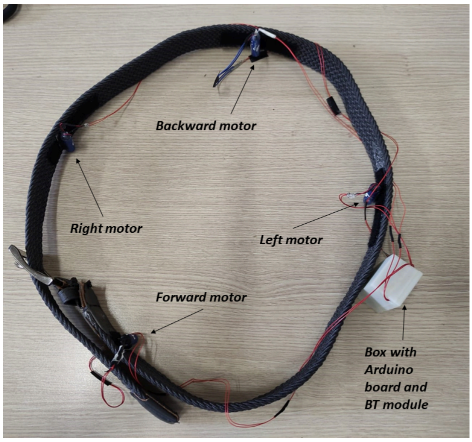


Figure 3.2: Haptic belt prototype with vibratory motors.

rather than being informed about the optimal direction to follow. Vibrations provided by haptic bands were used to simulate contact with people in a crowd during a VR exploration task [111]. The experimental results highlighted behavioral changes due to the increased realism perceived by the participants when the haptic stimulation was active, and the aftereffect of the haptic stimulation.

This work targets and investigates a general use-case, *i.e.*, a formation of two people carrying a bulky object by hand, to compare the effectiveness of the proposed haptic policies in communicating and displaying informative contents for cooperation purposes. The design of the three conditions was based on the requirement of intuitiveness of the haptic cues, and leveraged three different assumptions: in the first case the human locomotion was modeled as holonomic, in the second case the human walking was modeled with a nonholonomic behavior, while in the third condition the haptic patterns displayed the direction to the target area, but the members of the formation were in charge of deciding the strategy to reach it. The haptic policies were tested in an experimental campaign (Fig. 3.1) to measure the performance under different conditions in terms of quantitative and subjective measures.

3.1.2 System overview

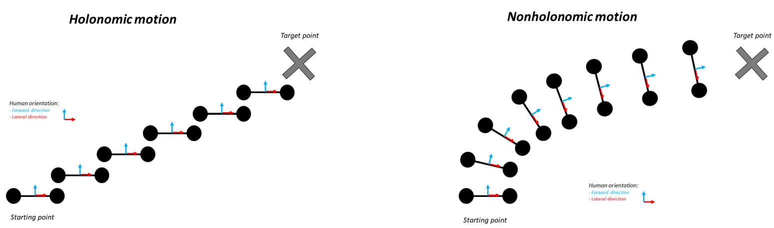
In this subsection, we describe the entire system, detailing the theoretical background (models of human walking and proposed guidance approaches), the exploited hardware, and the developed software.

Wearable haptic device

The haptic device developed in this work consists of a belt equipped with four vibratory motors, an Arduino Pro-Mini with a Blueooth module, and a Li-Po battery. The electronic components are contained in a 3D printed case (Fig. 3.2). The vibro-motors are controlled by applying a certain amount of voltage which determines both frequency and amplitude. As the human maximal sensitivity to vibrations is achieved around 200-300 Hz, the motors (Precision Microdrives 9mm Vibration Motor) have a vibration frequency range of 100-300Hz, amplitude between 3 and 9 g, lag time of about 20ms, rise and stop time of 35ms.

The four motors, placed at the cardinal points of the belt, can be activated to display vibratory patterns, *e.g.* rotations or directions on the horizontal plane. Directions that do not lay on the frontal or lateral axis are displayed as a combination of two vibrations from adjacent actuators, whose intensity depends on the component of the desired stimulus displayed by each motor. The choice of using four actuators only is to design a lightweight and easy-to-use device. Although a larger number of motors would result in better spatial resolution for displaying directions, the small distance between the actuators would make it difficult to clearly distinguish adjacent sources [64].

Haptic Guidance Policies



(a) (b)

Figure 3.3: Sketches of holonomic (a) and nonholonomic motion (b).

The aim of this work is investigating the most suitable strategy to support the users during the task, *i.e.*, assessing whether guiding the formation through a sequence of actions is more effective than augmenting human perception though the haptic channel. The guidance of human locomotion is tackled by using an approach that is typical for robotics. In most cases the human walking can be considered as nonholonomic (see Fig. 3.3). Indeed, humans usually walk in the forward feet direction and avoid lateral translations. As a consequence, the body orientation, and particularly the shoulders, follow the trajectory travelled in a very precise way [112]. The only exception to the nonholonomic nature in human walking takes place when the point to be reached is close to the human. In this case, it is more convenient to perform lateral or diagonal steps instead of rotating and then taking few steps forward [113]. In the specific context of a formation of humans moving a bulky object, it has to be taken into account that the arrangement of the formation due to the object geometry may force some users to walk backwards or sideward. According to holonomic and nonholonomic models two guidance policies were designed, which leveraged patterns of tactile stimuli to provide instructions.

A further experimental condition investigated the possibility of improving the task execution by augmenting the human's perceptions through haptics. Instead of instructing the users to perform specific motions, the haptic interface provides the users with necessary information about the task objective, which are then elaborated and exploited by the members of the formation relying on their own experience. Although the haptic patterns used are similar to the guidance conditions, in this case the paradigm of interpretation of the stimuli is notably different, as it is based on a sensory augmentation though haptics.

Given that during the experimental trials the users were blindfolded and were asked not to communicate verbally, the haptic feedback given by the exchange of mutual forces mediated by the object represents their only communication channel. Although they cannot exchange proper conversation, it allows the two users to perceive the partner's motions and intentions, and then to perform a synchronized action.

Based on these assumptions, the three haptic guidance policies are presented as follows:

Holonomic Guidance (H): Motors continuously vibrate indicating the direction to the goal. The user is instructed to perform a translation in the suggested direction. No rotation is allowed. Vibration stops for 1 s when target is reached.

Nonholonomic Guidance (N): Users are provided with either of the two different haptic pattern representing the following instructions:
i) Rotate: motors vibrate alternately clockwise or counterclockwise. The formation has to rotate in the suggested direction. The haptic pattern is active until the frontal direction (of the formation) is close to the target direction, with a $\pm 20°$ tolerance.
ii) Translate: motors vibrate to indicate the direction to the target. The user is instructed to perform a translation in the suggested direction. Vibration stops for 1 s when target is reached.

Sensory Augmentation (SA): Motors continuously vibrate to indicate the direction to the goal. The user is instructed to choose the strategy to reach the target. Vibration stops for 1 s when target is reached. Communication with the partner is mediated by the tactile channel, as participants can exchange forces through the handheld object.

The *Holonomic* condition is based on the assumption that the formation is able to travel goal-by-goal maintaining a fixed orientation, that can be beneficial when transporting objects that require stability (*e.g.* objects that are fragile or are difficult to balance). Although humans do not naturally take lateral or backward steps while walking, in the specific case of a formation carrying an object, humans exhibit holonomic features for complying to the task achievement.

