin. In our research, we use an array of wide-bandwidth actuators for presenting rich haptic signals to the skin on the forearm. The display portion of our system is wearable but still tethered to equipment that has yet to be miniaturized. We assume that speech recognition technologies are available to extract phonemes from oral speech in real time, and therefore use phonemes as the input to our system. We have designed and tested a set of distinct haptic symbols representing phonemes of the English language. The present work focuses on the exploration of training protocols that facilitate the learning of phonemes and words in hours as opposed to days or weeks. Our initial results with six participants have been very promising. In the rest of this paper, we present the background for our approach followed by methods and results from one pilot study and two experiments. We conclude the paper with guidelines for effective learning of speech communication through the skin via man-made systems.

3.2 Related work

There is a long history of research on the development of synthetic tactile devices as speech-communication aids for persons with profound hearing loss (e.g., see reviews [53], [54], [55]) that continues to the present day [56], [57], [58]. From a signal-processing point of view, many devices have attempted to display spectral properties of speech to the skin. These displays rely on the principle of frequency-to-place transformation, where location of stimulation corresponds to a given frequency region of the signal. Another approach to signal processing has been the extraction of speech features (such as voice fundamental frequency and vowel formants) from the acoustic signal prior to encoding on the skin. For both classes of aids, devices have included variations in properties such as number of channels, geometry of the display, body site, transducer properties, and type of stimulation (e.g., vibrotactile versus electrotactile).

A major challenge in the development of tactile aids lies in encoding the processed speech signals to match the perceptual properties of the skin. Compared to the sense of hearing, the tactual sensory system has a reduced frequency bandwidth (20-20,000 Hz for hearing compared to 0-1000 Hz for touch), reduced dynamic range (115 dB for hearing compared to 55 dB for touch), and reduced sensitivity for temporal, intensive, and frequency discrimination (see [23]). The tactual sense also lags behind the auditory sense in terms of its capacity for information transfer (IT) and IT rates [52]. For example, communication rates of up to 50 words/min are achieved by experienced operators of Morse code through the usual auditory route of transmission, compared to 25 words/min for vibrotactile reception of these patterns [50]. Taking these properties of the tactual sense into account, certain principles may be applied to create displays with high IT rate. One such principle is to include as many dimensions as possible in the display, while limiting the number of variables along each dimension.

Another challenge is present in the need to provide users with adequate training in the introduction of novel tactile displays. Compared to the structured training and many years of learning associated with the Tadoma method, most tactile aids have been evaluated within the context of relatively limited exposure in a laboratory setting. Recent advances arising out of the literature in language-learning [59] and memory consolidation [60] offer insight into improved approaches for training that may be applied to the use of a novel tactile speech display. Results from a multi-modal, game-based approach to language learning have shown that observers are able to learn to categorize auditory speech sounds when they are associated with the visual stimuli needed to perform the task. In addition, studies of learning suggest that following exposure to a new task, the initial memories associated with this task may be further consolidated by activations of these memories during periods of wakefulness and sleep. Thus, learning can occur between the laboratory sessions with no explicit training involved.

Based on the literature, we explored several strategies of facilitating learning and training such as breaking up training time into smaller intervals, sounding a phoneme/word out while feeling the corresponding haptic symbols, and keeping the task doable but challenging at all times.

3.3 General Methods

3.3.1 Participants

A total of six participants (2 females; age range 20-30 years old) took part in the present study. All were right handed with no known sensory or motor impairments. The participants came from diverse language backgrounds. While all participants spoke English fluently, only one was a native English speaker. Other languages spoken among the participants included Korean, Chinese, Tibetan, Hindi and German. Most of the participants also received early childhood music training including piano, clarinet, violin and percussion.

3.3.2 Apparatus

The experimental apparatus consisted of a 4-by-6 tactor array worn on the non-dominant forearm. The 24 tactors form four rows in the longitudinal direction (elbow to wrist) and six columns (rings) in the transversal direction (around the forearm). As shown in figure 3.2 below, two rows (i and ii) reside on the dorsal side of the forearm and the other two (iii and iv) on the volar side. The tactor positions were adjusted so that the rows form straight lines and the columns are evenly distributed from the elbow to the wrist. The tactors were attached to a sleeve via adjustable Velcro strips. The sleeve was then wrapped around the forearm with a snug fit to ensure good contact between the tactors and the skin.

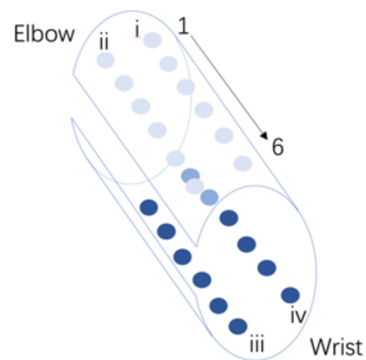


Fig. 3.2.: Illustrations of tactor layout in the experimental apparatus.

A wide-bandwidth tactor (Tectonic Elements, Model TEAX13C02-8/RH, Part #297-214, sourced from Parts Express International, Inc.) was used as the actuator. It has a flat frequency response in the range 50 Hz to 2 kHz with a resonant peak close to 600 Hz. A MOTU audio device (MOTU, model 24Ao, Cambridge, MA, USA) was used for delivering 24 channels of audio waveforms to the 24 tactors through custom-built stereo audio amplifiers. A Matlab program running on a desktop computer generated the multi-channel waveforms corresponding to the haptic symbols for phonemes, presented a graphic user interface for running the experiments, and collected responses from the participants.

With this setup, the tactors can be driven independently with programmable waveforms and on-off timing. The stimulus properties include amplitude (specified in dB sensation level, or dB above individually-measured detection thresholds), frequency (single or multiple sinusoidal components), waveform (sinusoids with or without modulation), duration, location, numerosity (single tactor activation or multiple tactors turned on simultaneously or sequentially), and movement (smooth apparent motion or discrete salutatory motion varying in direction, spatial extent, and trajectory).

The participants sat comfortably in front of a computer monitor. They wore a noise-reduction earphone to block any auditory cues emanating from the tactors. The participants placed their non-dominant forearm on the table with the volar side facing down. The elbow-to-wrist direction was adjusted to be parallel to the participants torso. The participants used their dominant hand to operate the computer keyboard and mouse (Figure 3.3).

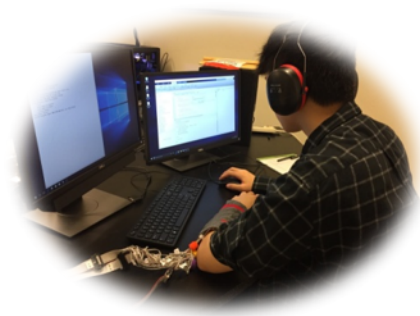


Fig. 3.3.: Experimental setup.

3.3.3 Phoneme Codes

English words are pronounced by a sequence of sounds called phonemes [61]. Table 3.1 shows the IPA (international phonetic alphabet) symbols of the 39 English phonemes used in the present study and example words that contain the corresponding phonemes. The list consists of 24 consonants and 15 vowels.

Vibrotactile patterns using one or more of the 4-by-6 tactors were created, one for each phoneme. The mapping of the phonemes to haptic symbols incorporated the articulatory features of the sounds, balanced by the need to maintain the distinctiveness of the 39 haptic symbols. For example, place of articulation was mapped to the longitudinal direction so that the wrist corresponds to the front of the mouth and the elbow the back of the mouth.

Table 3.1.: The thirty-nine (39) English phonemes used in the present study.

IPA symbol	Example word	IPA symbol	Example word	IPA symbol	Example word	IPA symbol	Example word
/i/	meet	/U/	hood	/k/	key	/ʒ/	azure
/ei/	mate	/ʌ/	hut	/p/	pay	/tʃ/	chew
/u/	mood	/OU/	boat	/t/	tea	/dʒ/	jeep
/aI/	might	/ɔI/	boy	/b/	bee	/h/	he
/aɐ/	mat	/aU/	pouch	/g/	guy	/r/	ray
/ɑ/	father	/d/	do	/f/	fee	/l/	lie
/ɔ/	bought	/m/	me	/ʃ/	she	/j/	you
/ɛ/	met	/s/	see	/θ/	think	/n/	new
/ɚ//ɝ/	bird	/w/	we	/v/	voice	/ŋ/	sing
/I/	bid	/ð/	the	/z/	zoo		

Therefore, the consonant /p/ was mapped to a 100-ms 300-Hz pulse delivered near the wrist (front of the mouth) whereas the consonant /k/ was mapped to the same waveform delivered near the elbow (back of the mouth). Their voiced counterparts, /b/ and /g/, were mapped to the 100-ms 300-Hz pulse modulated by a 30-Hz envelope signal delivered near the wrist and elbow, respectively. The modulation resulted in a rough sensation that signified voicing. Details of the phoneme mapping strategies and the resultant haptic symbols can be found in [6].

3.3.4 Intensity Calibration

In order to control the perceived intensity of vibrotactile signals at different frequencies and different locations on the forearm, signal amplitudes were calibrated in two steps in Exp. I. First, individual detection thresholds were taken at 25, 60 and 300 Hz for the tactor on the dorsal side of the forearm near the center (row i, column 4 in 3.2). A one-up two-down adaptive procedure was used and the resulting detection threshold corresponds to the 70.7 percentile point on the psychometric function [45]. Signal amplitudes were then defined in sensation level (SL); i.e., dB above the detection threshold at the same frequency.

In the present study, signal amplitudes were set to 30 dB SL for a clear and moderately-strong intensity.

Second, the perceived intensity of the 24 tactors was equalized using a method of adjustment procedure. A 300-Hz sinusoidal signal at 30 dB SL was sent to the tactor used in the detection threshold measurements (see the black tactor in figure 3.4). The participant selected one of the remaining 23 tactors, say the upper-left tactor in figure 3.4, and adjusted its vibration amplitude until the vibration felt as strong as that of the black tactor. This was repeated for all the tactors. The equalization results for one participant are shown in figure 3.4. The numbers below each tactor indicate the signal amplitudes in dB relative to the maximum amplitude allowed in the Matlab program for a 300-Hz vibration at 30 dB SL. For example, this participants detection threshold at 300 Hz was -54 dB relative to the maximum allowable amplitude. The amplitude for the black reference tactor was therefore at -24 dB for a 30 dB SL signal. The number -23 below the tactor in the upper-left corner indicates that its amplitude needed to be 1 dB higher than that of the black tactor to match its strength for a signal at 30 dB SL. In other words, the skin near the elbow under this tactor was slightly less sensitive than the skin under the black tactor. Generally speaking, the skin on the dorsal side was more sensitive than that on the volar side and the wrist area was more sensitive than the elbow area.

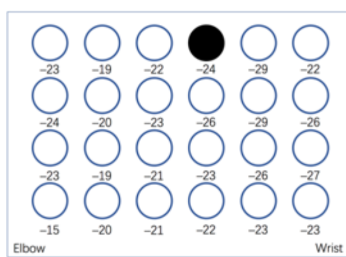


Fig. 3.4.: Tactor intensity equalization.

3.3.5 Data Analysis

The experiments reported in this paper consisted of learning and testing of phonemes through individual haptic symbols and words through sequences of haptic symbols. Test results were organized as stimulus-response confusion matrices where each cell entry is the number of times a haptic symbol was recognized as a phoneme label. Table 3.2 below shows an example of a confusion matrix for a 6-phoneme stimulus set. As was typical of most tests, a majority of the trials fall on the main diagonal cells (i.e., correct answers). Therefore, the results could be well captured by the percent-correct scores (48/50 = 96%) and the error trials /i/→/ei/and /i/→/u/. Therefore, in the rest of the paper, we report the percent-correct scores and error trials for each test.

Table 3.2.: An example confusion matrix for a 6-phoneme recognition test. Each cell represents the number of times a haptic symbol is recognized with a phoneme label.

		Phoneme labels (responses)					
		/m/	/d/	/s/	/ei/	/u/	/i/
Haptic symbols (stimuli)	/m/	9	0	0	0	0	0
	/d/	0	5	0	0	0	0
	/s/	0	0	7	0	0	0
	/ei/	0	0	0	5	0	0
	/u/	0	0	0	0	13	0
	/i/	0	0	0	1	1	9

3.4 Pilot Study: Learning of 6 Phonemes and 24 words

The purpose of the pilot study was to gain initial experience and insight into the learning process. One participant (P1) practiced and learned phonemes and words over a period of 21 days and took detailed notes. Within the 21 days, 4 days fell on a weekend, there was a break of 3 days after the 5th learning day, and a break of 2 days after the 11th learning day. Thus, there were a total of 12 experimental sessions.

Learning Materials

The materials included 6 phonemes and 24 words made up of the phonemes. The six phonemes were /i/, /ei/, /u/, /d/, /m/, and /s/. The words consisted of 10 CV (consonant-vowel) words (e.g., may, see) and 14 CVC (consonant-vowel-consonant) words (e.g., moose, dude).

Time Constraints

The learning and testing was self-paced. The participant took a break whenever needed.

Procedure

The following list shows the tasks performed by P1 over the 12 learning days. For each task, he practiced first with an individual phoneme (or word), then with a random list of phonemes (or words), followed by a recognition test. The numbers in parentheses indicate the total time spent on each learning day.

- Day 1-3 (20, 10, 15 min): 6 phonemes;
- Day 4-7 (5, 5, 5, 17 min): 10 CV words;
- Day 8-11 (5, 24, 5, 30 min): 14 CVC words;
- Day 12 (24 min): test with all 24 words.

Results

Participant P1 achieved a performance level of 200/200 (correct-trials/ total-trials) with the 6 phonemes on Day 3, 198/200 with the 10 CV words on Day 7, 200/200 with 7 of the 14 CVC words on Day 9, 200/200 with the remaining 7 CVC words on Day 11, and 198/200 with all 24 words on Day 12.

Insight Gained

The results of the pilot study indicate clearly that participant P1 was able to learn the 6 phonemes and 24 words almost perfectly after 165 min. He intuitively progressed from easier to harder tasks, each time learning and testing himself before moving onto more difficult tasks. Since the task was highly demanding, it was challenging for P1 to maintain a high level of concentration after about 20 min. Therefore, it was necessary and more productive to spread the learning and testing over many days instead of spending a long time continuously. Furthermore, P1 found that his performance did not deteriorate after a 3-day gap between Day 5 and Day 6.

Encouraged and informed by the results of the pilot study, two experiments were conducted. Experiment I tested four new naïve participants with 10 phonemes and 51 words. Experiment II tested one more naïve participant with the full set of 39 phonemes and tested the memory consolidation theory explicitly.

3.5 Experiment I: Learning of 10 Phonemes and 51 Words

3.5.1 Methods

Four new naïve participants (P2-P5) took part in Exp. I. Each participant spent a total of 60 min to learn 10 phonemes and 50 words made up of the 10 phonemes.

Learning Materials

The ten phonemes included the six used in the pilot study and four more: /w/, /ð/, /k/, /aI/. The words consisted of the 24 words used in the pilot study plus 27 additional words (13 CV and 14 CVC words).

Time Constraints

The learning time was capped at 10 min on each day, with no break. The design ensured that each participant could maintain full concentration during the time spent learning the phonemes and words, and took advantage of memory consolidation by spreading the one-hour learning period over multiple days.

Procedure

The following list shows the tasks performed by each participant over the six learning days. On each day, the participant practiced for 5 min, followed by a test with trial-by-trial correct-answer feedback for another 5 min.

- Day 1 (10 min): 6 phonemes;
- Day 2 (10 min): 24 words made up of the 6 phonemes;
- Day 3 (10 min): 4 new phonemes learned, all 10 phonemes tested;
- Day 4 (10 min): 27 new words;
- Day 5-6 (10, 10 min): all 51 words.

3.5.2 Results

The results in terms of percent-correct scores are shown in Figure 3.5. Due to the fact that the participants reached near-perfect performance level on Day 3 (10 phonemes) and Day 6 (51 words), we do not report the few error trials. The results are organized by day (and cumulative training time in min). Data for the four participants are shown in different color patterns.

Several observations can be made. First, it was relatively easy for the participants to learn the phonemes. Performance was near perfect on Day 1 (6 phonemes) after 5 min of learning and on Day 3 (10 phonemes) after 5 min of learning 4 new phonemes. Second, the transition from phoneme to words took some getting used to, as seen comparing the results from Day 1 (6 phonemes) and Day 2 (24 words made up of the 6 phonemes). This indicates that additional learning was required in order to process phonemes delivered in a sequence. Third, despite the initial dip in performance on Day 2 and Day 4 when the participants transitioned from phonemes to words, word-recognition improved quickly as seen in the rising performance from Day 4 to Day 6. The most significant improvement occurred with P5 who reached 62.5%, 77.5% and 97.5% correct from Day 4 to Day 6, respectively. Finally, regardless of individual differences among the four participants in earlier days, all participants succeeded in identifying the 51 words with very few errors by the end of the 60-min period.

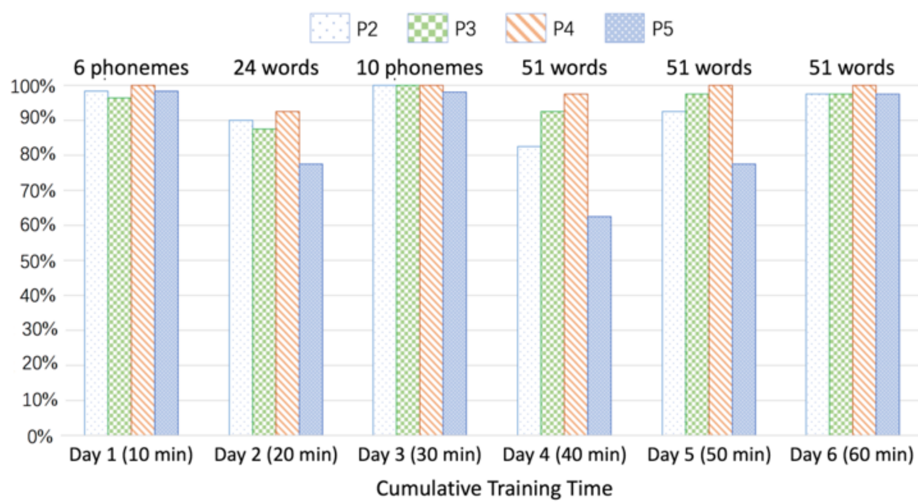


Fig. 3.5.: Results of Exp. I (10 phonemes and 51 words).