In the *Nonholonomic* assumption, the users can translate only forward or backward, avoiding lateral steps. Moreover, they are allowed to steer the direction during walk in a nonholonomic way, according to the haptic instructions.

The design of the *Sensory Augmentation* condition took into account the fact that humans are not as precise as robots when it comes to perform a motion, neither they need accurate step-by-step instructions to execute a complex task. Instead, they are able to unconsciously select the most efficient and natural set of actions to reach a goal, thanks to their experience and innate motion planning capacity.

The haptic pattern used to convey rotation instructions (used in the *Nonholonomic* condition) was selected in a user-centred preliminary experiment. Three different haptic patterns were presented to 10 naive users, that were asked to select the most intuitive pattern to convey the rotation direction. The participants had to take into account the requirement of easily distinguishing the rotation instruction from the translation instruction, described above. Moreover, the angle threshold used to switch between rotation and translation instructions in the *Nonholonomic* condition was selected according to natural walking strategies adopted by 10 participants in preliminary tests.

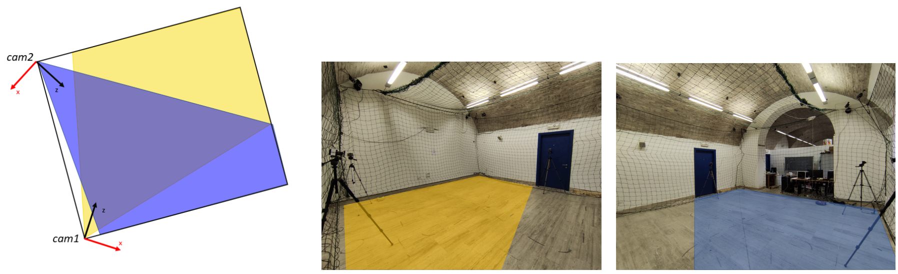
Grasping theory to control the formation

The problem of keeping the formation during the task has been addressed by means of the grasping theory. Each element of the formation is modeled as a contact point under the assumption that internal forces are neglected, as each user is in a configuration of power grasp to firmly hold the object. The grasping matrix, that relies on measures related to contact points, can be computed by using geometric parameters about the formation. Through the grasping matrix the software computes the forces to move the object toward the target location for each contact point, providing the directions of motion for each user. This solution allows to generalize the algorithm to any number of users. A detailed explanation on the grasping approach used in this work is reported as follows. Consider a rigid object, denoted by $\mathcal{O}$, and a group of N users, denoted by U_1 ...U_N. The configuration of $\mathcal{O}$ is defined by the position $p_0 \in \mathbf{R}^3$ of the center of mass of the object and by the rotation matrix $R_0 \in SO(3)$ defining the orientation. We assume that a tracking system records in real-time the motion of the object. In this framework, we consider the N users acting as N fingertips of a robotic hand that performs a grasp. The contact between the user and the object is modeled as a contact point with friction, also referred to as soft finger contact model [114]. According to this model, forces and torsional moments can be exchanged at the contact point. Furthermore, the contact force at the tooltip are denoted with $\lambda_i \in \mathbf{R}^3$. Denoting by $w \in \mathbf{R}^6$ the resulting wrench applied to $\mathcal{O}$ by $\lambda_1 ... \lambda_N$, the relation between w and $\lambda_1 ... \lambda_N$ depends on the geometry of the contact points and is given by $w = G\lambda$, where $\lambda = (\lambda_1^T ... \lambda_N^T)^T \in \mathbf{R}^{3N}$ is the stacked vector of contact forces and $G \in \mathbf{R}^{6\times 3N}$ is the grasp matrix. Solving the equation for the contact forces, it is possible to distinguish two contributions. The first, called internal forces (λ_{int}), does not contribute to the motion of the object and can be obtained through the null-space projector, while the second component of λ is defined as $\lambda_{ext} = \lambda - \lambda_{int}$ is the part of the force that produces the motion of the object. It has been proven in [115] that the choice of the internal forces λ_{int} and the control of the object motion via λ_{ext} can be considered independently in the control design. Thus, through our guidance approaches we move the object by applying external forces exploiting the human motion.

The direction of the contact force estimated for each operator is used to calculate the next instruction to reach the target, according to the current haptic policy. While in the H and SA conditions the direction of the force is equal to the direction suggested to the user, in the N condition the instruction is selected depending on the angle between the formation frontal direction and the force direction. If that angle is larger than $20°$ the rotation pattern is displayed, otherwise the translation pattern provides the users with the direction to the target.

Smartphone-based tracking system

In order to test and compare the effectiveness of the three haptic guidance policies, the participants and the carried object have to be located inside the experimental workspace. To track the users and the carried object, we leveraged RGB cameras and fiducial markers. We selected a tracking approach that is versatile, low-cost, and easy to put into practice in any environment. Indeed, this tracking system can be easily implemented using common equipment, such as smartphones (or webcams) for frame capture and streaming, ArUco fiducial markers, and a central processing unit that extracts meaningful data from the video frames to guide



(a) (b) (c)

Figure 3.4: In a) the map of the workspace, in b) and c) view from first and second camera, respectively.

the users. The *OpenCV* [116] library was used to estimate the orientation and position of each marker detected in the camera frames. In order to be uniquely identified, each element of the formation was identified with a different ArUco marker. The custom tracking system was selected to test whether the human collaboration task could be accomplished by using a simpler and more flexible (and less precise) setup than Vicon Tracking System and similar alternatives (Optitrack, *etc.*).

In the experimental campaign we exploited a Huawei P20-Lite and Huawei P10-Lite as RGB video sources. The video frames were captured and streamed to the central processing unit using an *ad-hoc* application. The framerate used was 10fps. Even if the frequency is relatively low, it was sufficient for the purpose. Indeed, the average frequency of human walking at comfortable condition is 1 Hz, that results in almost 10 updates between two consecutive steps. The smartphones were placed at two adjacent corners of a 5x5m room that was equipped with a Vicon Tracking System (consisting in 10 Bonita cameras), used as a ground truth to evaluate the capabilities of the camera-based tracking. The use of two cameras was enough to cover the entire workspace, as shown in Fig. 3.4. Two markers were placed on opposite faces of the object to overcome the problem of the occlusion. The camera arrangement and the object marker redundancy ensure that the object is always detected by at least one camera.

The central processing unit analyses the video frames to detect the visible markers and estimates their position and orientation, then it updates the grasping matrix, computes the forces to move the object toward the target, and sends haptic stimuli to the operators according to the haptic policy under test. The process is handled by a software based on the Python libraries *OpenCV* [116] and *ArUco* [117].