Compared with the pilot study, participants in Exp. I learned more phonemes and words in less time. This is probably due to the strict control of learning time per day in order to maintain a high level of concentration on the part of the participants. In addition, the mapping from phoneme to haptic symbols was improved based on the feedback from P1 in the pilot study. The new haptic symbols were more distinct and easier to learn than those in the pilot study. The continued improvement from Day 4 to Day 6 for all participants, especially P5, led us to speculate that memory consolidation may have played a big role in the learning process. We therefore designed Exp. II to explicitly test the effect, if any, of memory consolidation. We also used Exp. II to test whether all 39 phonemes could be learned and how long the learning process would take.

3.6 Experiment II: Learning of 39 Phonemes

3.6.1 Methods

The objectives of Exp. II were to (1) test memory consolidation explicitly, (2) gain experience and insight into learning all 39 phonemes, and (3) record the time it takes to learn the phonemes and the attainable performance level. One new naïve participant (P6) took part in Exp. II for a total of 14 consecutive days, including the weekends.

Learning Materials

All 39 phonemes shown in Table 1 are included in Exp. II. The 39 phonemes were divided into 8 groups. In addition to the first two groups (containing 6 and 4 phonemes, respectively) that were used in Exp. I, an additional 6 groups of phonemes were created with 5 new phonemes per group except for the last group that contained 4 new phonemes.

Time Constraints

As in Exp. I, the learning time was capped at 10 min on each day, with no break. The participant took detailed daily notes on his observations afterwards.

Procedure

For testing the memory consolidation theory, the participant always ended a day and began the next day with the same test, except for Day 1 when there was no test at the beginning of the day. The participant then spent 3-4 min on learning new phonemes and the rest of the 10 min on testing all phonemes learned so far, with trial-by-trial correct-answer feedback. In addition, the participant sounded out a phoneme during the learning phase. The learning plan is shown below, with the total number of phonemes learned/tested each day clearly marked (Figure 3.6). The participant had to achieve a percent-correct score of 90% or higher before he could move on to the next group of new phonemes. As shown below, the participant was able to learn one group of phonemes per day during the first 8 days, and was tested with all 39 phonemes from Day 9 to 14.

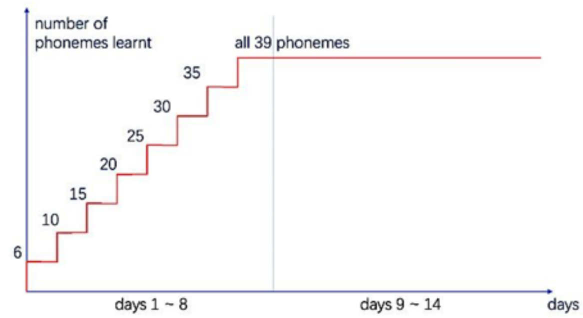


Fig. 3.6.: Learning plan for Experiment II.

3.6.2 Results

The results are presented in two parts. We first show the percent-correct scores for phoneme recognition from Day 1 to Day 8, including the test conducted on Day 9 that repeated the last test on Day 8 (Figure 3.7). It is clear that when the same test was conducted again the next day, the performance level either remained the same or improved. This provides a clear evidence for the memory consolidation theory in the sense that performance improved (with the exception of Day 6 to Day 7) after a period of activities not related to phoneme learning. A second observation is that the participant had no difficulty learning between 4 to 6 new phonemes a day, presumably due to the highly distinctive haptic symbols and the easy-to-learn mapping from phoneme to haptic symbols.

Starting on Day 9, the participant was tested with 4 runs of 50 trials on all 39 phonemes for six consecutive days. The daily percent-correct scores from the 200 pooled trials are shown in Figure 3.8. Overall, P6 was able to maintain a high performance level (93.8% ± 3.8% correct).

A stimulus-response confusion matrix was constructed from all 1200 trials (50 trials/run $\times$ 4 runs/day $\times$ 6 days) to examine the most-confused phoneme pairs. The following is the list, in order of descending number of confusions:

- /t/with /k/(9 times);
- /ae/with /l/(8 times);
- /b/with /d/(4 times);
- /g/with /d/(4 times);
- /i/with /z/(4 times);
- /t/with /p/(4 times);
- /u/with /ai/(3 times);
- /g/with /m/(3 times);
- /n/with /h/(3 times).

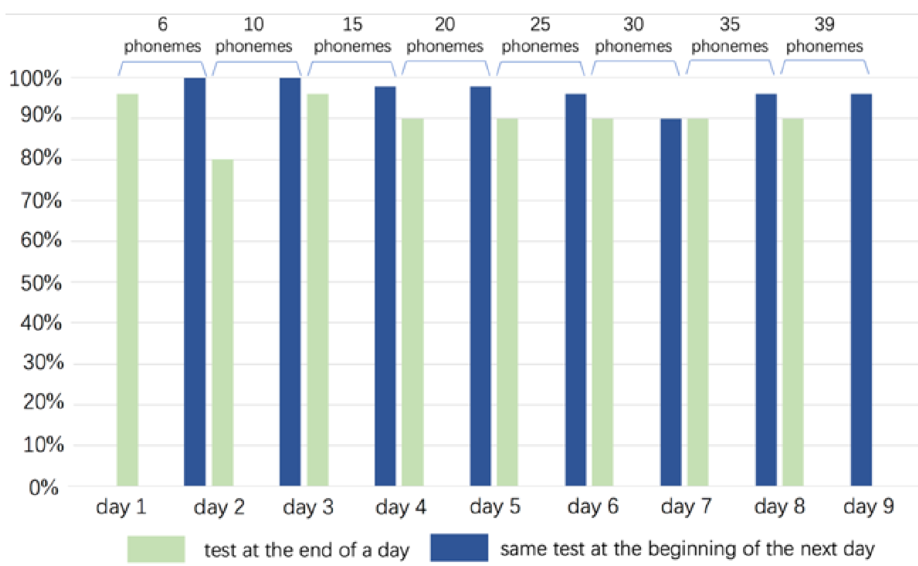


Fig. 3.7.: Phoneme recognition performance from Day 1 to Day 9.

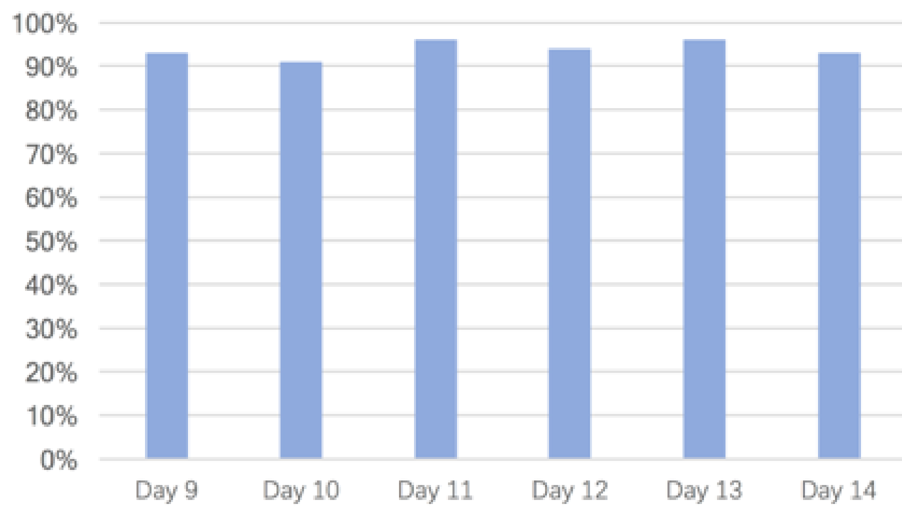


Fig. 3.8.: Daily phoneme recognition scores from Day 9 to Day 14.

The rest of the errors occurred twice or less and are not listed. The confusion pattern served to guide further refinement of the design of haptic symbols.

3.7 Concluding Remarks

The present study offers evidence to support the claim that speech communication through the sense of touch is an achievable goal. The participants received speech information (phonemes and words) presented on their forearm through a sequence of haptic symbols encoding phonemes. The results of Exp. I show that four nave participants were able to recognize 51 words in 60 min. The results of Exp. II show that all 39 English phonemes could be learned in 80 min. We demonstrated memory consolidation in Exp. II by showing an improvement in phoneme recognition performance when the participant P6 was tested a day after he learned the phonemes.

Several guidelines can be provided based on the experimental results and the insights we have gained from the present study. First, the learning time should be limited to 10 to 20 min per session. It was difficult for participants to maintain full concentration after 20 min. Second, it might be helpful for the participant to sound a phoneme out as the haptic symbol was delivered to the forearm although we did not collect data to prove that. Third, the task difficulty should be carefully managed so that the participant is challenged but able to make progress. Last but surely not the least, the results of Experiment II provide evidence for learning that occurred between laboratory sessions when the participant was not being trained. Therefore, we recommend spending a short period of time (e.g., 10 min) per day and taking advantage of memory consolidation for further improvement of learning outcome. Furthermore, the benefit of memory consolidation did not appear to be impacted when learning sessions were interrupted by one or more days.

In the future, we will continue to improve the haptic symbols to further reduce errors in phoneme recognition. We will also compare different learning strategies such as reversing the order in which phonemes and words are introduced to the participants. Our ultimate goal is to develop a haptic speech communication system for people with all levels of sensory capabilities.

Acknowledgments

This work was partially supported by a research grant from Facebook Inc. The authors thank all participants for their time and dedication, and Keith Klumb for project management.

4. ACQUISITION OF FOUR-PHONEME WORDS THROUGH THE SKIN

4.1 Introduction

In this study, we concentrate on the performance of learning longer words using TAPS. Our long-term goal is the use of TAPS as a practical means of communication in the daily life of a user. To achieve this goal, an understanding of how learning progresses from short to long words, phrases, sentences, and eventually passages of any length is necessary.

Bryan and Harter (1899) studied the acquisition of telegraphic language (i.e., Morse code), and observed that people go through several stages in learning Morse code language [62]. They closely monitored students of telegraphy over months and years on their learning of Morse codes. They found that the students experienced several periods of improvement and plateaus in the learning curve of receiving a discourse. Initially, the students were forced to attend to letters and words The students showed an improvement in receiving discourse as they learned to receive letters and words more proficiently. However, they reached a plateau in the middle of the learning curve for receiving sentences before resuming improvements in their performance again. The plateau was regarded as evidence that it was necessary to learn and master letters to learn words, which are combinations of letters, and before learning to receive phrases and sentences [62], [50]. Hence, the purpose of the present study was to obtain learning curves for longer words using TAPS, to expand upon past studies of TAPS that used shorter words.

During this experiment, a total of three experienced participants received training on the acquisition of 4-phoneme words. Past studies focused more on 2- or 3-phoneme words [6], [13], [7]. In the beginning of the experiment, the participants reviewed the 39 phonemes and a list of 500 words that they had learned in an earlier study [7] that contained from 1- to 5-phoneme words with most words consisting of 2 and 3 phonemes. Then, they learned one hundred 4-phoneme words over 10 days. Finally, the last two days were spent to assess the generalization of learning 4-phoneme words using a new list of one-hundred words.

4.2 Methods

This section describes the experimental methods that were unique to the present study. The same *TActile Phonemic Sleeve* (TAPS) described in Section 3.3.2 was used in this study. The intensity calibration and tactor equalization procedures were the same as described in Section 3.3.4.

4.2.1 Participants

A total of 3 young adults (P1, P2 and P3) participated in this study. Each participant was right handed without any tactual sensory impairments. Their age ranged from 23 to 26 with an average of 24 ± 1.73 years. There were two male participants and one female participant. P1 is a native Korean speaker who started learning English at the age of 8 years old. He also speaks German as a second language. P2 is a native Spanish speaker who learned English when he was eight. Both participants, P1 and P3, speak English fluently. P3, on the other hand, is a native English speaker who also speaks Chinese as a second language. The participants had experience with musical instruments such as violin, piano, and clarinet.

All the participants had used TAPS earlier in the previous studies and were familiar with the haptic codes for the 39 phonemes [6], [13], [7], [12]. They also learned a set of 500 words in [7].

4.2.2 Learning Materials

The participants began with the review of the haptic codes for the 39 phonemes, and then proceeded to the review of the 500-word list. The list composed of CV (consonant vowel), VC, CCV, CVC, VCC, CCVC, CVCC, and other consonant-vowel combinations.

Examples of the words in the 500-word list are shown in Table 4.1 with different consonant and vowel combinations. For example, the word “ace” in Table 4.1 is composed of a leading vowel followed by a consonant.

Table 4.1.: Example words for different combinations of consonants (C) and vowels (V) from the 500-word list used in [7].

Consonant & Vowel Combinations	Example Words
VC	ace
CV	do
CCV	blue
CVC	but
VCC	and
CCVC	bread
CVCC	build
CVCV	daughter
VCVC	above
CVCVC	tonight

The 500-word list that the participants learned in [7] contains words composed of 1 to 5 phonemes. Among the 500 words, 2 words (0.4% of the 500-word list) had only one phoneme, 89 words (17.8%) had two phonemes, 359 words (71.8%) were made up of 3 phonemes, 49 words (9.8%) were composed of 4 phonemes, and one word (0.2%) contained five phonemes. The number and proportion of words consisting of different number of phonemes in the 500-word list is shown in Table 4.2 using the information from [7].

Table 4.2.: Number and percentage of words composed of different number of phonemes in the 500-word list (from [7]).

Number of Phonemes	Number of Words	Percentage (%) of Words
1-phoneme	2	0.4
2-phoneme	89	17.8
3-phoneme	359	71.8
4-phoneme	49	9.8
5-phoneme	1	0.2

Two-hundred 4-phoneme words were compiled from words in CUNY sentences developed for testing reception of conversational speech [63]. The 200 words were randomly assigned to two 100-word lists. The first 100-word list (List #1) as shown in Table 4.3 was used for studying learning over time. The second 100-word list (List #2) presented in Table 4.4 was utilized after training had been completed to test the participants' generalization of learning on 4-phoneme words. That is, the participants were tested with List #2, which they had never perceived previously, without any training on the list. Both word lists (List #1 and List #2) consisted primarily of CVCV, CVCC, and CCVC in terms of vowel (V) and consonant (C) composition. 31% of the words in List#1 and List#2 were CVCC words. In addition, 31% and 22% of the words in both word lists were CCVC and CVCV words, respectively. 6% of the words were VCVC words. Remaining 10% of the words were composed of VCCV(4%), CVVC(3%), VCCC(1%), VCVV(1%), and CCVV(1%) words.

A timing parameter called Inter-Phoneme Interval (IPI) was determined prior to the experiment. The TAPS system converts the words into a series of haptic symbols designed for the 39 phonemes and conveys them to the skin (see [6]). The IPI is defined as the pause between phonemes. In this experiment, an IPI of 150 ms was used for all three participants.

4.2.3 Learning and Testing Procedures

Participants reviewed the haptic codes for the 39 phonemes first. Even though they were highly experienced with the TAPS system, almost a year had passed since they last participated in an experiment involving TAPS (i.e., [7]). It was also decided that the participants could benefit from reviewing the 500 words they learned in the previous study [7], to help them get up to speed. This was followed by the learning and testing with the two 100-word lists that contained only 4-phoneme words. The participants received training on the 39 phonemes and the 500-word list for 1 and 2 days, respectively. Then, they practiced and tested on the first 100-word list (List #1) for a period of 10 days. Finally, the participants spent 2 days being tested with List #2 to assess generalization of learning on the four-phoneme task.

Table 4.3.: The 100 four-phoneme words in List #1.

List #1				
able	clean	first	mirage	steak
after	clothe	flower	moved	stock
apple	color	fourth	movie	store
aunts	cooked	fried	often	summer
baby	cost	from	only	sweet
band	cover	glove	oven	tailor
belt	crack	golf	paper	tank
berry	cream	green	park	teacher
black	cruise	guest	plan	tent
block	diet	hard	puppy	throat
bond	dollar	honey	quit	train
borrow	dress	huge	ranch	trick
break	drop	kids	rest	trip
bring	ducks	last	sand	turkey
brush	easier	later	silk	under
butter	easter	lend	singer	water
card	eaten	lion	skirt	weeks
change	even	lunch	slice	wind
chimp	fever	major	slope	years
city	field	milk	snake	zipper

The entire experiment was conducted over 15 days, with practice limited to 10 to 20 minutes per day to avoid fatigue. Our past research has shown evidence of memory consolidation. Improved approaches for learning language ares suggested by studies demonstrating that an initial memory is encoded by repetitive reactivation during both wakefulness and sleep [60]. It is also demonstrated in [12] that breaking up learning time into small daily chunks and allowing participants to have a period of inactivity regarding learning haptic symbols improved participants' performance. It was therefore more efficient and effective to spread learning over a period of time with concentrated practice time each day.

For each learning material except for List #2, participants finished two tasks for most of the days: **practice** and **test**. For **practice**, the participant engaged in either "free play" or "mock test" activities. During *free play*, the participant could choose any phoneme or word and then feel it on the left forearm. During *mock test*, the computer delivered a randomly-

Table 4.4.: The 100 four-phoneme words in List #2.

List #2				
about	cloth	fresh	neighbor	soccer
alter	coffee	fries	office	sold
asked	coins	fruit	onto	stall
avoid	cookie	garage	open	still
ballet	crash	gift	paint	stop
bank	cross	great	pairs	style
bark	cute	grill	parade	sunny
best	desk	guest	past	swim
better	dinner	heavy	pepper	tax
bleach	doing	horse	place	think
blood	drive	into	press	toast
boots	dryer	july	ready	today
breathe	eating	least	rent	track
bridge	enjoy	letter	risk	treat
broke	enough	lights	shirts	trim
busy	every	liver	shutter	twice
cats	fast	lost	skate	washer
child	fatty	money	sleeve	weather
chilly	find	month	slip	wrist
clear	fix	must	smoke	yard

selected phoneme or word from the learning materials of the day, the participant responded with a phoneme or word, and correct-answer feedback was provided on each trial. At the end of each day's practice, the participant's performance was assessed in a **test** where no feedback was given.

An overview of the tasks performed by the participants on each day is shown below. Detailed descriptions of each day's activities are presented after the list.

- Day 1: Practice with all 39 phonemes (10 min) and test (25 trials/block × 2 blocks). Then, practice with 500-word list (10 min.)

- Day 2: Practice with 500-word list (20 min) and test (25 trials/block × 2 blocks).

- Day 3: Same as Day 2.

- Day 4: Practice with 100-word List #1 (20 min).

- Day 5: Practice with 100-word List #1 (10 min) and test (25 trials/block × 2 blocks).
- Days 6 & 7: Same as Day 5.
- Day 8: Practice with 100-word List 1 (10 min) and test (25 trials/block × 3 blocks)
- Days 9 to 13: Same as Day 8
- Days 14 & 15: Test with 100-word List #2 (25 trials/block × 4 blocks)

On the first day of the experiment, the participants reviewed all 39 phonemes used to form any English words. It was important for the participants to be proficient with phoneme recognition again so that they could proceed with the learning of words composed of a sequence of phonemes. They started with practicing 39 phonemes. Participants were allowed to practice with free play and mock test for 10 minutes in the beginning. Trial-by-trial correct-answer feedback was provided during the mock test, which counted as part of learning. After 10 minutes of practice, they were tested with 39 phonemes with 2 blocks of 25 trials. No feedback was provided during the test. Then, the participants practiced the 500-word review list for 10 min.

On Days 2 and 3, the participants continued with practicing the 500-word review list to familiarize themselves with combining a sequence of phonemes into words. The first 20 minutes were spent on practice in the form of free play and/or mock test. Afterwards, two blocks of 25-trial tests were conducted where no feedback was provided. The participants were expected to become proficient with word recognition after these two days.

On Day 4, the participants practiced with a new 100-word list (see List #1 in Table 4.3). As explained earlier, each of the 100 words contained four phonemes. The participants spent all 20 minutes on learning in the form of free play and/or mock test. No word recognition test was conducted on Day 4.

From Day 5 to Day 7, the participants spent 10 minutes each day with the practice of word List#1 and then performed word recognition without feedback in two blocks of 25 trials each.

Starting on Day 8, the participants again spent 10 minutes each day practicing with word List#1. The word recognition test without feedback increased from two to three blocks of 25 trials. The same procedure continued from Day 9 to Day 13.

The generalization word recognition tests were conducted during the last two days of Day 14 and Day 15. On each day, the participants performed the word recognition test without feedback on a new list of 100 words (see List#2 in Table 4.4). The 4-phoneme words in List #2 contained new words that did not appear in the 500-word review list or List#1. Each participant conducted four blocks of 25 trials on each day. At the conclusion of the 15 days, each participant completed a total of 600 trials with List#1 (24 25-trial runs) and 200 trials with List#2 (8 25-trial runs), of recognizing four-phoneme words without any feedback.

4.2.4 Data Analysis

Data on word recognition tests without feedback were collected from Day 5 to Day 15. Response time (RT) on each trial was logged and the participants' word responses were recorded. The RT was defined as the duration between the end of a stimulus presentation and the start of the key-down event of a typed response. The recorded responses were analyzed to compute percent-correct (PC) scores for words. For error trials, we also compiled position-based phoneme errors. Four measures were calculated:

- RT vs. Day
- Word PC Score vs. Day
- No. Phoneme Errors
- Phoneme Errors vs. Position in Word

RT vs. Day: RTs from correct responses recorded on the same day were used to create box plots of RT as a function of test day, from Day 5 to Day 15. It was expected that RT would decrease from Day 5 to Day 13 as the participants became more proficient at recognizing words in List#1. An increase was expected at Day 14 when the participants switched to the new List#2.

Word PC Score vs. Day: Percent-correct scores for word recognition were computed for each test day. We expected PC scores to increase from Day 5 to Day 13, and show a drop on Day 14 when the participants switched from word List#1 to List#2. A t-test was used to verify if there was a significant decrease from the word PC score on Day 13 to that on Day 14 with an alpha value of 0.05.

No. Phoneme Errors: Number of phoneme errors for the four-phoneme words could vary from 1 to 4 per word. They were computed from data collected from word recognition tests without feedback, separately for word List#1 and List#2, from Day 5 to Day 13, and from Days 14 and 15, respectively. It was expected that the relative frequency of making 2 errors at the middle positions of the 4-phoneme words would be the highest among the 4 different error types.

Phoneme Errors vs. Position in Word: In addition to the total number of phoneme errors per word, incorrect responses were further analyzed to observe phoneme error patterns as a function of their positions in each word. We expected higher errors in the second and third positions of a word as compared to those in the first and last positions. Analyses were performed separately with data collected with word List#1 (Day 5 to Day 13) and those with word List#2 (Days 14 and 15).

To verify an effect of number of phoneme errors and position-based errors on error rates, a chi-square test was used. If the chi-square test showed a significant effect, a t-test was used to compare the significant difference between two categories (i.e., significant difference between the proportion of words with 1 phoneme error and those with 2 phoneme errors).

4.3 Results

4.3.1 RT

The RT results as a function of test days are shown in Figure 4.1, separately for each participant. Shown on each subplot are the number of correct responses for List #1 (452, 474, 526 out of 600 for P1, P2, P3, respectively) and the same for List #2 (161, 140, 134 out of 200 for P1, P2, P3, respectively). Data for P1 show a clear decreasing trend from Day 5 to Day 13 when using List #1, with both the medians and the variance of the box plots decreasing and reaching plateaus (see Figure 4.1A). The median of RTs decreased from 2.55 sec on Day 5 to 0.45 sec on Day 13. The average RT of Day 5 and Day 13 were $3.03 \pm 2.00 sec$ and $0.46 \pm 0.23 sec$, respectively.

In Figure 4.1B, the RTs for P2 started at around the same level as P1's RT on Day 5, but stayed higher for the initial 4 test days before decreasing to a level that was about twice of P2's RT on Day 13. P2 showed a decrease of RTs from 3.18 sec on Day 5 to 1.06 sec on Day 13. Average RT values were $5.72 \pm 7.27 sec$ and $1.33 \pm 1.36 sec$ on Day 5 and Day 13, respectively.P3 showed the lowest RTs among the three participants tested, and the smallest decrement in RTs from Day 5 to Day 13 (see Figure 4.1C). Compared to median RT of 1.28 sec on Day 5, it decreased to 0.66 sec on Day 13. Whereas the average RT on Day 5 was $1.60 \pm 1.02 sec$, the average RT on Day 13 was $1.07 \pm 1.46 sec$.

These results show that in general, RTs improved (became lower) throughout the 9 test days as the participants became more proficient at recognizing the four-phoneme words in List #1. The RT results for List #2 collected on Day 14 and Day 15 are also displayed in Figure 4.1 in orange-colored box plots.

In Figure 4.1A, P1 shows a decreasing trend from Day 14 to Day 15 with both the median and variance decreasing. The median RT dropped from 1.05 sec on Day 14 to 0.72 sec on Day 15. P1's minimum RTs with List #2 remained approximately constant over the last two testing days. The average RTs on Day 14 and Day 15 were $1.22 \pm 0.90 sec$ and $0.85 \pm 0.49 sec$, respectively. Data for P2 showed a similar trend as P1. Both medians for Day 14 and Day 15 were higher than those of P1's, and the variance and median decreased from Day 14 to Day 15 (see figure 4.1B). A median RT of 2.24 sec on Day 14 decreased to a median RT of 1.37 sec on Day 13. The average RT on Day 14 was 2.68 ± 1.63*sec* and that on Day 15 was 1.71 ± 1.30*sec*. P3's data in Figure 4.1C also displayed an analogous trend as the previous two participants. The variance decreased from Day 14 to Day 15, and the median remained nearly constant. The median RTs on Day 14 and Day 15 were 1.13 sec and 1.11 sec, respectively. The average RT on Day 14 and Day 15 were 1.46 ± 0.88*sec* and 1.32 ± 0.82*sec*, respectively. For all participants, it was observed that there was a cost in RT when the participants switched from the learned List #1 on Day 13 to the new List #2 on Day 14. All participants also showed a decrease in their maximum RTs from Day 14 to Day 15.

4.3.2 Percent-Correct Scores for Words

Figure 4.2 presents the word recognition percent-correct scores on each day for the three participants based on List #1 collected from Day 5 to Day 13, and List #2 on Days 14 and 15. Data for P1, P2, and P3 are shown in Figures 4.2A, 4.2B, and 4.2C, respectively. P1 achieved an average PC score of 74.7 ± 10.4% for List #1 and 80.5 ± 11.2% for List #2. Data for P2 shows an average PC score of 78.3 ± 13.2% and 69.5 ± 11.7% for List #1 and List #2, respectively. P3 achieved the highest average word PC score of 87.7 ± 8.3% for List #1 and the lowest average PC score of 67.0 ± 10.2% for List #2 among the three participants. The chance level from Day 5 to Day 13 was 1% (1 out of 100 for the closed-set list of 100 words) while the chance level for Day 14 and Day 15 was near 0% as List #2 testing was done in an open-set format. It is clear from Figure 4.2 that the participants' word recognition percent-correct scores were well above the chance level for both List #1 and List #2, demonstrating that the participants did remarkably well at identifying 4-phoneme words using the TAPS system. It is further observed that the participants P1 and P3 showed an increase in percent-correct score from Day 14 to Day 15 while P2's performance remained similar. Comparing this result with the data in Figure 4.1, there appears to be a general trend that the response time decreased as the percent-correct score increased. Additionally and as expected, there

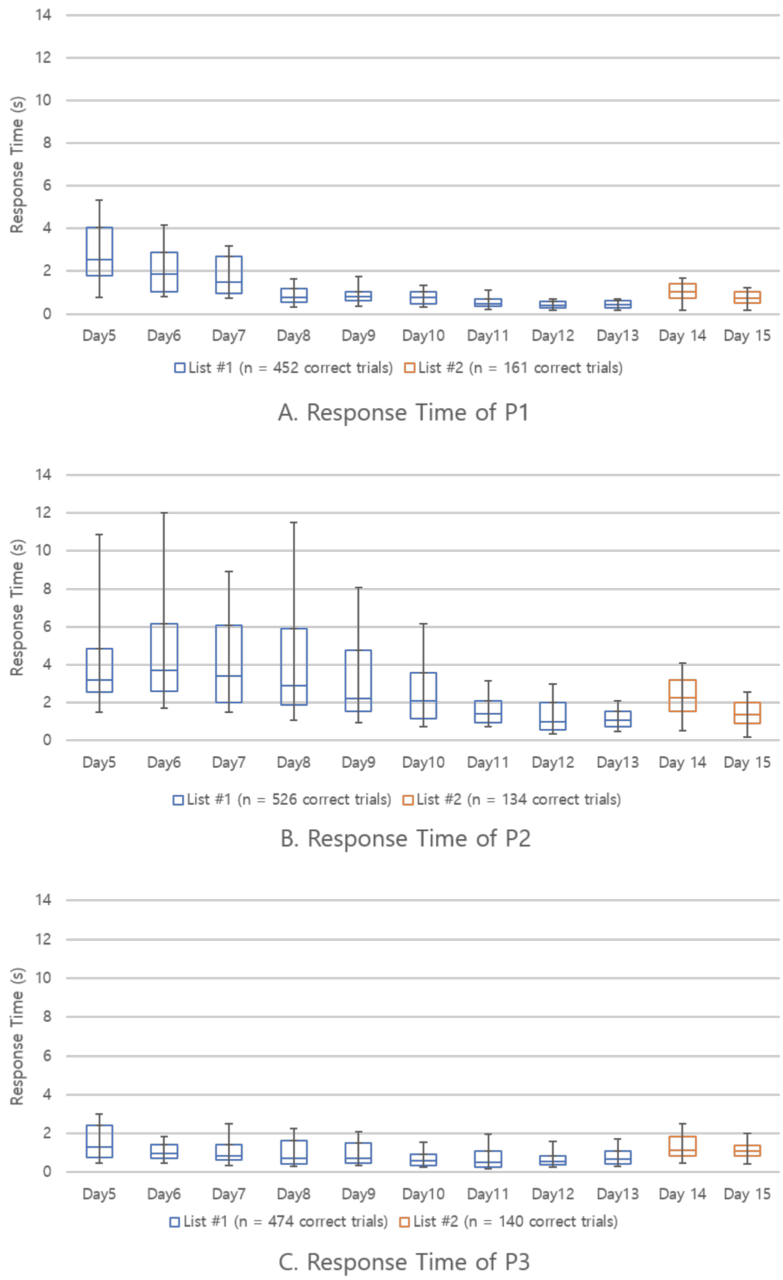
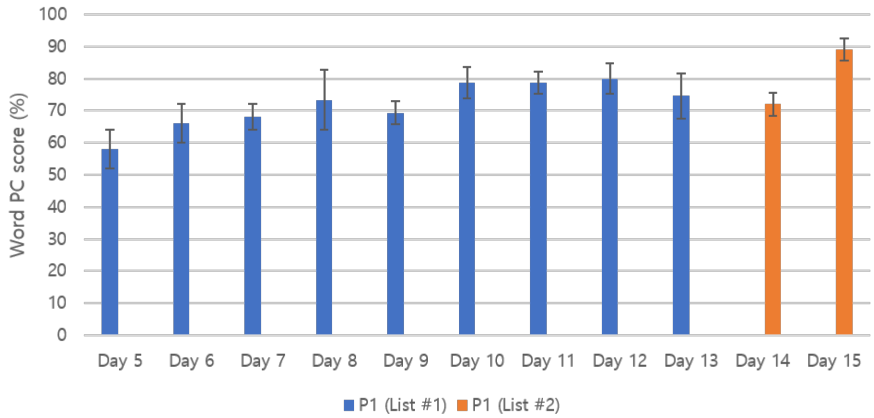
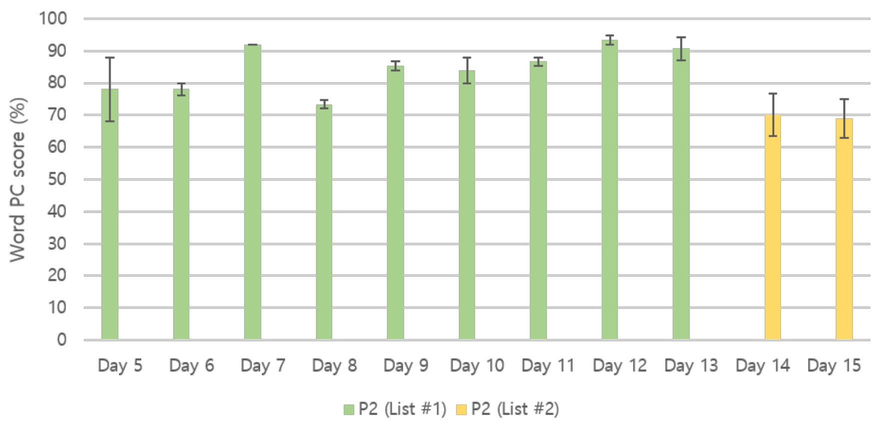


Fig. 4.1.: Box plots of daily RTs for participants P1 (A), P2 (B), and P3 (C).

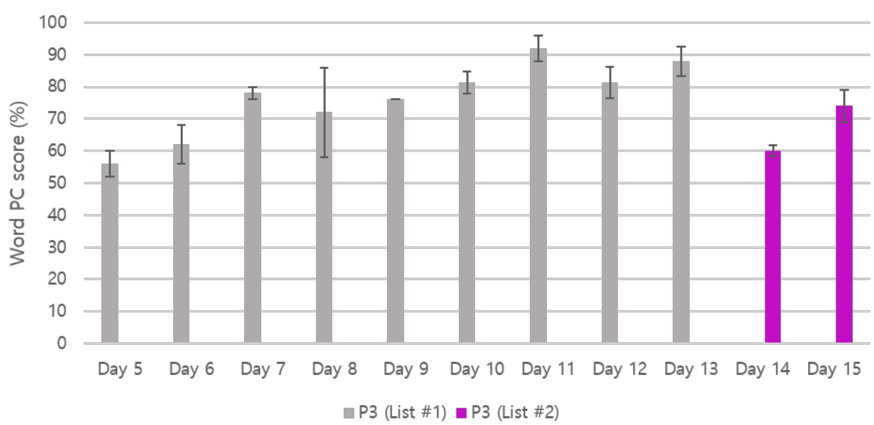
was a drop of word PC score from Day 13 to Day 14 when the transition from using List #1 to List #2 occurred. P1's data indicated that there was no significant between the word PC score on Day 13 and that on Day 14 ($t(3) = 1.6832$, $p = 0.1801$). Participant P2 showed a similar result which there was no significant difference between the word PC score on Day 13 and Day 14 ($t(5) = 2.1201$, $p = 0.0893$). Participant P3 showed a significant drop of word PC score from Day 13 to Day 14 ($t(2) = 6.0806$, $p = 0.0184$). However the drop was relatively small considering the near-zero chancel level, and the participants generally improved on Day 15. This showed that the participants were able to correctly identify a majority of the words by using TAPS.



A. Word PC score of P1



B. Word PC score of P2



C. Word PC score of P3

Fig. 4.2.: Percent-correct scores for identification of 4-phoneme words as a function of test days.

4.3.3 Number of Misrecognized Phonemes per Word

It is also important to know how many phonemes were not received correctly within each 4-phoneme words. To analyze this, incorrect responses from the three participants were collected. Within the 4-phoneme word, participants could make 1 to 4 errors regarding the identification of the 4 phonemes in the word. Figure 4.3 displays the percentage of each type of errors for each participant based on words in List #1. In Figure 4.3A for example, 15.5% of the incorrect responses made by participant P1 consisted only one mistake among 4 phonemes, and 29.1% of that had 2 errors among 4 phonemes, etc. As a result, the percentage values for each error type add up to 100% for each participant. Figure 4.3D shows the average of the proportion of each error type across the three participants with the error bars representing standard errors. As demonstrated in the figure, incorrect responses with 3-phoneme errors were the most common for all the participants. In comparison, 1-phoneme and 4-phoneme errors had the two lowest proportions. Through chi-square test, the error rate was dependent on the number of phoneme errors ($\chi^2(3, 348) = 69.8506$, $p = 4.5947 \times 10^{-15}$). In addition to the chi-square test, the t-test verified that a significant difference only occurred between the % of errors for words with 3 phoneme errors and those of the words with 4 phoneme errors (t(2) = 4.3390, p = 0.0492).