The main features of the developed software are:

- Estimating intrinsic and extrinsic parameters: The camera calibration procedure estimates the distortion coefficients and the camera matrix of each of the two cameras exploited in this project. Moreover, the relative orientation and translation between the two cameras is measured.
- Estimating the position and orientation of the formation in the workspace. The workspace is defined according to the reference frame of the main smartphone camera (indicated as *cam1*). Thus, in order to merge the tracking data coming from the secondary smartphone camera (indicated as *cam2*), data go through a change of reference system. The rotation and translation between cam1 and cam2 are known *a priori* from the calibration procedure. For any marker not detected in the current frame by the two cameras, position and orientation are reconstructed using the previous values of relative translation and orientation with the object. This assumption is due to the fact that the operators are engaged in a power grasp with the object, so their placement in the formation is fixed.
- Computing the haptic signals to inform the users about the direction to follow. Exploiting the theory of grasp, each user is considered a contact point with the object, thus forces to be applied are calculated to move the object to the target location while maintaining the grasp.

3.1.3 Experimental validation

The experimental validation was aimed at assessing the intuitiveness and efficacy of the three guidance policies described in Subsect. 3.1.2

The approach adopted can be extended to handle a variable number of subjects that carry together a bulky object. In this work we validated the approach with 2 users. For the sake of simplicity, and without loss of generality, we assume two people walking with a relative body rotation of 180°, so that the nonholonomic condition can be held by having one user walking forward, and the other backward. In case the formation is composed by more than two users, the nonholonomic condition still holds if the relative rotation between all the users is either 0° or 180°. This topology can be applied by having the participants holding the object on two opposite sides.

The experiments were designed to evaluate the task execution performance in the three conditions. Additionally, participants' opinions and subjective perceptions were collected using a post-experiment questionnaire to retrieve valuable information about usability and comfort of the proposed approaches. During each trial, two blindfolded participants were asked to hold the bulky object and walk along a predefined set of goals while being guided by the haptic cues.

Each subject gave her/his written informed consent to participate to the experiments and was able to discontinue participation at any time during experiments. The experimental evaluation protocols followed the declaration of Helsinki.

Ten participants (5 males, 5 females, age 25-55) took part to this experiment. Two cameras were placed on tripods and their mutual distance and orientation were computed through the static calibration procedure. Two markers were attached to the bulky object used in the task in such a way that one of them was always visible by at least one camera. The object dimensions were $100 \times 60 \times 90$ cm. The distance and orientation between the markers were measured and considered *a priori* knowledge. The users were asked to wear the haptic belt and attach the ArUco marker on their back. To test the three policies, we randomly generated 4 paths, consisting of a starting point, an end point, and three or four breakpoints. The available workspace was a $25\,\mathrm{m}^2$ room. The threshold beyond which the goal was considered reached was chosen after 12 pilot experiments. Starting from $0.2m$ up to $0.7m$, different values have been tested. Taking into account the accuracy of the camera tracking, the error due to the interpretation of the tactile cues by the humans, and the need to move as close as possible to the virtual target, a threshold of $0.5m$ was selected.

In the main campaign, the trial execution consisted in the following steps:

i) The trial started with an initialization phase to register the relative distance and orientation between the two participants and the object (necessary for the calculation of the grasping matrix). The relative rotations and translations were stored in a file for post-processing analysis. The purpose of the initialization phase was to record the topology of the formation before starting the real trial, and thus to initialize the main parameters for the grasping matrix, *i.e.*, the translation vector between the object and each user, and their relative rotation matrices.

ii) After the initialization phase, the software iteratively computed position and orientation of the object and users, and the proper instructions to guide the formation toward the target. The instructions were continuously transmitted to the users through the haptic interfaces. Whenever a target point was reached (within the $0.5m$ threshold), the motors stopped vibrating for $1s$ to inform the users, then displayed haptic cues toward the next goal.

iii) Once all the target points were reached, the trial was over. Position and orientation of the object with respect to the world reference system, computed by the OpenCV software and tracked by the *Vicon* system during the main phase of the trial, were stored into a file for post-processing analysis.

The aforementioned procedure was repeated by each couple of participants for the four trajectories and the three type of haptic guidance (H, N, SA conditions), resulting in 12 experiments per couple. Subjects were paired randomly in order to avoid any bias.

The experiment started with a 5 minutes training phase for the users to get familiar with the system, then performed the 12 trials in pseudo-random order to avoid learning effects. The haptic policy was communicated before each trial, while the trajectory was random. At the end of the experiment, they were asked to fill a questionnaire, reported as follows, to evaluate their experience and to provide helpful feedback. The items refer to the haptic policy under examination:

- EASE OF USE
  1) It is intuitive (can be used without written instructions)
  2) Using it is effortless
  3) I dont notice any inconsistency as I use it
  4) I can recover from mistakes quickly and easily
- EASE OF LEARNING
  5) It was easy to learn how to use it
  6) I easily remember how to use it
  7) It is easy to learn to use it
  8) I quickly became skillful with it
- COMFORT
  9) I felt comfortable in following the instructions
  10) I felt that the instructions allowed me to synchronize my movements with the partners
  11) I felt like the instructions were suggesting movements I would have done without
  12) I felt the instructions suggested natural movements

The questionnaire was aimed to evaluate qualitative aspects of the experiment: *Ease of use*, *Ease of learning* and *Comfort* regarding the three conditions. Items concerning *Ease of use* and *Ease of learning* derive from the *USE* questionnaire [73]. The items on *Comfort*, instead, have been designed to evaluate the movements naturalness. Participants' ratings were registered through a 7-points Likert scale, where 1 indicates "strongly disagree" and 7 "strongly agree". We report the reliability index calculated using the Cronbachs alpha for each questionnaire subsection: $\alpha = 0.77$ for the *Ease of Use*, $\alpha = 0.78$ for the *Ease of Learning*, and $\alpha = 0.78$ for *Comfort*.

3.1.4 Results

Comparison between Vicon and Camera tracking

In order to evaluate the accuracy of the tracking system based on the OpenCV library, the Vicon tracking system was considered as a ground truth. For each trial, the coordinates of the object position recorded by the Vicon system (indicated as P_{Vicon}) were compared with the coordinates recorded by the Camera-based tracking system (indicated as P_{cam}). To compute the average error, Vicon data were mapped in the Camera-based reference frame through the roto-translation of the trajectory points. The centroids of the two sets of points (C_{Vicon} and C_{cam}) were calculated and subtracted from the sets, so that points were centered around the origin:

Then, the covariance matrix was calculated as the product of the sets of points after the subtraction of the centroid:

$$H = (P_{vicon} - C_{vicon})(P_{cam} - C_{cam})^T \in \mathbb{R}^{2\times 2}$$

By applying the Singular Value Decomposition to the covariance matrix, it is possible to obtain the rotation matrix, and then reconstruct the translation vector between the two reference system. The Singular Value Decomposition expresses the original matrix H as the matrix product of three matrices: $H = U\Sigma V^T$. In this case, being $H \in \mathbb{R}^{2\times 2}$, also the three matrices obtained $U, \Sigma, V \in \mathbb{R}^{2\times 2}$.