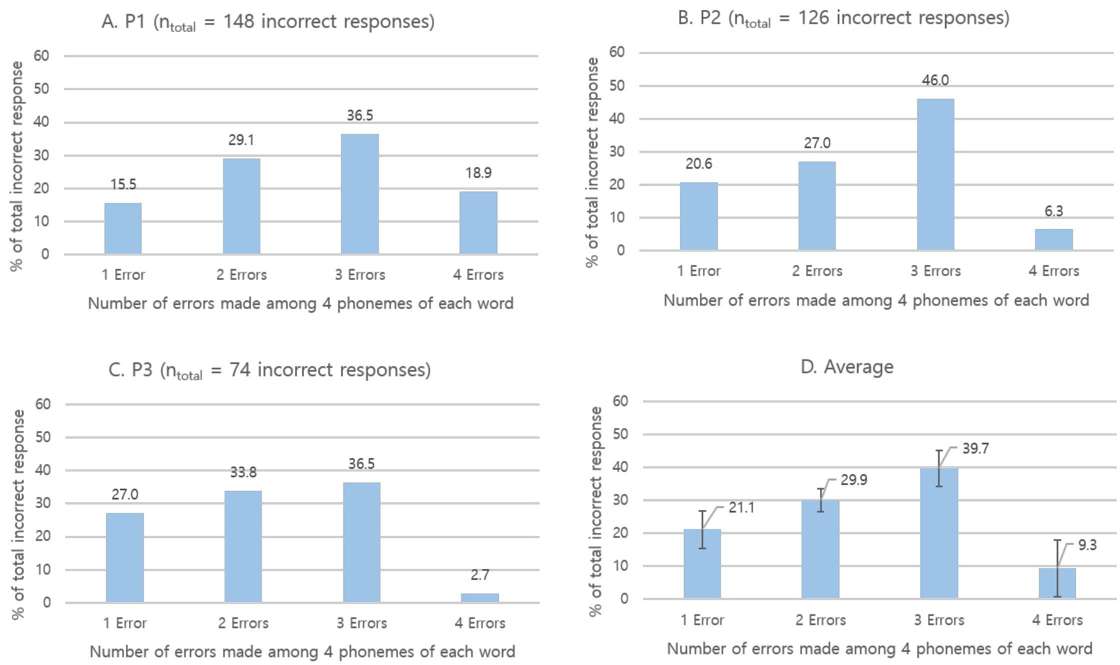


Fig. 4.3.: Percentage of total incorrect response as a function of number of phonemes misidentified in a 4-phoneme word based on List #1. Individual data for P1, P2, and P3 are shown in subplots (A)-(C), respectively. Average was taken from individual data and presented in subplot (D).

What would the phoneme error patterns be for words in List #2 with which the participants did not have a chance to practice and learn? Figure 4.4 shows the percentages of errors for the words in List #2, for individual participants and the average data. A different trend emerged for List #2. While the percentages of errors increased from 1 phoneme error to 3 errors in a word and then dropped for 4 phoneme errors for the words in List #1 (Figure 4.3), they decreased continuously from 1 phoneme error to 4 errors for the words in List #2 (Figure 4.4).

In Figure 4.4-(D), 44.1% and 37.1% of the incorrect responses were associated with misidentifying one or two phonemes in a word, respectively. Words with all four phonemes misidentified covered only 2.5% of the total incorrect responses. The individual participant's results followed the same trend as shown in Figure 4.4-(A), (B), and (C), except for P3's data for 1 phoneme error. The chi-square test indicated that the error rate was dependent on the number of phoneme errors in List #2 ($\chi^2(3, 348) = 71.1612, p = 2.4075 \times 10^{-15}$). Furthermore, the t-test verified that there was significant difference between % of errors for words with each phoneme error. For instance, there was a significant difference between % of errors for words with one phoneme errors and those of the words with two ($t(2) = 4.9822, p = 0.0380$), three ($t(2) = 5.3489, p = 0.0332$), or four ($t(2) = 8.2269, p = 0.0145$) phoneme errors. There was a significant difference between % of errors for words with two phoneme errors and those of the words with three ($t(2) = 5.2538, p = 0.0344$) or four ($t(2) = 9.1648, p = 0.0117$) phoneme errors. Lastly, a significant difference occurred between % of errors for words with three phoneme errors and those of the words with four ($t(2) = 11.0129, p = 0.0081$) phoneme errors.

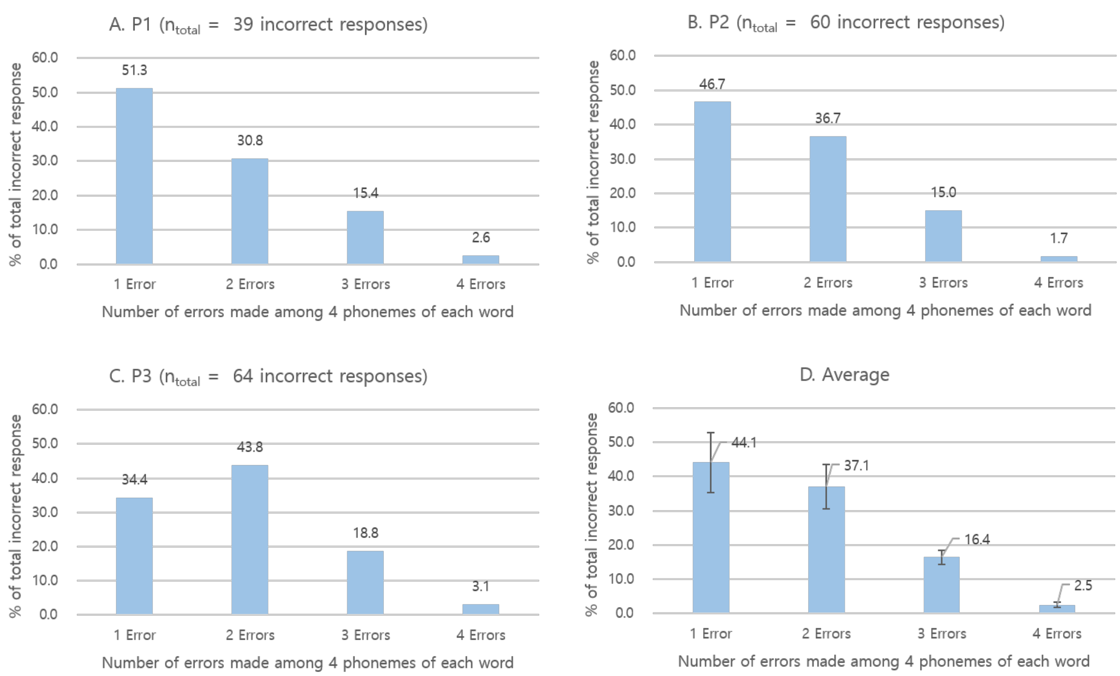


Fig. 4.4.: Percentage of total incorrect response as a function of number of phonemes misidentified in a 4-phoneme word based on the list #2. Individual data for P1, P2, and P3 are shown in figures (A)-(C), respectively. Average was taken from individual data and presented in (D).

4.3.4 Phoneme Errors by Position

The incorrect responses were then analyzed by looking at the positions of the phonemes that were incorrectly identified. Shown in Figure 4.5 there are the proportions of errors made at each phoneme position out of the total number of errors based on words in List #1. For example, data from participant P1 in Figure 4.5-(A) shows that most errors (30.8%) occurred at the 3rd phoneme in the word, followed closely by those at the 4^{th} phoneme position (29.2%). Figure 4.5-(D) illustrates the average percentages of errors made in each phoneme position across the participants with error bars indicating standard errors. It highlights that the majority of errors occurred at the 3^{rd} ($30.3 \pm 1.7\%$) and 4^{th} ($30.0 \pm 1.8\%$) phonemes and the least errors at the 1st phoneme position ($13.8 \pm 0.1\%$). The trend of the average data followed that of each participant's individual results. The chi-square test showed that the effect of errors based on phoneme positions in Figure 4.5 was significant on the error rates ($\chi^2(3, 842) = 60.4679, p = 4.2737 \times 10^{-13}$). Then, the t-test indicated that there were significant differences between the % of errors at 1st position and those at 2^{nd} ($t(2) = 110.5070, p = 8.2549 \times 10^{-5}$), 3^{rd} ($t(2) = 16.2977, p = 0.0037$), or 4^{th} ($t(2) = 15.3694, p = 0.0042$) positions.

Figure 4.6 presents similar results for the words in List #2, for individual participants and their average data. The overall trend for the words in List #1 and List #2 were similar, except that most errors occurred at the 3^{rd} position for the words in List #2. The chi-square test the error rate was dependent on the errors based on phoneme positions in Figure 4.6 ($\chi^2(3, 291) = 38.9079, p = 1.8155 \times 10^{-8}$). In addition, the t-test indicated that there were significant differences between % of errors at each position except between % of errors at 3^{rd} position and those at 4^{th} position. There were significant differences between % of errors at 1^{st} position and those at 2^{nd} ($t(2) = 4.6020, p = 0.0441$), 3^{rd} ($t(2) = 19.1537, p = 0.0117$), or 4^{th} ($t(2) = 20.9313, p = 0.0023$) positions. Significant differences were observed between % of errors at 2^{nd} position and % of errors at 3^{rd} ($t(2) = 4.3602, p = 0.0448$), or 4^{th} ($t(2) = 16.4.3240, p = 0.0495$) position.

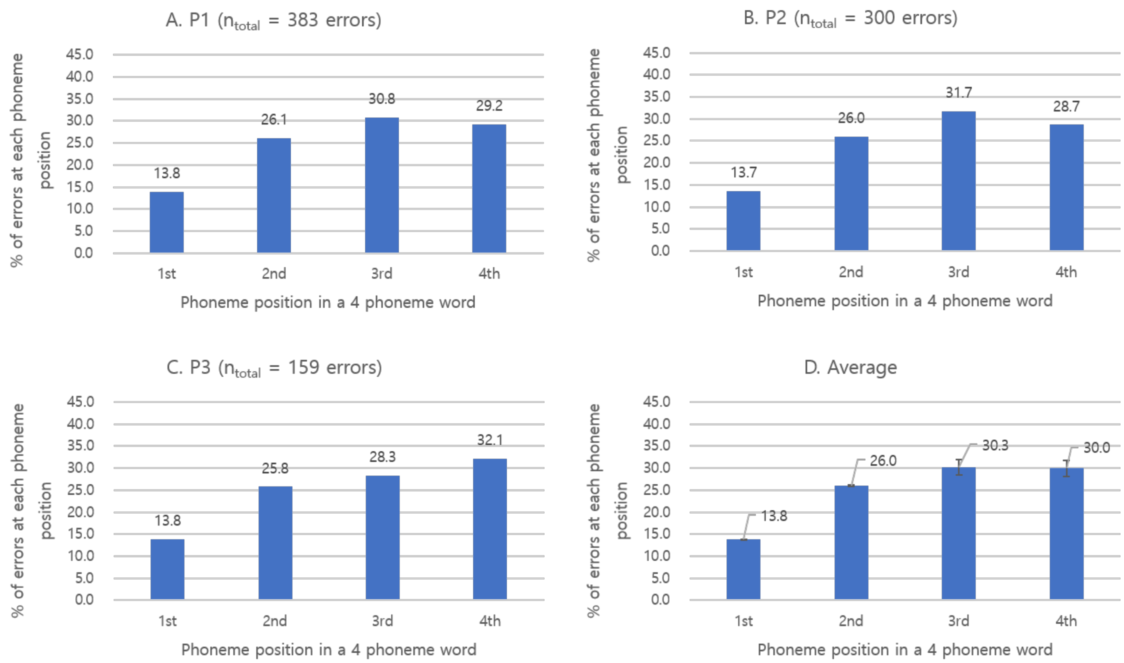


Fig. 4.5.: Percentage of errors by position based on words in List #1. Individual data for P1, P2, and P3 are shown in subplots (A)-(C), respectively. Average data are presented in subplot (D).

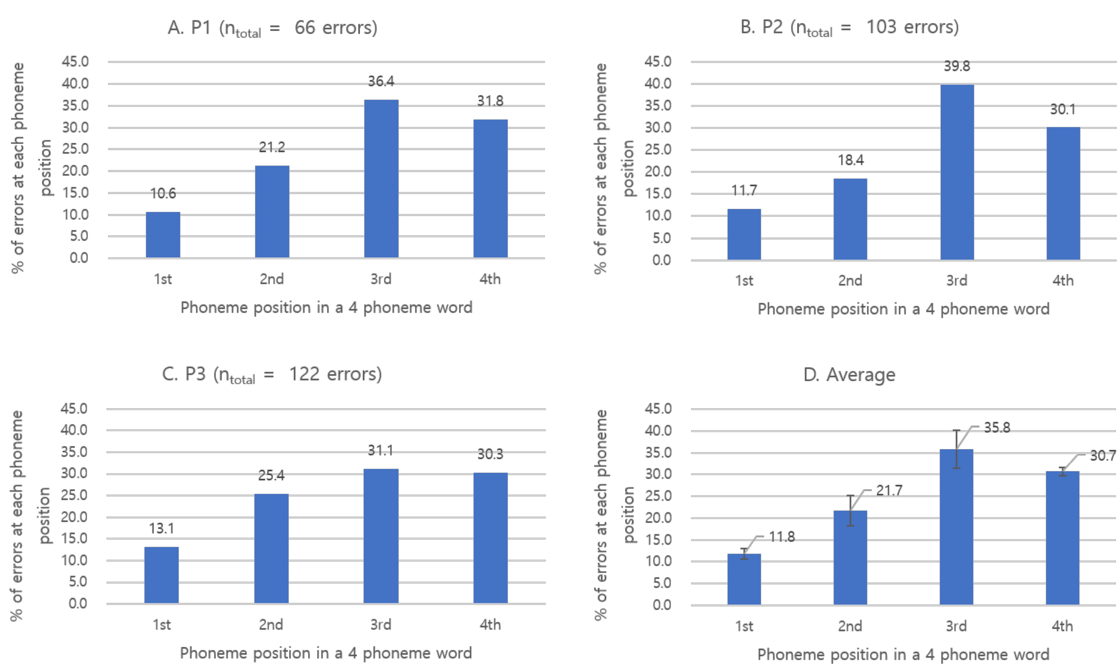


Fig. 4.6.: Percentage of errors by position based on words in List #2. Individual data for P1, P2, and P3 are shown in subplots (A)-(C), respectively. Average data are presented in subplot (D).

4.4 Discussion

In this study, three experienced participants learned 4-phoneme words over a period of 10 days and were tested on their generalization of learning using TAPS. From Day 4 to Day 13, P1, P2, and P3 practiced and were tested with 100 four-phoneme words in List #1, and the word percent-correct (PC) scores were $74.7 \pm 10.4\%$, $78.3 \pm 13.2\%$, and $87.7 \pm 8.33\%$ for the three participants, respectively. From Day 14 to Day 15, they were tested on a new 100 word list (List #2) that also contained 4-phoneme words to investigate the generalization of learning. The word PC scores were 80.5 ± 11.2% for P1, 69.5 ± 11.7% for P2, and 67.0 ± 10.2% for P3. Based on the PC scores that the participants achieved, they were clearly performing well above the chance level of 1% (for List #1) or near 0% (for List #2). In addition, participant P1 and P2 indicated that they could apply learning acquired from List #1 to a new word list (List #2) which they did not have significant difference between word PC score on Day 13 and Day 14. However, participant P3 showed a cost of word PC score from List #1 (Day 13) to List #2 (Day 14) which was expected since the testing material was changed. Furthermore, the participants showed a decreasing trend in response time with a concurrent increasing trend in word PC scores. The incorrect responses were analyzed by looking at the positions of missed phonemes in each 4-phoneme word and the number of missed phonemes in each word.

Based on the results from Figure 4.3, participants showed that the error rates were dependent on the number of phoneme errors. In addition, it was observed that they tended to make 3 errors among four phonemes rather than misrecognizing all 4 phonemes. Based on this result, it was deduced that this error pattern could be caused by using one of the phonemes to guess the words. In other words, participants might have focused more on perceiving one of the phonemes than focusing on all phonemes equally well. Especially for List #1, it was possible that the participants could have used mainly the first phoneme to identify the words due to the fact that they had already been exposed to the word list before testing and knew what words to expect after receiving the first phoneme. Figure 4.5 supports this hypothesis in that the smallest percentage of errors was made in the first position for all three participants. This indicates that the participants identified the first phonemes mostly correctly as compared to phonemes in other positions.

The strategy discussed above might have been applicable not only to words that the participants have experienced already but also to words that participants had not felt. In Figure 4.4, the % of errors for words with 1 phoneme errors was the highest whereas those of the words with 4 phoneme errors was the lowest. This result from Figure 4.4 suggests an error pattern that the participants used most of the phonemes to guess the words. For instance, the participants could have used the first few phonemes to identify the words since the participants have not felt the words in List #2. Figure 4.6 further supports this claim by showing that the % of the errors at 1st phoneme position was the lowest and % of the errors at 2nd phoneme position was higher but lower that those at 3rd or 4th phoneme position. This presents that the participants invested their concentration mostly in the 1st phoneme position followed by less concentration in the 2nd. And they focused least on the 3rd or 4th phoneme position. This can be caused since the participants were assessed with new 4-phoneme words and could not anticipate what words they were being tested on. Hence, they could have put effort to identify not only the first phoneme but also the 2nd, 3rd, and 4th phonemes in a word.

The error patterns of the present study showed a different trend compared to those in another study of TAPS [14] that investigated the error rates of words consisting of 2 and 3 phonemes which were based on the same 39 haptic symbols used in the present study. In Reed et al.'s (2020) study [14], seven participants using TAPS were assessed with a 100-word list. Of the 100 words, 31 were 2-phoneme words and the remaining 69 were 3-phoneme words. The participants initially had training sessions for learning the 100-word list. The training session continued until the participants achieved a word PC score of above 70% with trial-by-trial correct-answer feedback. Then, the participants were assessed with three 50-trial runs of words that were randomly selected from the same word list. The average word PC score across the participants with an IPI of 150ms was 88.9%. It was observed that the error rates were not dependent on the number of misrecognized phonemes or phoneme position. It was concluded that the participants attended to the "complete phonetic make-up" [14]. In comparison, the results of the present study for 4-phoneme words demonstrated that the participants tended to focus more on at least one of the phonemes of a word. This difference in results might have been due to the change in the length of words. Compared to words with 2 or 3 phonemes, 4-phoneme words have a longer duration which required a greater amount of memory. The participants might have preferred to concentrate more on identifying the initial phoneme and deduce the words while receiving other phonemes rather than focusing equally on receiving all the four phonemes which demanded a higher cognitive load.

The word PC scores obtained in the present study can be compared to two past studies of the TAPS system on word acquisition using a list of 100 English words [13] and 500 words [7], both made up of all 39 phonemes. The 100 words used in Jiao et al. (2018) were composed of 2- and 3-phoneme words. The average PC score of 8 participants who were in the phoneme-based learning group (i.e., they learned phonemes first before learning words, as opposed to learning words from the beginning) within a learning period of 100 min was 80.6% [13]. The 500 words used in Tan et al. (2020) contained mostly words composed of 3 phonemes, with less than 10% of the 500 words composed of 4- and 5-phoneme words. The average word recognition PC score of 11 experienced participants was 80.7%. In comparison, the 100 words used in the present study contained more phonemes than those in Jiao et al.'s (2018) and Tan et al.'s (2020) studies and the PC scores were generally comparable to those obtained in the past two studies.