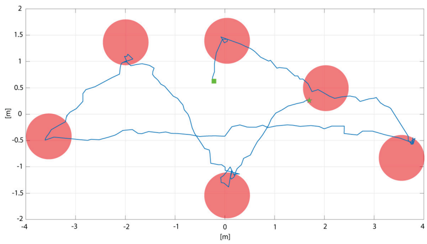
The rotation matrix between the two point clouds can be obtained from the decomposition of the covariance matrix: ${}^{Vicon}R_{cam} = VU^T$. Then the translation vector T can be obtained as:

$$T = C_{vicon} - ({}^{Vicon}R_{cam})C_{cam}$$

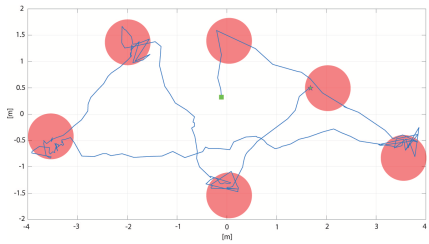
After computing the translation vector and the relative position between the two reference frames, the Vicon data can be represented in the Camera-based system reference frame to compare the tracking error. The average tracking error of the Camera-based system, estimated within the entire set of trials, was $12.7 \pm 8.3cm$.

Haptic Guidance Policies

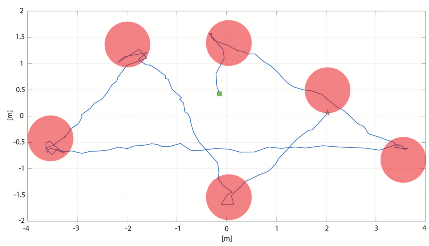
Smoothness of the trajectory, trial execution time, and questionnaires about intuitiveness of the haptic cues and comfort in movements have been taken into account in the comparison of the haptic strategies. The alpha-level adopted for the statistical analyses is $\alpha = 0.05$.



(a) Holonomic Policy



(b) Nonholonomic Policy



(c) Sensory Augmentation Policy

Figure 3.5: Representative user's trajectories for the first path exploiting the *Holonomic*, *Nonholomic* and *Sensory Augmentation* policies are reported in a), b), and c) respectively. Starting points are represented with a green square, while end points are marked with a green star.

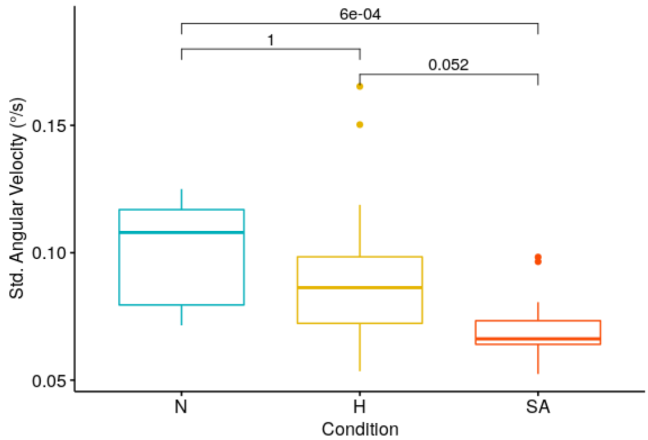


Figure 3.6: Standard deviation of the angular velocity measured in the three conditions.

Paths and trajectories The object trajectories recorded during the trials were analysed to extract meaningful information about the smoothness of the point-to-point movements in the three conditions.

We extracted the epochs representing the point to point movements between two adjacent target points, excluding the alignment phase. The smoothness was calculated for each segment as the standard deviation of the angular velocity around the vertical axis. Indeed, the variation of the rotation around the vertical axis (yaw) represents the change in direction of the formation. The standard deviation of the angular velocity (expressed in degree/s) provides an estimate of the orientation changes of the formation. Shapiro-Wilk's test revealed that data were not normally distributed ($p < 0.05$) (see Fig 3.6), thus we relied on non-parametric tests to compare the distributions. The Kruskall-Wallis test reported statistical significance ($\chi^2 = 12.464$, $df = 2$, $p = 0.002$). The pairwise comparison were performed using Paired Wilkoxon rank sum test, applying the Bonferroni correction for family-wise error rate. The difference between N and SA conditions was statistically significant ($p < 0.001$), while the difference between N and H was not significant ($p = 1.00$). The comparison between H and SA conditions obtained a *p-value* slightly larger than the alpha-level used as a reference in the text ($p = 0.0518$).

The tracking of the object carried under the three haptic policies along the same trajectory is represented in Fig. 3.5.

	H	**N**	**SA**
Trial times [s]	$72.0(\pm\ 17.4)^{**N}$	$90.7(\pm\ 27.1)^{**H,**SA}$	$73.2(\pm\ 16.0)^{**N}$
Norm. times [s/m]	$4.84(\pm\ 1.70)^{**N}$	$6.77(\pm\ 2.35)^{**H,**SA}$	$5.00(\pm\ 1.26)^{**N}$

(*) $p < 0.05$, (**) $p < 0.01$, the letter defines the pairwise comparison

Table 3.1: The table reports the average time to complete the trials and the average normalized time calculated per each condition.

Time Completion time was used as metric for comparing the effectiveness of the three policies. Both the overall trial time and the time to reach the single target were evaluated. Starting with the overall task time, the average time to perform the trials is reported in Table 3.1. Trial durations per condition were statistically analyzed using Shapiro-Wilk test and Levene test to check the normality of distributions and homogeneity of variance, respectively. Both tests were not significant ($p > 0.05$) for the considered combination of trials and conditions. The difference between conditions was assessed using One way Repeated Measures ANOVA, taking into consideration the effects of conditions on the trial duration. The test showed a statistically significant effect of the condition on the trial duration($p < 0.001$, $F = 12.48$, $df = 2$). The pairwise t-test performed on the three combinations of conditions, adjusted with the Bonferroni correction, showed a statistically significant reduction of the trial duration in the H ($p < 0.001$) and SA conditions ($p < 0.001$) with respect to the N condition. On the other hand, there was no statistical difference between the H and SA trial duration distributions ($p = 1.00$).