Our results can also be compared with those reported by other past studies on tactile devices that transmitted English words through tactile stimuli. One of the early tactile devices is called the tactile vocoder. The tactile vocoder delivers and filters acoustic waveforms into 16 vibrotactile actuators placed on the skin [2]. One of the initial studies using the tactile vocoder was on learning 250 words. Participants who had experienced the vocoder for 55 hours with word training and 20 hours of generalization training were tested on 150 words that they had already learned and an additional 100 new words. During the experiment, it took an additional 25.5 hours to achieve a percent-correct score of 76% with the 100 new words. The overall word recognition percent-correct score for the 250 words from 2500 trials was 75.6% [8]. A further study on the vocoder was conducted on a large word list with 1000 words. A participant experienced with the vocoder reached a word PC score of only 8.8% after 196 hours [2]. Zhao et al. (2018) introduced another haptic device worn on the forearm that delivers English words. The device delivers tactile symbols through 6 voice-coil actuators and the symbols were matched to 9 phonemes. The participants were tested on 20 words with 2- or 3-phonemes and scored a word recognition percent score of 83% (study 2) after roughly 47 min of phoneme recognition and word recognition tests [4]. Novich et al. (2015) devised and tested a spectral-based haptic vest which uses 27 tactors to deliver 50 English words. A test was done based on a four-alternative forced-choice (4AFC) identification method. After training for 11 to 12 days, their participants reached percent-correct scores that ranged from 35% to 65% compared to a chance level of 25% (one out of four response alternatives) [3]. Another study to deliver English words was done based on a device that used an electrotactile speech communication device called the Tickle Talker. In this study, the participants had already spent 12 to 33 hours in word recognition training by using the device. They were then tested on a word list that contained 20 trained words. The average word recognition percent-correct score was 42.5% (chance level 5%) for trained words [64]. Compared to the results obtained with other tactile devices, the results from the present study were similar or much better in terms of learning time and PC scores.

The present study demonstrates the efficacy of the TAPS system further in its learning generalizability. A generalization study done by Galvin et al. (2000) through the Tickle Talker presented a word PC score of 22.5% (chance level near 0) for 20 untrained words, which was about half of that for 20 trained words (42.5%, as cited above) [64]. Compared to this result, even though the TAPS system showed a similar decreasing trend in word PC score from Day 13 (List #1) to Day 14 (List #2), the average word PC score for List #2 was an impressive 72.3%, which was much higher than the word PC score of 22.5% for the 20 untrained words in Galvin et al.'s (2000) study. Hence, the TAPS system is more effective for generalization of learning in terms of a more than three-fold increase in PC score given a five-fold increase in vocabulary size for untrained words.

Overall, the present study has shown that experienced participants were able to acquire longer (4-phoneme) words with the TAPS system and that the learning was generalizable to a new list of 4-phoneme words. Compared to the earlier studies that used TAPS (see [13] and [7]), the average word recognition PC score of this study did not drop notably even though the participants were tested on the 4-phoneme words that were longer than the majority of the words tested in the previous studies. The results for the generalization study showed a slight drop for participants P1 and P2 and a significant drop for the participant P3 in word PC score that nonetheless remained impressively high, indicating that the participants were able to generalize their learning of longer words through the TAPS system. The results presented in this chapter also demonstrated that the learning time and word PC scores were similar or better when compared with the results of studies using different tactile aids. Hence, the present study supports the conclusion that users of the TAPS system can learn longer words within a reasonable amount of time.

5. EVALUATION OF TWO-WAY COMMUNICATION THROUGH TAPS

Prior research with the TAPS (up to Tan et al. 2020 [7]) and the experiment on four-phoneme word acquisition (Chapter 4 of this thesis) have demonstrated the remarkable ability of English speakers at acquiring any English word on the forearm at a practical learning rate of one word per minute [7]. The study described in this chapter investigated the ability to conduct two-way communication using two identical TAPS systems worn by two experienced participants. In a setup that mimicked "texting," participant A typed a text message on a computer screen that was received by participant B via TAPS. Participant B interpreted the tactile message, occasionally "replayed" the message on TAPS for clarification, and responded by typing another text message on a second computer screen that was sent to participant A's TAPS. This process continued for a predetermined period of time on a testing day, during the Spring and Fall of 2019 and Spring of 2020. The details of the "conversation" in terms of the text messages sent and time spent were recorded for further analysis. The research presented in this chapter sheds light on the feasibility of TAPS for tactile communication of spontaneous speech, which is the ultimate goal of any device-mediated tactile speech communication system.

5.1 Methods

This section illustrates the experimental methods for the two participants to send text messages and perceive messages through the TAPS system. The same TActile Phonemic Sleeve (TAPS) system was used as presented in Section 3.3.2 in this thesis. The procedures for intensity calibration and tactor equalization were the same as those described in Section 3.3.4.

5.1.1 Participants

Two participants, P1 and P3 who had taken part in the study presented in Chapter 4, completed this study. They were 25 and 23 years old, respectively, at the beginning of this study. They were chosen for their extensive experience with the TAPS systems and their availability to conduct the two-way communication experiment over multiple semesters. Their characteristics were explained in Section 4.2.1.

5.1.2 Experimental Setup

Two identical stations were set up on two adjacent tables. Each station consisted of a PC, a monitor, and a complete TAPS system (see Section 3.3.2). The same calibration procedures regarding detection threshold estimation and tactor equalization as elaborated in Section 3.3.4 were followed.

A TCP server-client protocol was implemented to enable the communication between the two PCs. At the beginning of each experiment, the TCP client was connected to the TCP server. The client initiated a message to the TCP server, and the TCP server responded. The communication continued afterwards.

The FLITE system developed by Carnegie Mellon University was adapted as the text-to-speech (TTS) front-end to TAPS. The FLITE is an open source TTS system intended for speech synthesis on small embedded machines [65]. It converts text input into a phonemic transcription that is fed into a speech synthesizer. In our case, the phonemic transcription was used as the input to the TAPS system that converted the phoneme stream into haptic symbols.

Two timing parameters needed to be determined prior to the experiment: the Inter-Phoneme Interval (IPI) and the Inter-Word Interval (IWI). When an input message is transmitted to TAPS, it is converted into a series of haptic symbols designed for the 39 phonemes and transmited to the skin of the forearm (see [6]). The IPI refers to the pause between phonemes. The IWI indicates the temporal gap between words. These two values are important in tactile speech communication since it affects the communication rate and the reception of words. If the IPI or IWI was too short, masking between the stimuli may occur and the participants may not have enough time to process each phoneme before the next one arrived. If the IPI or IWI was too long, the transmission rate would be reduced and the participants may find it harder to retain the memory of the previous stimulus while receiving and processing the next stimulus [14]. Considering that the participants may wish to start with longer intervals and then reduce them as they became more proficient, the user interface allowed these two parameters to be set manually at the beginning of each "texting" session, and their values were recorded. Based on the results from an earlier study (i.e., [14]), both participants started with an IPI of 150 ms and an IWI of 300 ms. The 150-ms IPI value was higher than the minimum value used in [14] (0 ms) and the IWI was the same as the minimum value used in [14]. The longer IPI value was chosen for the two-way communication task which was expected to be more difficult than the phoneme and word identification tasks used in the earlier study. After the first 10 days of two-way communication, however, the participants decided that 300 ms was too short for the IWI and increased it to 500 ms for the rest of the experiment from Day 11 to Day 45.

Another modification that turned out to be very useful was the use of a haptic signal prior to the transmission of each text message. This helped the participant receiving the message to be ready. We used the "knocking" signal designed by Shim & Tan (2020) for getting the receiver's attention [66]. The duration of the knocking signal was approximately 1.2 seconds.

5.1.3 Procedures

Between March 18, 2019 and February 24, 2020, the two participants engaged in two-way texting communications using the TAPS on 45 days when they were both available. The daily time spent ranged from about 8 min to 47 min with an average of 28 ± 7 min. The participants spent more time on the first day (47 min) than on the remaining 44 days to get used to the software interface used for sending text messages and entering responses. The daily time dropped to 8 to 31 min (average: 16 ± 6 min) for the next 44 testing days.

There was no specific time requirement per day. Instead, the participants worked together until they felt tired or lost concentration, which typically lasted 10 to 20 min. An overview of the time spent on each testing day as a function of calendar dates over the entire period from March 2019 to February 2020 is shown in Figure 5.1. The two gaps corresponded to summer vacation in 2019 and the following winter break. A similar plot of testing time on a monthly basis is shown in Figure 5.2.

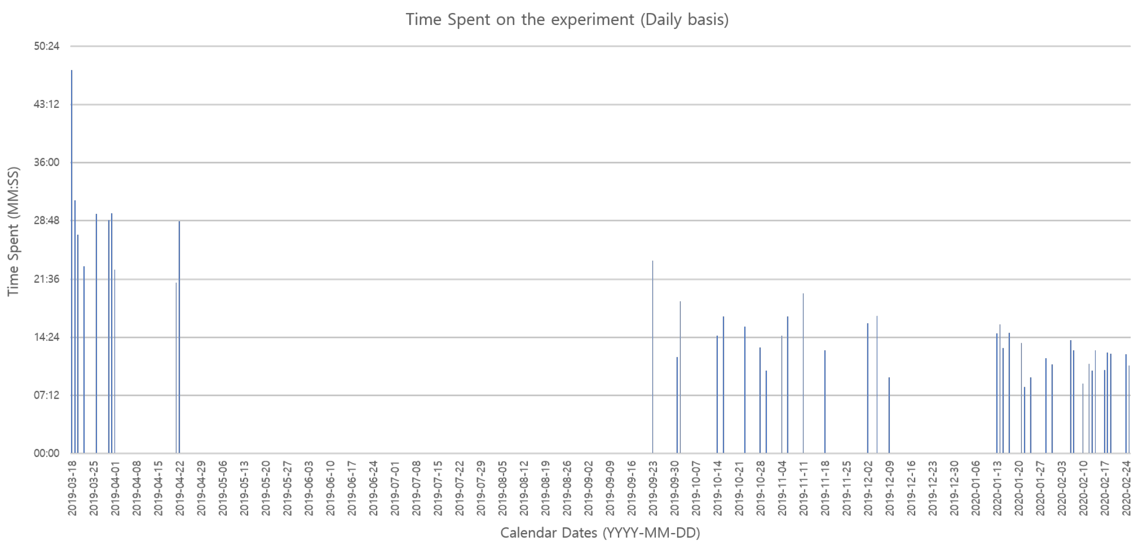


Fig. 5.1.: Daily time spent as a function of calendar dates.

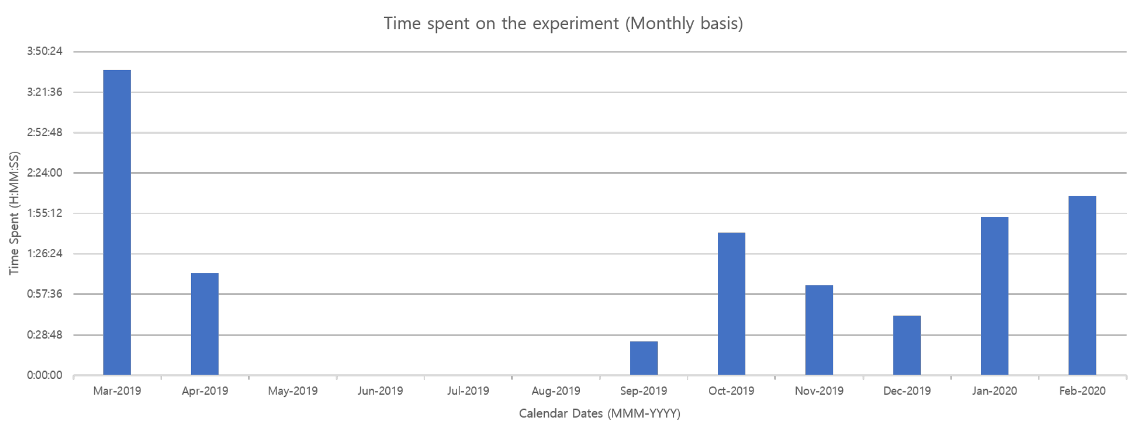


Fig. 5.2.: Monthly time spent as a function of calendar months.

On each day, the participants started by agreeing on a general topic for their conversation. The vocabulary was completely open: any English words could be used, and any number of words could be sent in a message. One participant, the "sender," would start by typing a message into a text box in the experimental user interface, and press a "send" button when finished. The other participant, the "receiver," would get the special "knocking" signal on TAPS, followed by the sequence of phonemes transcribed from the typed text message by the TTS software. The "receiver" had to type the received message into a text box so it could be compared with the sent message later. The "receiver" then becomes the "sender" and initiates the next message. As would happen with any conversation, the participants sometimes changed the topic in the midst of their chat and carried on with the experiment. An example dialogue that lasted about 12 min is shown below, with the number of times that the receiver replayed the messages shown in parenthesis.

- Start of the dialogue -

P1: so sorry

P3: no worries (repeated 1 time)

P1: how is your project

P3: should start testing (repeated 2 times)

P1: how many participants (repeated 1 time)

P3: I think five (repeated 1 time)

P1: how long does it take

P3: I think three days (repeated 3 times)

P1: starting this week (repeated 2 times)

P3: yes how about you

P1: next week probably (repeated 1 time)

P3: how is the interface (repeated 4 times)

P1: not bad still making (repeated 4 times

- End of the dialogue -

Figure 5.3 shows the same dialogue introduced above step by step. P1 initiated the conversation with "so sorry" and sent it to P3. Upon receiving the message, P3 recorded “so sorry” (correctly in this instance), and sent “no worries” back to P1. P1 was unable to identify the message upon its first transmission, and clicked on the “Replay last” button once on the computer screen to feel it again. After feeling the receiving the message once again, P1 then responded by typing “no worries” and saved it. The “Replay last” button could be used as many times as the participant wished and its number of usages was recorded along with the response. The number of times that the messages were repeated by the receiver were shown in parenthesis in the dialogue shown above this paragraph. After entering received message and saving it, the participants could click on the ”Show” button on the screen to check whether they were correct or not. This is useful in cases where the message could not be interpreted even after replaying it several times, so the dialogue could continue regardless of whether a message could be received.

P1 then proceeded to send a new message “how is your project” to P3. The message was correctly received by P3 upon its first transmission, the response was recorded, and P3 sent “should start testing” in response to P1’s question.

The following time stamps and text messages were recorded from each session between the two participants: start time of the “knocking” signal, start time of a message, the message itself (as text), end time of the message, start of response as indicated by the first key-down event following the message, number of repeats, and the response (as text). A timing diagram for a typical message is shown in Figure 5.4, where response time was defined as the time between the end of a transmitted message and the first stroke of typing a response. It the participant repeated any messages on TAPS, the time was included in the response time. The same information was logged for all messages transmitted between the participants and analyzed later. It was expected that as the experiment went on, the participants would improve their ability to receive text messages on their forearms in the form of increased accuracy and transmission rate.

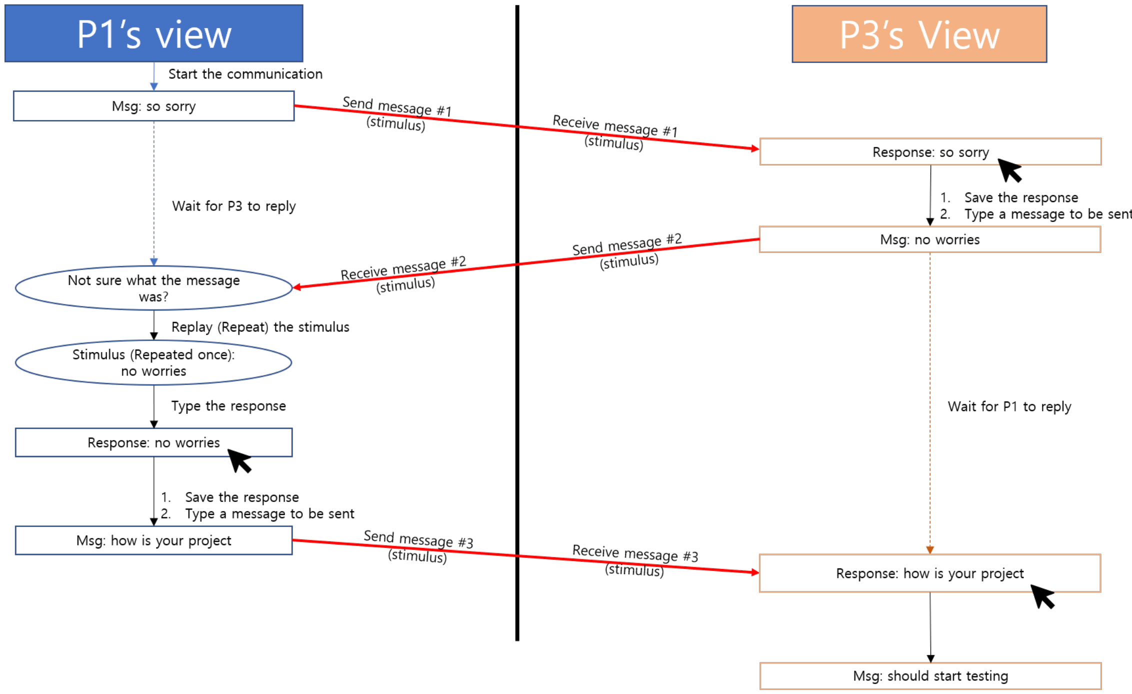


Fig. 5.3.: Step-by-step illustration of the dialogue shown above. Rectangles contain actual messages sent between the two participants. Ovals contain actions taken by a participant, such as repeating a stimulus.

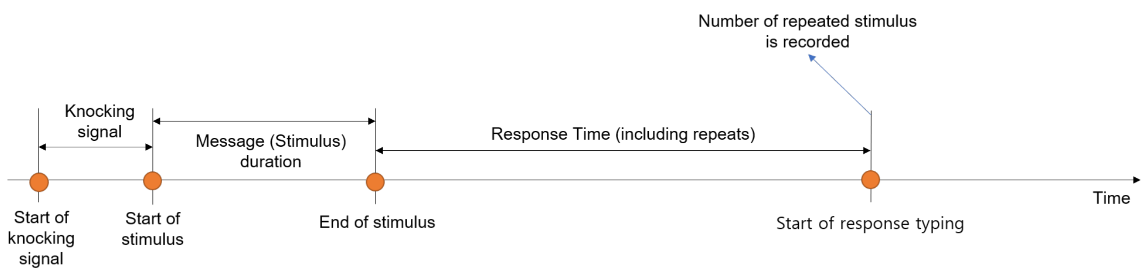


Fig. 5.4.: Timing diagram for a message.