For what concerns partial times, the time to travel each segment of the trajectory was normalized by the length of the segment itself in order to compare time values across the trials. As it is obvious that longer distances require larger amounts of time, the normalized time accounts for the differences in traveling short and long paths guided by the three haptic policies. Normalized times per condition have been compared using Kruskal-Wallis test, since Shapiro-Wilk test revealed that only data from SA condition were normally distributed ($p = 0.185$). The Kruskal-Wallis test by rank reported a statistically significant difference between groups ($\chi^2 = 50.948$, $p < 0.001$, $df = 2$) , that was further assessed through a pairwise comparison using the Wilkoxon rank sum test. After applying the Bonferroni correction, the statistical tests showed that the N condition was statistically different from the H ($p < 0.001$) and SA ($p < 0.001$) conditions, while the difference between H and SA condition was not statistically significant ($p = 0.40$).

	H	**N**	**SA**
Ease of use	$17.4(\pm\ 3.0)^{*N}$	$11.7(\pm\ 2.6)^{*H,*SA}$	$18.3(\pm\ 2.9)^{*N}$
Ease of learning	$20.1(\pm\ 2.7)^{*N}$	$15.1(\pm\ 2.6)^{*H}$	$17.1(\pm\ 3.0)$
Comfort	$11.9(\pm\ 3.6)^{*N,*SA}$	$16.8(\pm\ 1.8)^{*H}$	$17.8(\pm\ 2.3)^{*H}$

(*) $p < 0.05$, (**) $p < 0.01$, the letter defines the pairwise comparison

Table 3.2: Average score of the participants expressed as mean($\pm$ std). Three different one-factor statistics were calculated for the three questionnaire subsections, using the Haptic Policy as factor (rows are independent).

The normalized time samples were also exploited in a correlation analysis. Since the three investigated policies rely on different sequence of movements to reach the target, we hypothesised that the normalised time could depend on the length of the trajectory segment. Concerning the holonomic movements and the sensory augmentation condition, the p-values obtained for the correlation index ($p = 0.057$, $r = -0.19$ for H condition, and $p = 0.07$, $r = -0.24$ for SA condition, respectively) are greater than the reference alpha level used for this work ($\alpha = 0.05$). Since the difference from the accepted alpha level is low, we should not reject the correlation. Regarding the nonholonomic condition, there is a significant negative correlation ($r = -0.29$, $p = 0.0031$) between distance walked and normalized time.

Questionnaire After each experiment, the users have been asked to fill a questionnaire on the online form, resulting in 10 subjective evaluations. The sum of the ratings obtained in each questionnaire subsection (*Ease of use*, *Ease of learning* and *Comfort*) has been used as metric for the comparison. Since the rating of each item was on a 1-7 Likert scale, and each subsection is composed by 4 items, the overall rating for each subsection ranges between 4 and 28. For the sake of simplicity, we removed the baseline offset and brought the values in the range 0-24 by subtracting 4 to the final value. Table 3.2 reports the average score for each subsection.

A statistical analysis was performed on the questionnaire data to assess statistical relevance in the three subsections. The values obtained from the Likert scale should be treated as ordinals, so we performed a Kruskall-Wallis test for each questionnaire subsection, using the *Condition* (H, N, SA) as independent factor. The Kruskall-Wallis test revealed statistical significance between the rating distributions for the *Ease of Use* ($\chi^2 = 15.901$, $df = 2$, $p < 0.001$), *Ease* *of Learning* ($\chi^2 = 11.344$, $df = 2$, $p = 0.0034$), and *Comfort* ($\chi^2 = 12.831$, $df = 2$, $p = 0.0016$). Post-hoc analysis were carried out using pairwise Wilkoxon signed-rank tests adjusted using Bonferroni correction.

For what concerns the *Ease of Use*, ratings for the N condition were statistically different from H ($p = 0.017$) and SA ($p = 0.027$) conditions. The difference between H and SA conditions was not statistically significant ($p = 1.00$). For what concerns the *Ease of Learning*, ratings for the H condition were statistically different from N condition ($p = 0.017$). The SA condition was not statistically different from N ($p = 0.45$) and H ($p = 0.19$) conditions. For what concerns the *Comfort*, ratings for the H condition were statistically different from N ($p = 0.017$) and SA ($p = 0.032$) conditions. The difference between N and SA conditions was not statistically significant ($p = 1.00$).

3.1.5 Discussion

The experiments were conducted using the information provided by the camera-based tracking system. Although its tracking error was not negligible, all the participants still managed to complete the tasks in acceptable time. We might hypothesize that the tracking error was partially covered by the error due to the human interpretation and actuation of the haptic patterns, and by the threshold for reaching the target goal.

The correlation analysis on the normalized time revealed an overall negative correlation between normalized time and path length, that might refer to the optimization of the formation motion in longer paths.

The comparison of the deviation from the optimal trajectory estimated in the three conditions suggests that the Holonomic policy allows the users to follow the instructions more accurately than in the other conditions. This result does not surprise, as in the Nonholonomic and Mixed conditions a small angle error may cause wide deviations from the optimal trajectories.

Experimental results from overall and partial times suggest that the Nonholonomic condition was less effective in guiding quickly the formation through the paths. We hypothesize that the combination of rotation and translation, if performed separately, introduces delays in the task execution. Moreover, in the Nonholonomic condition the rotation cue does not indicate the desired orientation, but only the steering direction (clockwise or counterclockwise), so the users have to rotate until the rotation pattern is interrupted. For what concerns the performance of the Holonomic and Sensory Augmentation conditions, the trial duration was comparable. We may suppose that the small workspace dimension affected the trial execution time, as the correlation analysis highlights the diminishing of normalised time for travelling a segment with the increasing of the distance from the goal. The experimental trials should be repeated in a large workspace to assess the effects of distance on the performance related to the three conditions.

The analysis of questionnaire ratings adheres with the experimental results, as the Nonholonomic condition was rated the lowest in the *Easy To Use* and *Easy to Learn* items with respect to the other conditions. On the other hand, both the Nonholonomic and Mixed conditions achieved higher ratings than the Holonomic condition in the *Comfort* items. This may suggest that the users preferred the nonholonomic walking model as a paradigm for guidance, for what concerns comfort and naturalness of movements. On the other hand, the guidance policies that featured a single typology of haptic cue (Holonomic and Mixed) were rated as more intuitive and easy to learn.

If we consider the three conditions under the different metrics measured in this work, it is possible to say that the SA condition was time-efficient, reported the greatest trajectory smoothness, and received high ratings for what concerns *Ease of Use*, *Ease of Learning* and *Comfort*. Condition H was also time-efficient, while the smoothness was intermediate. Scores were high for *Usability* and *Ease of learning*, but low for *Comfort*. The N policy, despite the high *Comfort* ratings, had the worst results for what concerns time-efficiency and smoothness, and received low scores on *Usability* and *Ease of learning*.