5.1.4 Data Analysis

Data on communication through the use of TAPS between the two participants were collected for 45 days. Percent-correct (PC) scores for messages or words, PC_{msg} and PC_{word}, were computed. The total amount of time spent in the experiment on each day (session) was logged. Communication rates based on messages or words, $MsgPM$ and WPM, were also computed. The following paragraphs detail the calculations of the PC scores and communication rates.

PC for messages (PC_{msg}): For each transmitted message, the receiver's response was compared with the sender's original message. A score of 1 was assigned if the response was identical to the sent message. Otherwise, a score of 0 was assigned. The analysis was performed on all messages regardless of who was the sender or receiver. The daily PC_{msg} was calculated by dividing the number of correctly received messages by the total number of messages transmitted on each day. Due to the relatively few (entire) messages transmitted per day, the daily PC_{msg} scores were also averaged over multiple days in order to see the trend more clearly. It was expected that there would be an improvement of PC_{msg} scores as the participants became more proficient at the task.

PC for words (PC_{word}): It was recognized that the PC_{msg} measure was quite demanding, and that successful communication could be accomplished even if some words were not correctly received in a message. Therefore PC scores for words were also calculated. For each transmitted message, the PC_{word} score was calculated by dividing the number of correctly received words by the total number of words transmitted. The daily PC_{word} score was then obtained by averaging the PC_{word} scores for the same day. We expected that the PC_{word} score would increase as time went on.

Communication rate in messages per minute ($MsgPM$): Similar to PC scores, communication rates were calculated based on messages or words. The $MsgPM$ score was computed by dividing the number of correctly received messages per day by the total time in minutes for the day. It was hoped that the communication rate in $MsgPM$ would improve after several days of spending time with the TAPS and the interface.

Communication rate in words per minute (WPM): Similarly, the communication rate in words per minute (WPM) was calculated as the total number of correctly received words on each day divided by the total time in minutes for the day. We also expected that the transmission rate in WPM would improve over time.

To verify improvement of PC scores and communication over time, a linear regression was derived and its slope was compared to a slope of zero using a t-test. To conduct the t-test on PC score, the rationalized arcsine transformation of PC_{msg} and PC_{word} was used. This was because the PC score was based on a binomial distribution, meaning that the variance would be dependent on the mean of the data and disobey the assumption of the t-test. Hence, the rationalized arcsine transformation of PC scores was used to fit in the t-test [67], [68].

5.2 Results

The participants each spent a total of 762 minutes (12.7 hours) in this experiment. The results of PC_{msg} as a function of experimental days for both participants are presented in Figure 5.5. The blue bars represent the data of P1 while the orange bars show those of P3. Over the 45 days, the average PC_{msg} was 69.0% ± 18.9% for P1 and 77.8% ± 21.1% for P3. The minimum PC_{msg} for P1 and P3 were 25.0% (Day 1, Day 20, Day 25) and 20.0% (Day 20, Day 23), respectively. It is difficult to discern a clear pattern as the participants' performance fluctuated on a daily basis. There appears to be a "difficult day" (Day 20) in the middle of the 45 days where both participants had their respective minimum PC_{msg}. Daily log showed that the experiment was conducted early in the morning on that day, which may have contributed to the lower performance levels.

A 5-day moving average window was applied to the data in Figure 5.5 to smooth the data. The results are shown in Figure 5.6 and Figure 5.7 for P1 and P3, respectively, along with the linear regression lines. It can be seen that P1's PC_{msg} scores started and ended at similar levels, with a slight dip in the middle. It should then come as no surprise that a linear regression shows a R^2 value of 0.0019, and the slope of the linear regression compared to a slope of 0 is not significant (t(39)=0.3163, p=0.3767). In Figure 5.7, the 5-day PC_{msg} averages for P3 show a slight improving trend with a dip around Day 20 and a smaller dip around Day 33. A linear regression resulted in a modest R^2 value of 0.4586. A t-test comparing the slope with 0 indicates that the linear trend is significant (t(39)=6.1578, p<0.00001), suggesting that P3's ability to communicate through TAPS as measured by the PC_{msg} scores improved over time.

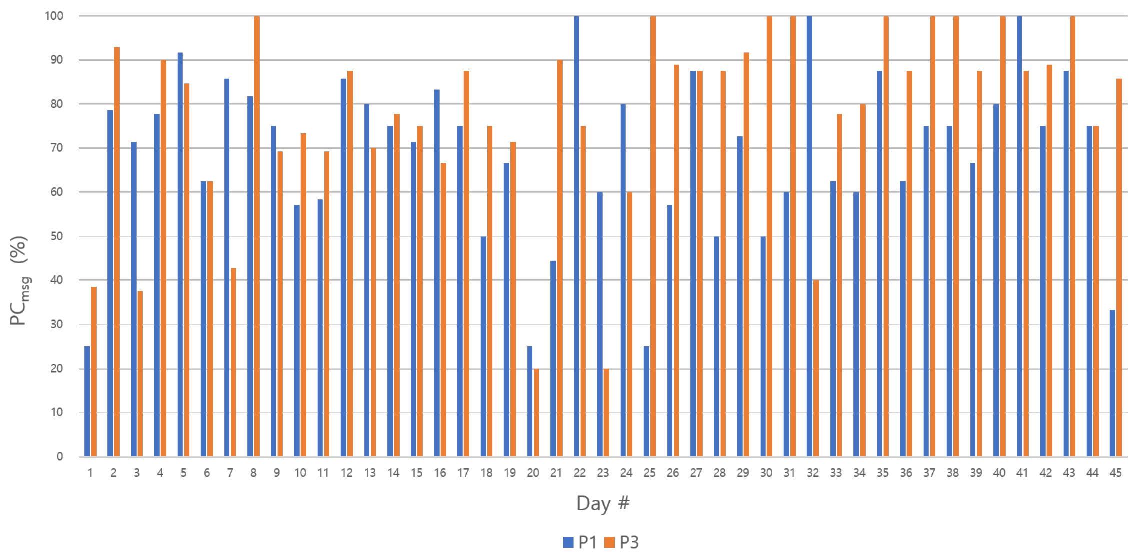


Fig. 5.5.: Daily PC_{msg} scores for both participants.

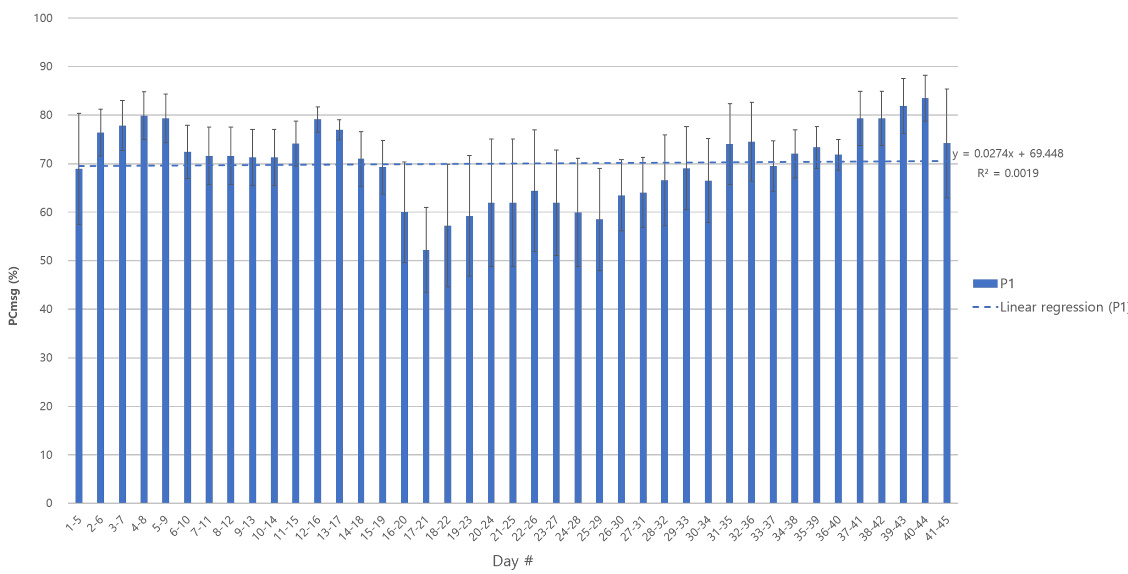


Fig. 5.6.: Moving average of P1's PC_{msg} scores with a 5-day window.

The PC_{word} scores show very similar patterns as the PC_{msg} scores. Figure 5.8 presents the PC_{word} scores as a function of test days. The average PC_{word} of P1 was 78.2% ± 13.3% whereas that of P3 was 85.2% ± 14.4%. Figures 5.9 and 5.10 show the 5-day running averages for P1 and P3, respectively. Again, P1's data were not well modeled by a linear

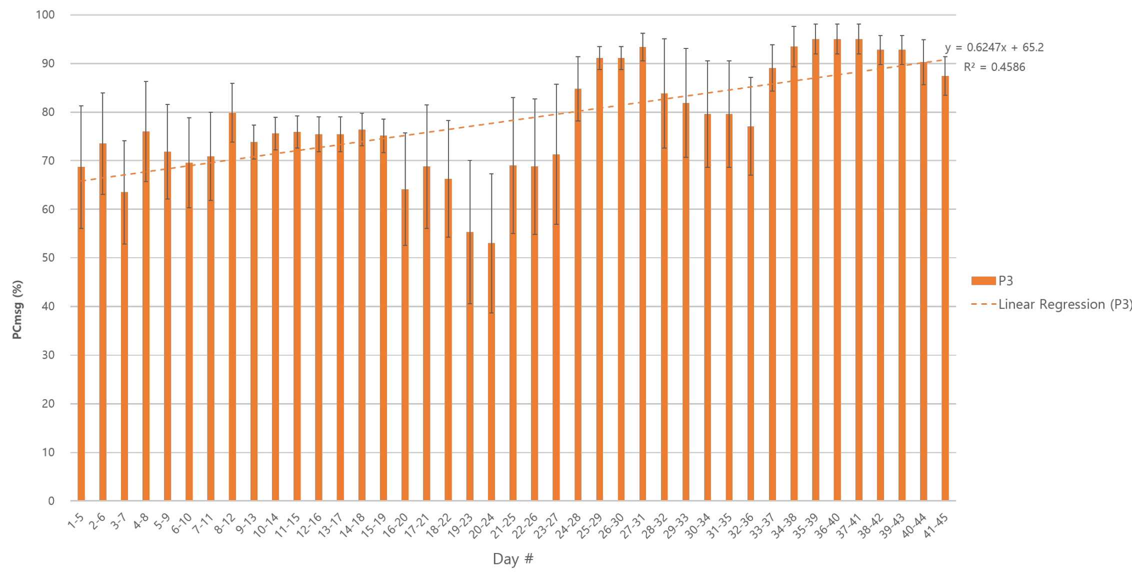


Fig. 5.7.: Moving average of P3's PC_{msg} scores with a 5-day window.

regression analysis (R^2 = 0.0033; t(39)=0.4008, p=0.3544). In Figure 5.10, P3's data show an increasing trend which turned out to be significant (t(39)= 6.2418, p<0.00001).

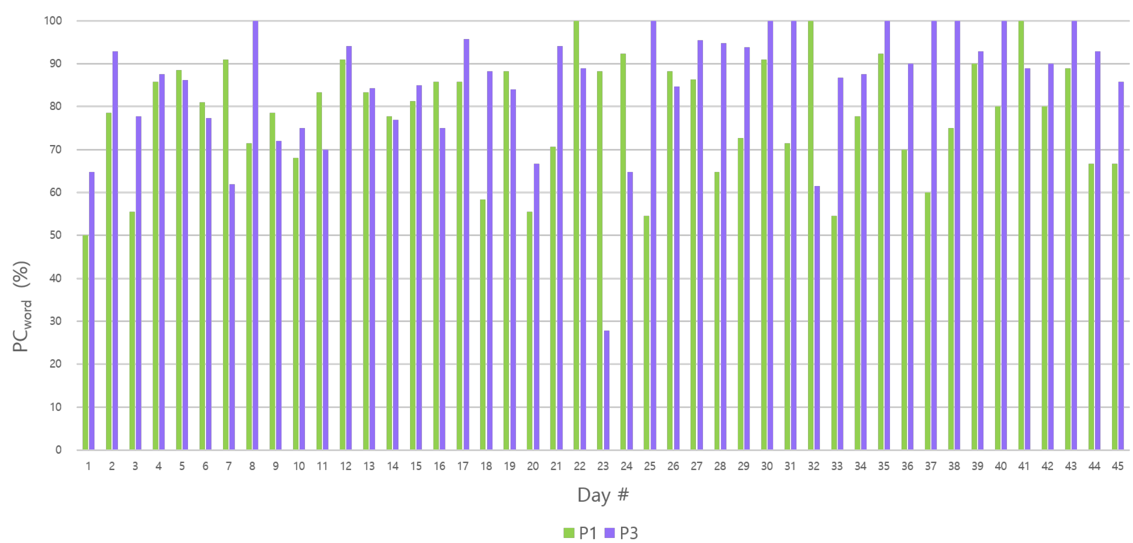


Fig. 5.8.: Daily PC_{word} scores for both participants.

The communication rates in messages per minute ($MsgPM$) were then analyzed as a function of test days as shown in Figure 5.11. The average communication rates were

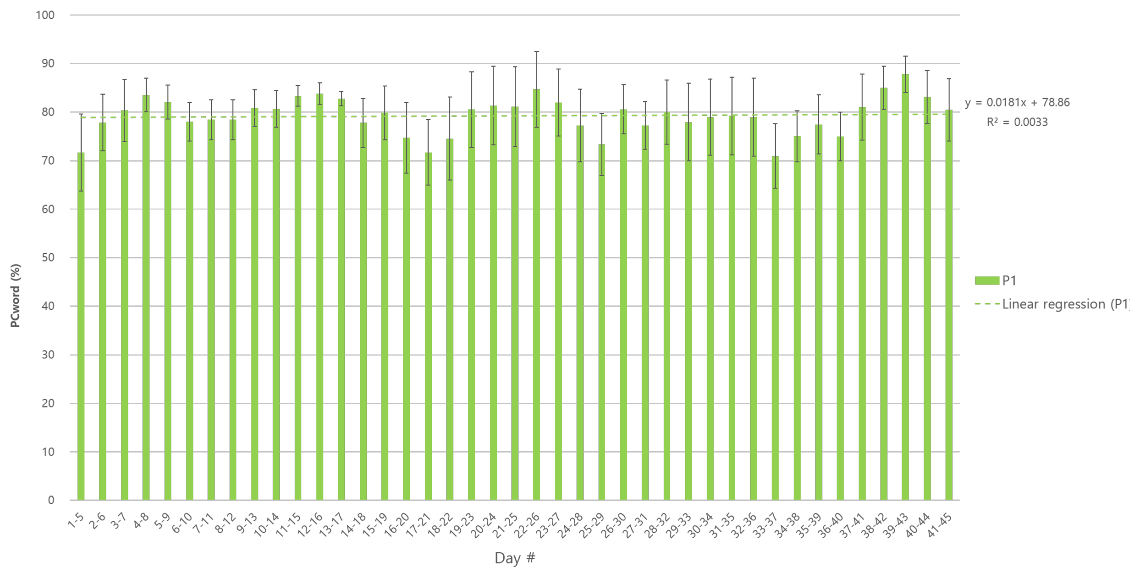


Fig. 5.9.: Moving average of P1's PC_{word} scores with a 5-day window.

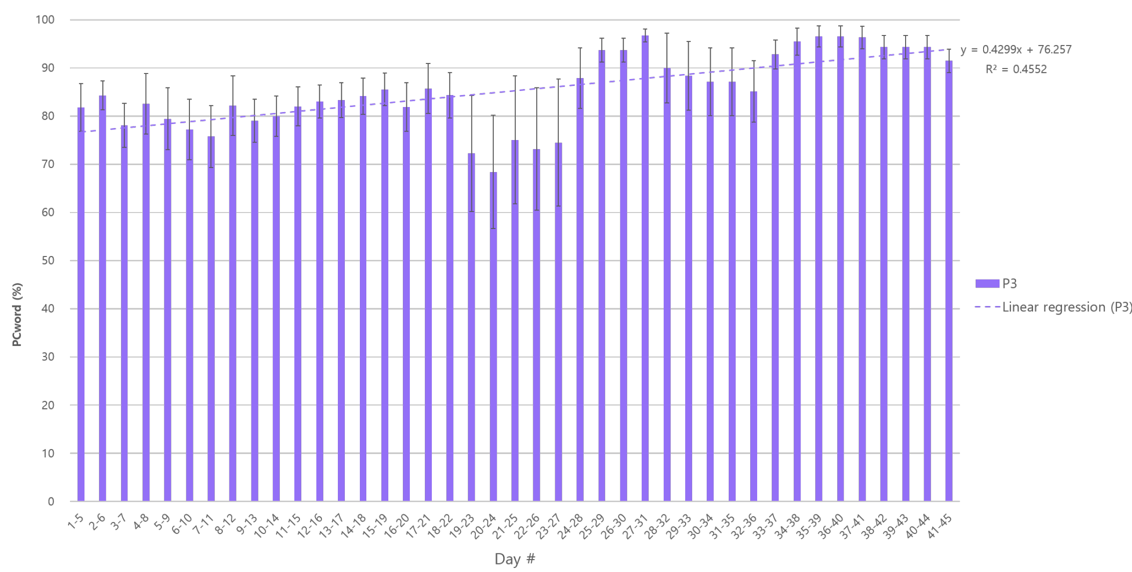


Fig. 5.10.: Moving average of P3's PC_{word} scores with a 5-day window.

calculated for the entire conversation on a day rather than for individual participants or individual messages. During the 45 days of the experiment, the minimum, maximum, and average communication rates were 0.118, 1.41, and 0.784 ± 0.312 $MsgPM$, respectively. The 5-day averages are plotted in Figure 5.12 to smooth the daily fluctuations in $MsgPM$.

It can be seen that the communication rates in messages per minute have a pattern that is strikingly similar to those of PC_{msg} and PC_{word} scores. A linear regression analysis resulted in a modest R^2 of 0.5314, and the increasing trend was significant as indicated by a t-test comparing with a slope of 0 (t(39)=6.6696, p<0.00001).

Fig. 5.11.: Daily communication rates in $MsgPM$.