During the *a posteriori* interpretation of the results, we isolated the relevant independent factors that affected the properties and performance of the three haptic policies: *i)* Continuous Availability of the target direction during the task, *ii)* Freedom of motion, *iii)* Holonomic or Noholonomic walking model, *iv)* Number of haptic patterns.

The main traits of SA condition are that the users were provided with a single haptic pattern, they were always aware of the goal position (pointed by the haptic pattern), and were free to select their walking strategy. Although not backed by experimental results, the visual inspection of the experimental trajectories showed that in the SA condition the users rarely selected the holonomic strategy whenever the target direction was not parallel to the goal.

H condition also featured a single haptic pattern, the users were always aware of the destination (pointed by the haptic pattern), but the movements were constrained to holonomic policy.

N condition had double pattern, the users were not always aware of the destination (because the rotation pattern concealed the indication of the target, represented by the translation cue) and the movements were constrained to nonholonomic policy.

The envisaged independent factors, selected by comparing the design of the three conditions, are:

I) Continuous Availability of the target during the task *VS* Target availability partially impaired;

	SA	**H**	**N**
Destination Available	✓	✓	
Freedom of Motion	✓		
Holonomic Strategy		✓	
Single haptic pattern	✓	✓	
Time-efficiency	✓	✓	
Smoothness	✓		
Ease of Use/Learning	✓	✓	
Comfort	✓		✓

Table 3.3: the table rows represent the **independent factors**, in bold, and the *dependent factors*, or measured metrics, in italic. Columns represent the three haptic policies.

II) Free selection of the walking strategy *VS* Motions constrained by the instructions;

III) Holonomic *VS* Nonholonomic strategy of motion;

IV) 1 haptic pattern *VS* 2 haptic patterns.

We hypothesized that the independent factors might have played a role in the task performance, thus being reflected in the calculated metrics. In order to have a visual comparison of the factors, we attributed a binary value to each independent factor according to the considered haptic policy. For instance, we might describe the SA condition as: Continuous Target Availability (true), Free selection of the walking strategy (true), Holonomic motion (false), single haptic pattern (true). For H condition: Continuous Target Availability (true), Free selection of the walking strategy (false), Holonomic motion (true), single haptic pattern (true). For N condition: Continuous Target Availability (false), Free selection of the walking strategy (false), Holonomic motion (false), single haptic pattern (false). The factors are reported in Table 3.3 as rows, whereas columns represent the considered haptic policy. Moreover, we enriched the table with the binary ratings of the exploited metrics, considered *dependent factors*, according to the three conditions.

The table shows the following trends:

1) **Destination availability**, **Single haptic pattern**, *Time-efficiency* and *Ease of Use/Learning* show a similar trend.

2) **Freedom of Motion** and *Smoothness* show a similar trend.

3) **Holonomic Strategy** and *Comfort* (negative correlation) show a similar trend.

According to the trends highlighted by the table, we formulated hypotheses on the relative dependence of factors.

1.a) The conditions that featured a single haptic pattern (SA and H) were also continuously displaying the goal direction. The effects of said two factors might be additive, in the sense that the increased simplicity of displaying a single pattern and the fact that the goal direction is always known might induce variation of the same sign on the dependent factors. On the other hand, we are more prone to hypothesize that the second pattern used in the N condition to instruct rotation was impairing the goal information availability. In this hypothesis, the relevance of the factor Number of pattern is secondary to the destination availability, in the sense that the availability of the target direction has a more relevant effect than the number of haptic patterns displayed.

1.b) We might hypothesize that always knowing the goal direction is relevant for time-efficiency, since the users can focus their efforts on the strategy to reach it.

1.c) We might hypothesize that always knowing the target location had an effect on the perceived usability and intuitiveness of the haptic policy.

1.d) Also the use of a single pattern might affect the perceived intuitiveness, since it is easier to interpret one haptic pattern than distinguishing between two.

2) The smoothness of trajectory refers to how steady the movement was during the task. We might hypothesize that condition SA was smoother because motions were not constrained, *i.e.*, following the instructions leads to less smooth point-to-point movements. Probably, also the Destination availability affected the trajectory smoothness, since knowing the goal location is necessary to walk steadily toward it. We hypothesize that the smoothness records the interaction between Freedom of movements and Destination availability factors here.

3) The increased comfort perceived by the users in the SA and N condition might refer to the walking model adopted, that is the Nonholonomic for the N condition, and is closer to Nonholonomic than to Holonomic in the SA condition (although not proven by numbers).

To extract new insights on the conditions, we expressed the three haptic policies as a combination of the independent factors: SA can be considered a control condition, that allowed participants to always perceive the direction to the goal and to move without constraints. In the H condition the direction to target was always available, but motions were constrained according to the holonomic model (by experimental protocol). In the N condition the direction to target was partially impaired, and the movement was constrained according to the nonholonomic model.

We hypothesize that constraining the movements (despite the model adopted) did not strictly affect the time-efficiency, while the smoothness was affected. Instead, impairing the perception/knowledge of the target direction had a relevant effect on the time to perform the task and on the smoothness of walked trajectories. The Ease of Use and Ease of learning might be linked to the number of haptic patterns, or, more probably, to the availability of goal location. We interpret this link as the intuitiveness of the haptic cues. In accordance with the literature about the human walking strategies, the condition featuring the holonomic model reported the lower comfort rating. Anyway, the implications of this result are not clear. Further works might analyse this aspect in the light of ergonomics or muscle fatigue.

Finally, the proposed hypotheses still need a deep and extended testing, but they might provide a basis for further studies on the guidelines for the design of haptic policies in human guidance. For instance, if the primary metric in the task is the time-efficiency, the haptic policies should be shifted toward the availability of relevant information. On the other side, a task requiring smoothness should also include the self-selection of the motion strategy. The guidance based on instructions, instead, might be preferable in task requiring greater coordination, *e.g.* guiding a larger formation.

3.1.6 Conclusions

This work investigated the problem of guiding a formation of humans mediated by haptics during a cooperative task, *i.e.* transporting a bulky object in a predefined workspace. Three guidance policies have been designed considering different models of human locomotion and intuitiveness of haptic patterns. The theory of grasp was used to model and generate the haptic cues for steering the group of humans. The experimental and subjective results revealed that using the haptic policies to provide the users with relevant information on the task, *i.e.* the direction to the goal, was the most efficacious factor for achieving time-effectiveness, smoothness of the walked trajectories and intuitiveness of the haptic patterns. Providing strict instructions was detrimental from the point of view of the trajectory smoothness, that might represent the steady flow of the task. Conversely, the instruction-based guidance could be preferred in controlling a larger formation, or in any scenario that requires accurate movement. The adoption of holonomic and nonholonomic motion paradigm is expected to affect the comfort perceived by the participants during the task. This aspect may be exploited for the ergonomics aspect of the guidance policies in the design process. Further experimental studies on realistic scenarios (presence of narrow passages, multiple users and larger workspace) will provide a more precise assessment. The grasping matrix approach can be improved by a finer implementation of the contact model, *e.g.* considering the relative distance of the users from the carried object to determine if the formation is breaking and provide an alert.