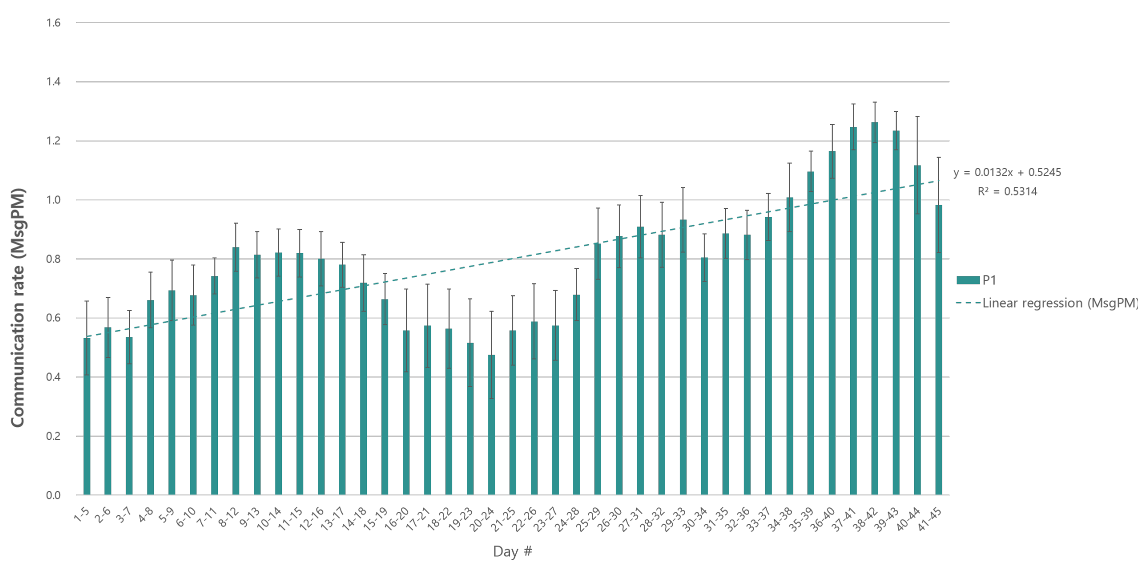


Fig. 5.12.: Moving average of the communication rates in $MsgPM$ with a 5-day window.

Curiously, the communication rates in words per minute (WPM) on a daily basis (Figure 5.13) and the corresponding 5-day running averages (Figure 5.14) show a pattern that is complementary to that of the rates in messages per minute ($MsgPM$). The WPM graphs both started at relatively low values, reached peaks in the middle of the 45-day period, and ended at intermediate levels. The minimum, maximum and average communication rates over the 45 days were 0.379 (Day 1), 3.456 (Day 22), and 1.596 ± 0.596 WPM. The 5-day running averages in Figure 5.14 do not fit a linear trend very well (R^2 = 0.067).

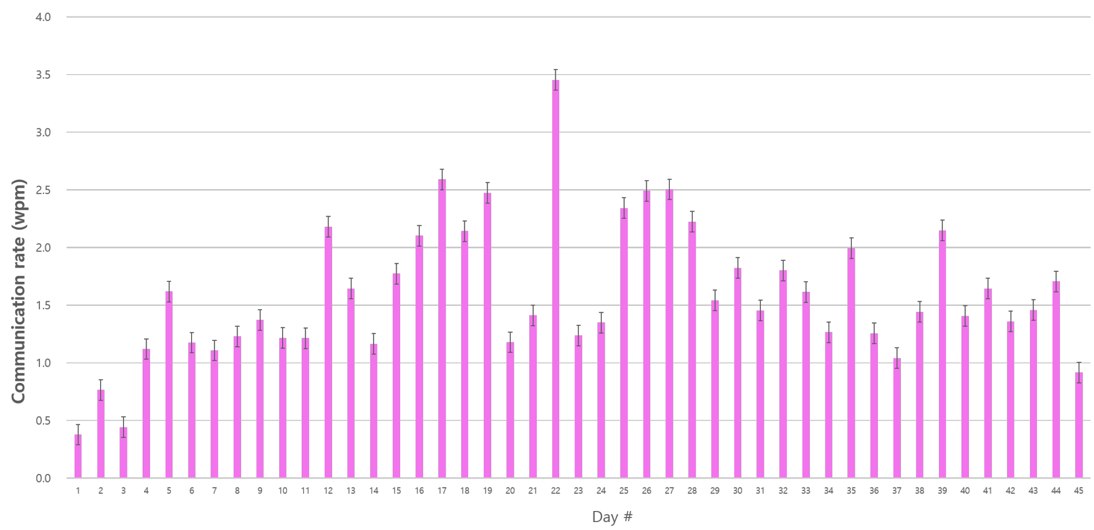


Fig. 5.13.: Daily communication rates in WPM.

To investigate the dissimilar trend between the communication rates in $MsgPM$ and WPM, an average words per message ($WPMsg$) in a dialogue as a function of experimental days is plotted as presented in Figure 5.15. Each dialogue on each day consisted of an average of 2.00±0.846 words per message. The maximum words per message of a dialogue was 3.67 WPMsg that occurred on Day 20, and the minimum $WPMsg$ was equivalent to 1 $WPMsg$ which could be observed on Day 2, Day 40, and Day 45. The fact that the maximum words per message were transmitted on Day 20 may be another reason why both participants reached a minimum in their PC_{msg} scores on that day (see Figure 5.5).

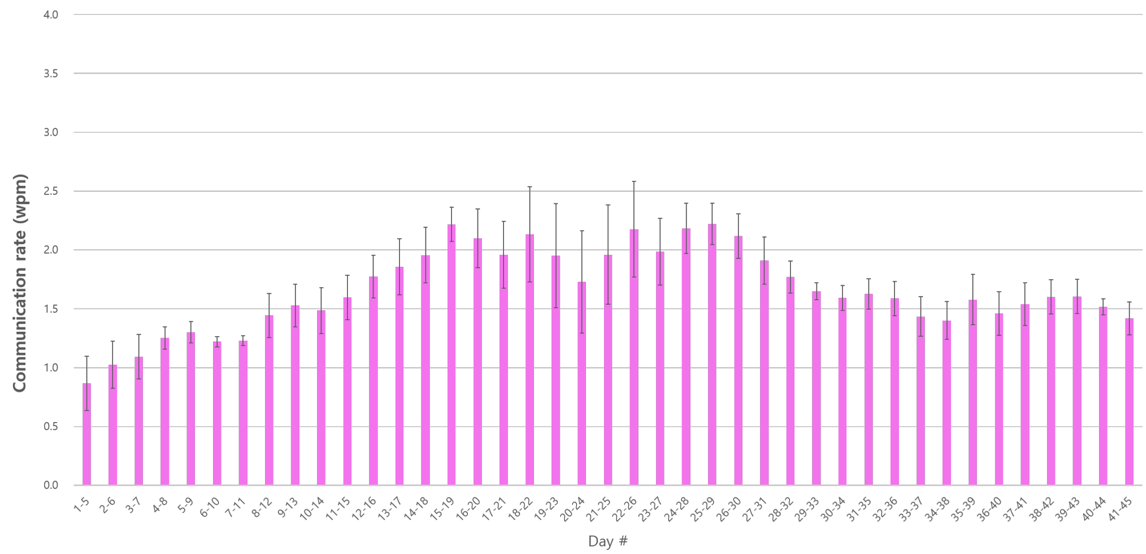


Fig. 5.14.: Moving average of the communication rates in WPM with a 5-day window.

In order to smooth the fluctuations existing in Figure 5.15, a 5-day average of words per message on each day is plotted in Figure 5.16. This plot shows a clear peak in words per message in the middle of the experimental period.

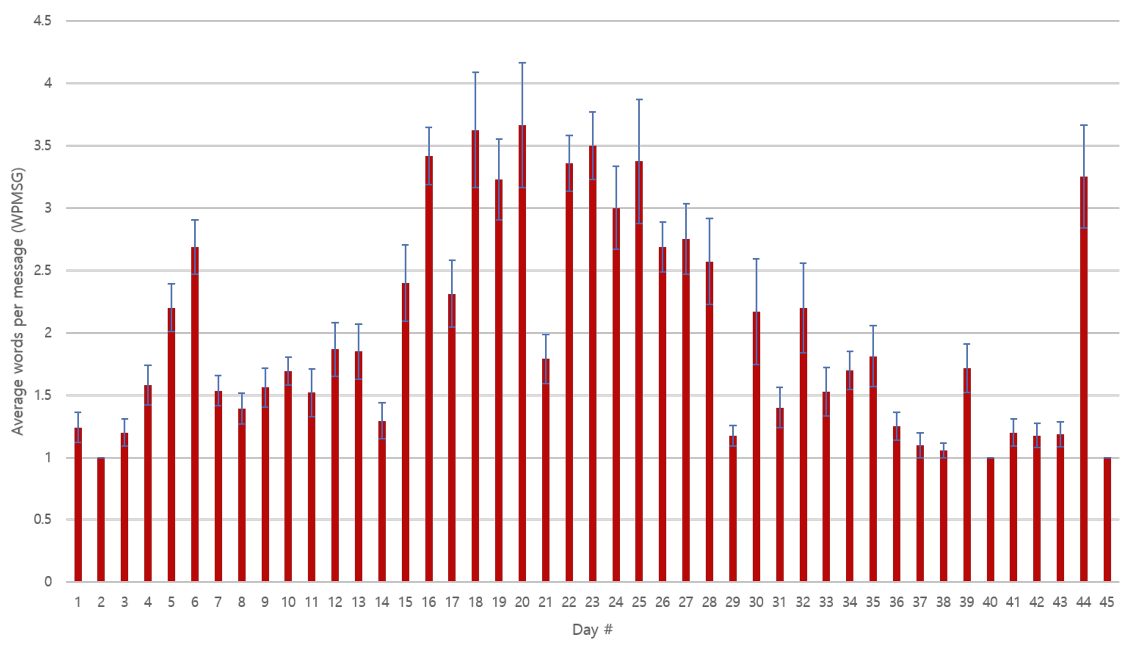


Fig. 5.15.: Average words per message on a daily basis.

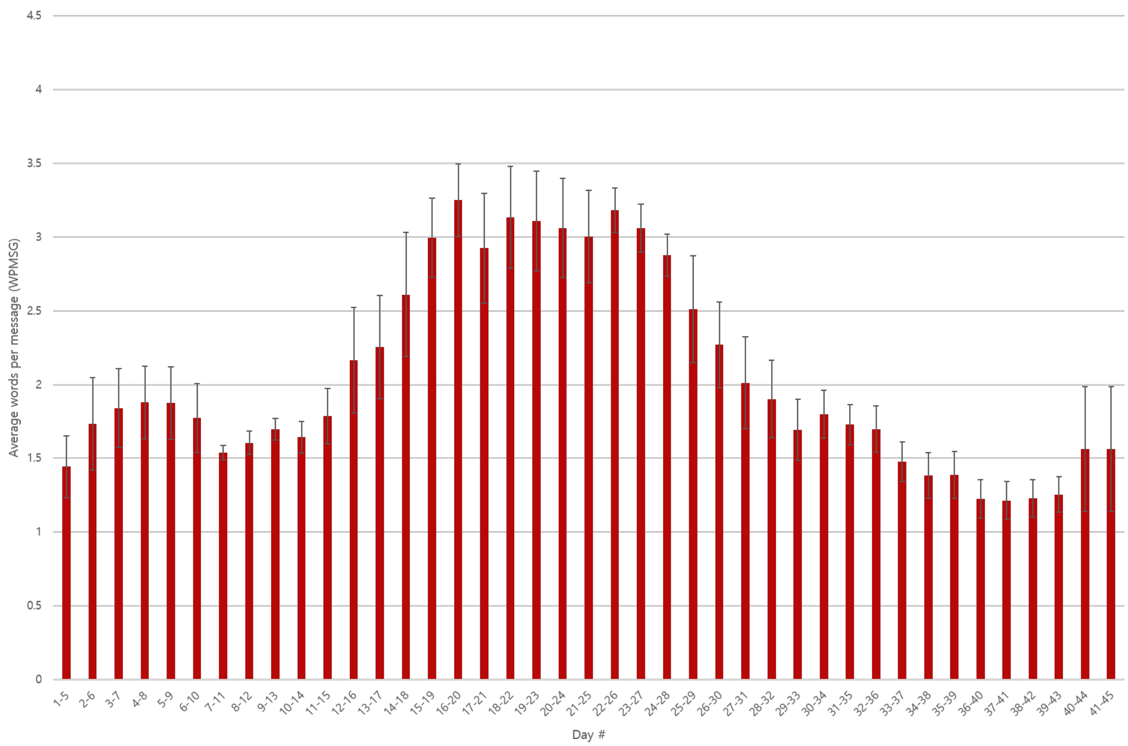


Fig. 5.16.: Moving average of words per message ($WPMsg$) with a 5-day window.

5.3 Discussion

Through the two-way tactile communication using TAPS done by the two participants, several measurements could be calculated to evaluate the feasibility of TAPS in practical use. The daily percent correct scores of identifying messages (PC_{msg}) and words in a message (PC_{word}) were calculated first. P1 acquired an average PC_{msg} score of 69.0% ± 18.9% whereas P3 obtained that of 77.8% ± 21.1%. The average PC_{word} scores for P1 and P3 were 78.2% ± 13.3% and 85.2% ± 14.4%, respectively. In addition to calculating the PC scores, the transmission rates of messages and words using TAPS were estimated. The transmission rate of messages was 0.784 ± 0.312 $MsgPM$ while word transmission rate was 1.596 ± 0.596 WPM. Regarding the PC_{msg} scores, P3 showed an improvement over the experimental period whereas P1 did not show a notable increase in PC_{msg} scores. A similar outcome appeared in the PC_{word} scores for both participants. Participant P3 demonstrated a significant linear increase in performance from spending time with TAPS as reflected in the increasing PC_{word} scores whereas P1 did not.

Compared to the expectation that both communication rates in $MsgPM$ and WPM would improve over time, the communication rates in $MsgPM$ showed an improvement over the study period (see Figure 5.12) whereas the WPM rates increased initially and then decreased after reaching peak values in the middle of the experimental period (see Figure 5.14). For example, the participants had a minimum rate of $MsgPM$ on Day 20 as seen in Figure 5.11, but the rate of WPM was similar to those on other days. To further investigate the different trends between the two communication rates, the two PC scores, PC_{msg} and PC_{word}, should also be considered since the daily $MsgPM$ and WPM rates are related to the correctly identified messages and words on each day. In Figure 5.5, both participants had minimum PC_{msg} message scores (25% for P1 and 20% for P3) on Day 20. However, the PC scores for words, PC_{word}, were much higher: 55.6% and 66.7% for P1 and P3, respectively. This could be interpreted as the participants missing most of the messages as a whole but correctly identified more words in the messages. This helps to explain why the trends of communication rates in WPM and $MsgPM$ were dissimilar. Scoring a high PC_{word} does not necessarily lead to high PC_{msg} scores. Instead, the participants might have correctly identified short messages with a small number of words but incorrectly identified longer messages that contained more words. In addition, participants might miss a single word in longer messages which could lead to a high word recognition score but a lower percent-correct score of identifying the message. Hence, this explains the difference in the trends between the two communication rates measured in $MsgPM$ and WPM.

The communication rates in terms of messages or words per minute obtained in the present study appear to be quite low, mainly due to the way that the experiment was set up. Whereas the participants of other past studies were typically instructed to respond as quickly and accurately as they could upon receiving pre-constructed stimuli, the two participants in the present study recorded their responses and took the time to think about the next message and typed it into a message box on the screen. In other words, the "conversation" during the present study was more "natural" and the messages were not pre-determined.

The time that the two participants took to record their responses and to think and type new messages penalized the calculation of communication rates in the present study.

The two-way communication performance using the TAPS between the two experienced participants in this study can be compared to the results from a previous study on TAPS by Tan et al. (2020) with both experienced and naive participants [7]. In Tan et al.'s study, 21 participants (11 experienced, including P1 and P3, and 10 naive) achieved an average word recognition score of 72.1 ± 3.5% with 500 English words. The average of P1 and P3, who were among the top-performers in Tan et al.'s study, was 91.7 ± 2.1%. The stimuli consisted of single words made up of one to five phonemes with most of the words consisting of 2 or 3 phonemes. In comparison, the word PC scores of the two experienced participants P1 and P3 in the present study were similar to the group average in Tan et al.'s study even though they dropped somewhat due to the vocabulary size expanding from a closed set of 500 words in Tan et al. (2020) to an open set of words in the present study. Considering the fact that the present study placed no restriction on the number of words in a message or the number of phonemes in a word, the task of identifying words in messages became more difficult in the present study as the chance level dropped from 0.2% (1 in 500 words) to near 0. Therefore the word PC scores achieved by P1 and P3 in the present study are still truly remarkable and demonstrates the two experienced participants' ability to receive impromptu text messages on the skin of their forearms via the TAPS system.

The results of the present study can also be compared to several other studies of delivering English words through tactile stimulation on the skin using devices other than TAPS. One of the early studies used the "vibratese language" in which 45 symbols such as English letters and numbers were encoded on five vibrators. Although the system was able to transmit 67 words per minute, it took about 12 hours for their participants to get trained for matching the signals with symbols and move forward to learning English words and short messages. It was reported that the participants could receive the "vibratese language" at a rate of 38 words per minute [69]. The transmission rate of the present study cannot be directly compared to that of the vibratese language system due to the differences in the way that the time variable was calculated, as will be explained later in this section.

Another well-known tactile aid is called the "Vocoder". The tactile vocoder takes the acoustic waveform and filters it through 16 channels. Each of them activates a vibrator on the skin of the forearm. One female participant took part in a series of experiments reported in Brooks & Frost (1983) [70], Brooks et al. (1985) [8], and Brooks et al. (1986) [2] that evaluated the use of the vocoder. In Brooks & Frost (1983), the participant took 55 hours to learn 150 words to reach a criterion of 80% correct [70]. She then spent about 20 hours in generalization training with different speakers of the same word list until she reached 80% correct again. Brooks et al. (1985) showed that the same female participant, who was now an experienced user of the vocoder (with a total of 75 hours of using the vocoder in Brooks & Frost's 1983 study), took an additional 25.5 hours to learn 100 new words and reached 76% correct [8]. The same female participant who had learned 250 words using the tactile vocoder in Brooks & Frost (1983) [70] and Brooks et al. (1985) [8] took part in another study that used a much larger set of 1000 words. The participant was able to reach an 8.8% PC score (88 out of 1000 words) for word identification after an additional 14 hours of testing [2]. Compared to these results, the two experienced participants in the present study spent less time with TAPS (about 19 hours in total for each participant) and reached an average word PC score of above 78% with an open vocabulary.

Novich et al. (2015) adopted the spectral-based approach to transmit acoustic signals to tactile stimuli through 27 actuators worn on the torso by using a device called the VEST. The participants were presented with a list of 50 words. Their task was to identify individual words they perceived through the VEST by choosing a response from four alternatives (i.e., with a chance level of 25%). After training for 12 days, the performance of word recognition ranged from 35% to 65% [3]. Based on these results, the performance with TAPS from the present study was better in terms of a higher word recognition score (81.7%) and a significantly lower chance level that was near zero.

Compared to other tactile aids, the TAPS system demonstrated impressive results. First, the study presented in this chapter showed remarkable results in terms of both PC_{msg} and PC_{word} scores. The average PC_{msg} and PC_{word} scores of the two participants were 73.4% ± 20.4% and 81.7% ± 14.2%, respectively. Considering that they used an open vocabulary for the experiment with a chance level close to zero, the PC scores were outstanding. It should be noted that most of the past studies used closed-set word lists yet produced similar results in terms of PC scores for word recognition. Furthermore, the experimental setup was also noteworthy. Unlike other past studies on tactile aids that used one-way communication to test the receiver's performance of word recognition in a well-controlled laboratory setting, the study reported here used a two-way "texting" setup that was more ecologically valid as far as daily communication between two people goes. Therefore, the results from the present study demonstrate that the TAPS system is a feasible and efficient tactile device for two-way speech communication, as prior studies using other tactile devices either achieved similar PC scores after a longer period of learning, or lower PC scores with a comparable period of using their tactile devices. It would be interesting to see if the IPI and IWI durations may be reduced with prolonged use of the TAPS system, as a previous study on TAPS has shown that the participants were able to receive haptic symbols with an IPI of 75 ms without a significant drop in word-recognition performance [14]. This would reduce the time for communicating between two people and increase the communication rates in both $MsgPM$ and WPM, thereby improving the performance with the TAPS system further.

6. CONCLUSION AND FUTURE WORK

Through three separate studies presented in Chapters 3, 4, and 5, this thesis has contributed to the evaluation of the feasibility of the TActile Phonemic Sleeve (TAPS) system for speech reception through the skin, including evidence of two-way communication.

In the first study described in Chapter 3, we have introduced a model of learning protocols using the TAPS system that converts English phonemes into haptic symbols. After a pilot study, two experiments were conducted where a total of six participants learned and were tested with phonemes and words that included different sizes of word lists by practicing for a short time on each day. The results led to two findings regarding TAPS. First, it was achievable for participants to learn and acquire phonemes within a short amount of learning time. In Experiment I, naive participants could learn 51 words made of 10 phonemes in 60 minutes. Experiment 2 showed that all 39 phonemes could be learned within 80 minutes. Secondly, evidence of memory consolidation was obtained in Experiment 2 where the participant improved in phoneme recognition after a certain interval of inactivity. Therefore, the first study demonstrated that it was feasible to learn all 39 phonemes, the basic building block of any English words, and this could be achieved by spending tens of minutes a day over a few days.

The second study presented in Chapter 4 expanded upon several studies that were conducted after the first study (i.e., [13], [6], [7]) by employing longer, four-phoneme words as the learning materials. Those studies on word recognition using TAPS were based on words that contained one to five phonemes, with a majority of the words containing two or three phonemes. After reviewing the 39 phonemes and a previously-learned 500-word list for the first 3 days, three experienced participants learned a list of 100 four-phoneme words over a period from Day 4 to Day 13, and were tested on a new set of 100 four-phoneme words for generalization of learning during the last two days of the experiment (Day 14 and Day 15). The three participants achieved an average word recognition score from 74.7%

to 87.7%. The average word recognition score for learning four-phoneme words from Day 4 to Day 13 was 80.2 ± 6.7%. For the generalization of learning, the average word recognition score was 72.3 ± 7.2% which was well above the chance level of near zero. It was demonstrated that the participants were able to learn 4-phoneme words with 170 min of practice and with word recognition scores that were well above chance levels. These results confirmed that the experienced users of TAPS could continued to learn longer English words within a relatively short amount of time as compared to other past studies.

Having ascertained that the participants were capable of learning longer words with an impressive result through the TAPS, the third study reported in Chapter 5 tested two experienced participants in their ability to freely communicate with each other by sending text messages through the TAPS for 45 non-consecutive days. During this experiment, the dialogue of each day was recorded, and the participants' PC scores and communication rates were computed and analyzed. The results show that the participants could definitely acquire words in messages with word PC scores that were significantly above the chance level of near 0%, considering that the participants could choose any English words for their text messages. The fact that an open vocabulary was used made the task much more difficult than those reported in other past studies. As far as we are aware, this was the first time that any tactile communication system has been used in two-way communication with an open vocabulary.

Future work can proceed in several directions. To facilitate the transmission of spontaneous speech, new haptic codes are needed to represent punctuation symbols in messages. Considering that our goal is to ultimately allow the TAPS system to be used in daily communication between people who may send multiple phrases or sentences in a message, it is necessary to have haptic symbols for punctuation marks that is used to separate phrases and sentences, and perhaps even haptic emojis. In some cases, the placement of a comma can alter the meaning of a sentence, so it is critical to know the punctuation marks for effective communication. The haptic symbols for punctuation marks will have to be short since one of the goals for tactile aids is to maximize the presentation rate. Furthermore, these haptic symbols should be very distinctive as compared to the 39 haptic symbols designed for the

39 phonemes. They also need to be easy to learn in order to minimally affect the learning period for all the haptic symbols used in the TAPS system. Another future direction is to start testing people with hearing impairments on their use of the TAPS system for speech reception. Many deaf people do not use oral English and use other means, such as American Sign Language, for communication. Now that we know that the TAPS system can be used for speech communication on the skin based on a phonemic coding approach, we can start working on a modified coding scheme based on the specific language requirements of people with sensory impairments. It is our ultimate goal to enable tactile communication of oral or written language using a system like the TAPS for people with all levels of sensory capabilities.

REFERENCES

REFERENCES

[1] J. M. Weisenberger and M. E. Percy, "The transmission of phoneme-level information by multichannel tactile speech perception aids." *Ear and Hearing*, vol. 16, no. 4, pp. 392–406, 1995.

[2] P. Brooks, B. Frost, J. Mason, and D. Gibson, "Continuing evaluation of the Queen's University tactile vocoder. i: Identification of open set words." *Journal of Rehabilitation Research and Development*, vol. 23, no. 1, pp. 119–128, 1986.

[3] S. D. Novich, "Sound-to-touch sensory substitution and beyond," Ph.D. dissertation, Rice University, 2015.

[4] S. Zhao, A. Israr, F. Lau, and F. Abnousi, "Coding tactile symbols for phonemic communication," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Montreal Quebec, Canada, April 2018, pp. 1–13.

[5] N. Dunkelberger, J. Sullivan, J. Bradley, N. P. Walling, I. Manickam, G. Dasarathy, A. Israr, F. W. Lau, K. Klumb, B. Knott, F. Abnousi, R. Baraniuk, and M. K. O'Malley, "Conveying language through haptics: A multi-sensory approach," in *Proceedings of the 2018 ACM International Symposium on Wearable Computers*, Singapore, Singapore, October 2018, pp. 25–32.

[6] 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, no. 1, pp. 2–17, 2019.

[7] H. Z. Tan, C. M. Reed, Y. Jiao, Z. D. Perez, E. C. Wilson, J. Jung, J. S. Martinez, and F. M. Severgnini, "Acquisition of 500 English words through a TActile Phonemic Sleeve (TAPS)," *IEEE Transactions on Haptics*, accepted, 2020.

[8] P. Brooks, B. Frost, J. Mason, and K. Chung, "Acquisition of a 250-word vocabulary through a tactile vocoder," *The Journal of the Acoustical Society of America*, vol. 77, no. 4, pp. 1576–1579, 1985.

[9] G. Plant, "Training in the use of the Tactaid VII: A case study," *KTH Computer Science and Communication (STL-QPSR)*, vol. 35, no. 1, pp. 91–102, 1994.

[10] R. M. Uchanski and L. D. Braida, "Effects of token variability on our ability to distinguish between vowels," *Perception & Psychophysics*, vol. 60, no. 4, pp. 533–543, 1998.

[11] G. Luzhnica, E. Veas, and V. Pammer, "Skin reading: Encoding text in a 6-channel haptic display," in *Proceedings of the 2016 ACM International Symposium on Wearable Computers*, Heidelberg, Germany, September 2016, pp. 148–155.

[12] 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 *International Conference of Design, User Experience, and Usability*, Las Vegas, NV, USA, July 2018, pp. 447–460.

[13] 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*, Pisa, Italy, June 2018, pp. 623–635.

[14] C. M. Reed, H. Z. Tan, Y. Jiao, Z. D. Perez, and E. C. Wilson, “Identification of words and phrases through a phonemic-based haptic display: Effects of inter-phoneme and inter-word interval durations,” *ACM Transactions on Applied Perception*, p. in review, 2020.

[15] M. A. Srinivasan, “What is haptics?” *Laboratory for Human and Machine Haptics: The Touch Lab, Massachusetts Institute of Technology*, pp. 1–11, 1995.

[16] J. M. Wolfe, K. R. Kluender, D. M. Levi, L. M. Bartoshuk, R. S. Herz, R. L. Klatzky, S. J. Lederman, and D. Merfeld, *Sensation & Perception*. Sinauer Sunderland, MA, 2006.

[17] 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.

[18] K. O. Johnson, “The roles and functions of cutaneous mechanoreceptors,” *Current Opinion in Neurobiology*, vol. 11, no. 4, pp. 455–461, 2001.

[19] R. S. Johansson and J. R. Flanagan, “Coding and use of tactile signals from the fingertips in object manipulation tasks,” *Nature Reviews Neuroscience*, vol. 10, no. 5, pp. 345–359, 2009.

[20] K. O. Johnson, T. Yoshioka, and F. Vega-Bermudez, “Tactile functions of mechanoreceptive afferents innervating the hand,” *Journal of Clinical Neurophysiology*, vol. 17, no. 6, pp. 539–558, 2000.

[21] 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.

[22] 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.

[23] 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, pp. 1–36.

[24] R. Verrillo, “Temporal summation in vibrotactile sensitivity,” *The Journal of the Acoustical Society of America*, vol. 37, no. 5, pp. 843–846, 1965.

[25] S. Weinstein, “Intensive and extensive aspects of tactile sensitivity as a function of body part, sex and laterality,” *The skin senses, Ed.D.R. Kenshale*, pp. 195–222, 1968.

[26] M. R. Longo and R. Sadibolova, “Seeing the body distorts tactile size perception,” *Cognition*, vol. 126, no. 3, pp. 475–481, 2013.

[27] B. G. Green, "The perception of distance and location for dual tactile pressures," *Perception & Psychophysics*, vol. 31, no. 4, pp. 315–323, 1982.

[28] T. Wong, R. Ho, and J. Ho, "Influence of shape of receptor organ on the horizontal-vertical illusion in passive touch." *Journal of Experimental Psychology*, vol. 103, no. 3, p. 414, 1974.

[29] M. R. Longo and P. Haggard, "Weber's illusion and body shape: anisotropy of tactile size perception on the hand." *Journal of Experimental Psychology: Human Perception and Performance*, vol. 37, no. 3, p. 720, 2011.

[30] B. Whitsel, O. Franzen, D. Dreyer, M. Hollins, M. Young, G. Essick, and C. Wong, "Dependence of subjective traverse length on velocity of moving tactile stimuli," *Somatosensory Research*, vol. 3, no. 3, pp. 185–196, 1986.

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

[32] S. Grondin, T. Kuroda, and T. Mitsudo, "Spatial effects on tactile duration categorization." *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, vol. 65, no. 3, p. 163, 2011.

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

[34] 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*, Vancouver, BC, Canada, May 2011, pp. 2019–2028.

[35] S. Lakatos and R. N. Shepard, "Constraints common to apparent motion in visual, tactile, and auditory space." *Journal of Experimental Psychology: Human Perception and Performance*, vol. 23, no. 4, p. 1050, 1997.

[36] J. H. Kirman, "The effect of number of stimulators on the optimal interstimulus onset interval in tactile apparent movement," *Perception & Psychophysics*, vol. 17, no. 3, pp. 263–267, 1975.

[37] 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 *Proceedings of the 2009 IEEE International Workshop on Haptic Audio Visual Environments and Games*, Lecco, Italy, November 2009, pp. 19–23.

[38] E. P. Gardner and W. A. Spencer, "Sensory funneling. i. psychophysical observations of human subjects and responses of cutaneous mechanoreceptive afferents in the cat to patterned skin stimuli." *Journal of Neurophysiology*, vol. 35, no. 6, pp. 925–953, 1972.

[39] F. A. Geldard and C. E. Sherrick, "The cutaneous" rabbit": A perceptual illusion," *Science*, vol. 178, no. 4057, pp. 178–179, 1972.

[40] H. Z. Tan, R. Gray, J. J. Young, and R. Traylor, "A haptic back display for attentional and directional cueing," *Haptics-e: The Electronic Journal of Haptics Research*, vol. 3, no. 1, (20 pages), June 2003.

[41] G. A. Gescheider, *Psychophysics: Method, Theory, and Application*. Lawrence Erlbaum, 1985.

[42] R. Yunus, S. Ali, Y. Ayaz, M. Khan, S. Kanwal, U. Akhlaque, and R. Nawaz, "Development and testing of a wearable vibrotactile haptic feedback system for proprioceptive rehabilitation," *IEEE Access*, vol. 8, pp. 35 172–35 184, 2020, doi:10.1109/ACCESS.2020.2975149.

[43] K. Bark, J. Wheeler, G. Lee, J. Savall, and M. Cutkosky, "A wearable skin stretch device for haptic feedback," in *Proceedings of the World Haptics 2009-Third Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*. Salt Lake City, UT, USA: IEEE, March 2009, pp. 464–469.

[44] 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, 2012.

[45] H. Levitt, "Transformed up-down methods in psychoacoustics," *The Journal of the Acoustical Society of America*, vol. 49, pp. 467–477, 1971.

[46] H. Robbins and S. Monro, "A stochastic approximation method," *The Annals of Mathematical Statistics*, vol. 22, pp. 400–407, 1951.

[47] 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.

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

[49] S. Ricco and C. Tomasi, "Fingerspelling recognition through classification of letter-to-letter transitions," in *Proceedings of the 2009 Asian Conference on Computer Vision*. Springer, 2009, pp. 214–225.

[50] 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.

[51] 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.

[52] C. M. Reed and N. I. Durlach, "Note on information transfer rates in human communication," *Presence*, vol. 7, no. 5, pp. 509–518, 1998.

[53] C. M. Reed, N. I. Durlach, and L. D. Braida, "Research on tactile communication of speech: a review." *American Speech-Language-Hearing Association Monographs*, no. 20, 1982.

[54] C. M. Reed, N. I. Durlach, L. A. Delhorne, W. M. Rabinowitz, and K. W. Grant, "Research on tactual communication of speech: Ideas, issues, and findings." *The Volta Review*, 1989.

[55] J. H. Kirman, "Tactile communication of speech: A review and an analysis." *Psychological Bulletin*, vol. 80, no. 1, p. 54, 1973.

[56] H. Yuan, C. M. Reed, and N. I. Durlach, "Tactual display of consonant voicing as a supplement to lipreading," *The Journal of the Acoustical Society of America*, vol. 118, no. 2, pp. 1003–1015, 2005.

[57] A. Israr, C. M. Reed, and H. Z. Tan, "Discrimination of vowels with a multi-finger tactual display," in *Proceedings of the 2008 Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, Reno, NV, USA, March 2008, pp. 17–24.

[58] S. D. Novich and D. M. Eagleman, "Using space and time to encode vibrotactile information: toward an estimate of the skins achievable throughput," *Experimental Brain Research*, vol. 233, no. 10, pp. 2777–2788, 2015.

[59] 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.

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

[61] D. H. Ecroyd, M. M. Halfond, and C. C. Towne, *Voice and Articulation: A Handbook*. Scott, Foresman & Company, 1966.

[62] L. B. William and N. Harter, "Studies on the telegraphic language: The acquisition of a hierarchy of habits." *Psychological Review*, vol. 6, no. 4, p. 345, 1899.

[63] A. Boothroyd, L. Hanin, and T. Hnath, *A sentence test of speech perception: Reliability, set-equivalence, and short-term learning*. City University of New York, Speech and Hearing Sciences Research Center, RCI10, 1985.

[64] K. L. Galvin, P. J. Blamey, R. S. Cowan, M. Oerlemans, and G. M. Clark, "Generalization of tactile perceptual skills to new context following tactile-alone word recognition training with the tickle talker¢," *The Journal of the Acoustical Society of America*, vol. 108, no. 6, pp. 2969–2979, 2000.

[65] A. W. Black and K. Lenzo, "Flite: a small, fast speech synthesis engine," 2005.

[66] S. Shim and H. Z. Tan, "Palmscape: Calm and pleasant vibrotactile signals," in *Design, User Experience, and Usability. Interaction Design*. Springer, 2020, pp. ch. 37, pp.1–17.

[67] R. L. Sherbecoe and G. A. Studebaker, "Supplementary formulas and tables for calculating and interconverting speech recognition scores in transformed arcsine units," *International Journal of Audiology*, vol. 43, no. 8, pp. 442–448, 2004.

[68] G. A. Studebaker, "A "rationalized" arcsine transform," *Journal of Speech, Language, and Hearing Research*, vol. 28, no. 3, pp. 455–462, 1985.

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

[70] P. Brooks and B. J. Frost, "Evaluation of a tactile vocoder for word recognition," *The Journal of the Acoustical Society of America*, vol. 74, no. 1, pp. 34–39, 1983.

PUBLICATIONS

PUBLICATIONS

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 – International Conference of Design, User Experience, and Usability,* Springer, 2018, pp. 447-460.

Y. Jiao, F. M. Severgnini, J. S. Martinez, **J. Jung**, H. Z. Tan, C. M. Reed, E. C. Wilson, F. Lau, A. Israr, and 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*, 2018, pp. 623-635.

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, no. 1, pp. 2-17, 2018.

H. Z. Tan, C. M. Reed, Y. Jiao, Z. D. Perez, E. C. Wilson, **J. Jung**, J. S. Martinez, and F. M. Severgnini, "Acquisition of 500 English words through a TActile Phonemic Sleeve (TAPS)," *IEEE Transactions on Haptics*, to appear, 2020.