Chapter 4

Minor projects on Haptic Guidance for Humans

The last chapter contains two minor projects that are not strictly related to walking or cooperation scenarios, but investigate applications of haptic stimulation to provide users with assistance during daily activities.

The application described in Sect. 4.1 deploys the haptic interfaces as detection and notification tools to prevent unwanted behaviors and guide the users toward healthy habits. The No Face-Touch project has been developed as a *Do It Yourself* solution to promote the user's safety and awareness toward unhealthy habits during the Covid-19 pandemics. We wanted to help the scientific and non-scientific community with effective and non-invasive tools, that were accessible for everyone with small efforts. While respecting social-distancing, hygiene and protection rules has the main impact on limiting the virus spreading, we addressed the self-inoculation issue, that happens when people touch contaminated objects and then transfer the virus to mucosal areas. The infection via self-inuculation determines a lower amount of infections when compared to air-mediated transmission, but is less predictable and thus more difficult to prevent. The solution we propose is detecting the face-touch occurrences, defined as contact between the user's hand and face (regardless the part touched), and alerting the user to reduce the incidence of dangerous events. The detection system leverages wearable devices with very low encumbrance to cope with the users' comfort requirements. We developed two detection approaches that can be implemented on a wide range of smartwatch and smartbands, leveraging either magnetic or inertial measurements. The system performance has been investigated with user studies on the effectiveness of the proposed approach in detecting face-touches in laboratory and real-life scenarios.

Sect. 4.2 proposes a solution to embed haptic stimulation in an Augmented Reality system designed for daily use, with the aim to support workers and general users during their activities by enhancing their environmental awareness and task performance through visual and haptic cues. The presented system implements Augmenting Reality, finger tracking and haptic feedback for smartphone. The work tackles the current lack of 'physical interaction' and haptic feedback (apart from the device vibration) during AR experiences on smartphones. In addition, it is designed to cope with industrial and daily-use scenarios to promote the use of the Augmented Reality in real-life applications. Differently from other approaches in literature, the setup does not require external cameras, head mounted displays (HMD) or cumbersome hardware. Instead, it uses a wearable haptic interface worn on the finger, and proposes a forearm montage of the smartphone to keep both the hands free. The smartphone placement, depicted in Fig. 4.7 can be counterintuitive at first, because the positioning resembles the one used for watches, but it provides several advantages for the AR purpose. Indeed, it simplyfies the tracking of the hand and the virtual object in the environment. By projecting the hand and the virtual objects in the same reference system, the virtual physical interactions can be measured and displayed on the user's skin. A major objective is unburdening the hands from holding the phone and the haptic interfaces to interact with the virtual environment. The former issue is addressed with the smartphone support worn on the forearm, that provides the user with a display firmly attached to the body. The haptic interfaces has a small form-factor with respect to common exo-skeletons or grounded interfaces, and can be optimized to become very small. When adopted for daily use, AR should not deliver an immersive experience because the users need to be aware of the surroundings. We propose an AR experience that provides support in daily tasks, but should be used 'on demand' to augment the human capabilities and guide the user only during the task. An envisioned application is, for instance, the manual assembly of parts in industrial scenarios.

4.1 No Face-Touch: Detecting and Notifying Unsafe Gestures

The alarming morbidity of COVID-19 has drawn the attention to the social role of hygiene rules, with a particular focus on the importance of limiting face-touch occurrences. To deal with this aspect, we developed No Face-Touch, a system able to estimate the hand proximity to face and notify the user whenever a face-touch movement is detected, promoting the user's awareness. In its complete setup, the system consists of an application running on the smartwatch and a wearable accessory. Its ease of implementation allows this solution to be ready-to-use and large-scale deployable. We developed two gesture detection approaches compatible with sensors embedded in recent smartwatches, *i.e.* inertial and magnetic sensors. After preliminary tests to tune target gesture parameters, we tested the two approaches and compared their accuracy. The final phase of this project consisted in exploiting the most robust approach in a daily living scenario during a 6-days campaign. Experimental results revealed the effectiveness of the proposed system, demonstrating its impact in reducing the number of face-touches and their duration.

4.1.1 Motivation

The epidemic of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused 10 357 662 laboratory-confirmed infections including 508 055 deaths all over the world by 1st July 2020 [118]. Among other factors, the speed of the outbreak has inevitably provoked national and global public health crisis. Not only has coronavirus disease (COVID-19) an alarming morbidity and mortality, but it has also an extended incubation period and a high variability in symptoms manifestation, which result in important implications for surveillance and control activities [119].

Among the policies carried out in response to COVID-19, individual protective behaviour has a great significance on the reduction of the index R_0, *i.e.* the average number of infections caused by a primary case in a population consisting only of susceptibles [120]. As a general rule, protective behaviours can be classified into three groups: preventive, avoidant, and management of disease [121]. The first group includes hygiene measures (such as hand washing, cough and sneeze etiquette, and surfaces cleaning), mask wearing and uptake of vaccinations. Observance of these behaviours has effects mainly on the risk of transmission factor. Avoidant behaviours are mostly represented by social distancing, *e.g.* avoid going to crowded places, maintaining at least 1 metre distance between ourself and others, working in compliance with quarantine restrictions. The last category includes following the directions of local health authority when seeking medical attention and staying home and self-isolate even with minor symptoms. In response to a pandemic flu, respecting hygiene measures becomes even more valuable in case virus transmission can occur by self-inoculation, *i.e.* by transferring contaminated material from hands to other body sites [122, 123]. Although the literature on the mechanisms of self-inoculation of common respiratory infections (*e.g.*, influenza, coronavirus) is limited [124–126], contaminated hands have been reported as having potential to disseminate respiratory infections [127], especially if associated to face-touches [128]. As regards SARS-CoV-2, if the virus is transferred to eyes, nose or mouth, it can enter the body and infect the subject [129], therefore avoid touching the face has to be a paramount prevention habit. In crucial contexts as health care settings, frequent face-touching is a potential mechanism of acquisition and transmission. A self-inoculation event may occur if a health care worker fails to comply with hand hygiene rules after patient contact or after contact with the patient's contaminated environment.

Although consequences on hygiene related aspects are the most evident, they are not the only valid reason to discourage people from touching the face. There exist behavioural disorders which are strictly connected to this repetitive movement. Onychophagia [130] (the habit of biting one's own nails), trichotillomania [131] (the repetitive pulling of one's own hair) and dermatophagia [132] (the habitual biting of the skin) are just few examples. From the patient's point of view, episodes of such bad habits are often unintentional and occur with little apparent control or awareness. This aspect is further supported by a behavioural observation study undertaken in [133] where 26 subjects were observed and videotape recorded to monitor the occurrences of face-touches. Using standardized scoring sheets, the frequency of hand-to-face contacts with mucosal or nonmucosal areas was tallied and analysed. On average, subjects touched their face 23 times per hour. Of all face-touches, 44% involved contact with a mucous membrane, whereas 56% involved non mucosal areas. Of mucous membrane touches observed, 36% involved the mouth, 31% involved the nose, 27% involved the eyes, and 6% were a combination of these regions.

A common behavioural intervention designed to reduce the manifestations of habit-based disorders is known in literature as habit reversal therapy (HRT) [134]. Its techniques can be organized in five phases: *(i)* awareness training, *(ii)* relaxation training, *(iii)* competing response training, *(iv)* motivation procedures, and *(v)* generalization training. In response to the need to improve awareness, several low-tech strategies including wearing heavy bracelets, perfume, gloves, *etc.* have been used [135]. On this direction, acoustic, visual, and haptic signals can be employed for providing alerts to users. As an example, in [136] a loud tone was used as a deterrent with a 36-years old woman who had been diagnosed with moderate mental retardation and hair pulling. This study demonstrated how an audio alert upon coming in contact can be experienced as aversive and may contribute to a reduction in bad behaviours.

As a matter of fact, audio and/or visual cues may be ineffective or undesired in some circumstances, especially when vision is temporarily impaired or background noise makes auditory feedback difficult to hear or understand. On the contrary, the sense of touch is not only the most robust and distributed of human senses, but it is also proximal, bidirectional, and private [137]. These features make the haptic channel particularly suitable to convey information in everyday environments, where visual and auditory modalities might be busy to effectively accomplish a task (*e.g.*, vision occupied in finding objects), impaired due to personal protective equipment (*e.g.*, worker wearing headphones), or inappropriate

Figure 4.1: The application No Face-Touch runs on the smartwatch. It estimates hand proximity to face and notifies the user with a vibration whenever a face-touch movement is detected.

(*e.g.*, student attending lecture, spectator during a public show). Vibrotactile anklets [138], dorsal and waist belts [139], bracelets [140], and rings [141] have been deeply studied and extensively exploited as haptic means for providing information to users. Examples range from encoding complex directional cues to human-robot collaboration, enhancing human-human social activities, limb guidance and situational awareness.

In this manuscript, we present "No Face-Touch", an open project exploiting haptic feedback for suggesting and training people to develop good habits, in order to limit further transmission of SARS-CoV-2, and more in general, to help people become more aware of their face-touching. To systematically detect all the times the hand approaches the face, an automatic system is required. We exploited widespread and off-the-shelf devices, such as smartwatches and smart bracelets, to track the human hand and notify the subject in case of face-touching. This solution will minimize the mental effort required to keep hands away from the face, catching also involuntary movements that would take place without the subject noticing. The concept is visually summarized in Fig. 4.1. The choice of the smartwatch as core technology came for a precise reason: we wanted to provide immediate help to people, without the requirement of buying or creating new hardware. From the literature, widely exploited is the use of cameras [142], magnetic technologies [143], and exoskeletons [144] to track arm and wrist pose.

Even though recognized as reliable, these setups require expensive and bulky equipment or complex and elaborated installation procedures. On the other hand, several methods have been developed for computing the absolute objects pose (*i.e.*, with respect to the world reference system) by means of Micro Electro-Mechanical Systems (MEMS) sensors, typically embedded in smartwatches [145].

This section is organized as follows. Subsect. 4.1.2 provides a description of the proposed system from an engineering perspective, including hardware and software specifications. The third subsection (Subsect. 4.1.3) presents in detail the algorithms we developed to detect face-touch events. Subsect. 4.1.4 describes the experiments performed to verify the objectives achievement and reports *a-posteriori* discussions, enriched with statistical analysis of results. Conclusions are drawn in Subsect. 4.1.5, along with a brief discussion on the range of possible new development directions that No Face-Touch may enable. Source code repositories, available releases, and compatible devices are listed in Subsect. 4.1.6. A conclusive Appendix contains the pseudo-code implementation of the algorithms detailed in Subsect. 4.1.3.

4.1.2 No Face-Touch system

The objective of the No Face-Touch project is identifying whenever the hand gets too close to the face, and alerting the user to stop the current motion. With the aim to develop a ready-to-use and large-scale deployable system, two design guidelines have been followed. Firstly, only technologies already available on widespread devices have been exploited. Secondly, the system implementation is thought to be highly plug-and-play, meaning that no complex installation and/or hardware assembly procedures are needed for the system to work. As a result, No Face-Touch is composed of three elements: *i)* a smartwatch worn by the user; *ii)* an application running on the smartwatch or on the companion smartphone; *iii)* a wearable accessory worn close to the face (like a necklace, a pair of earrings or a pair of glasses) embedding magnets to generate a detectable magnetic field. While the first two elements are essential for the system functioning, a configuration without the third element has been proposed and evaluated as alternative solution.

Two different policies have been developed. In one case, the algorithm leverages data coming from accelerometer and magnetometer sensors, while in the second case only acceleration measurements are used. Indeed, this second method accounts for the fact that many smartwatches do not feature a magnetometer. Although less robust (as reported in Subsect. 4.1.4), in our regards it was worth proposing an alternative approach to provide support to the largest possible population. From the software point of view, the application has been developed for different platforms to take into consideration the variety of smartwatch brands and operative systems available in commerce.

Obviously, both left and right arms can be perform face-touches. Even if the system has been characterized and tested while wearing a single device, *i.e.* monitoring a single arm only, we believe that its validity is not diminished. Indeed, it has been demonstrated that HRT is an efficacious long-term behavioural intervention [146, 147]. Thus, a valid alternative to wear two devices is interchanging the smartwatch position on left and right arm.

4.1.3 System overview

In this subsection, we present methods and algorithms implemented for carrying out the different stages of the experimental validation. For each phase, a different software has been developed to collect data from inertial and/or magnetic sensors. While the first evaluation is preparatory for the system functioning, and thus the *ad-hoc* application does not have any use outside the experiment, the algorithms used in the second and third experiments have been released in a public repository as the No Face-Touch application.

In what follows we report the rationale and description of the exploited methods, whereas the experimental evaluations are in Subsect. 4.1.4.

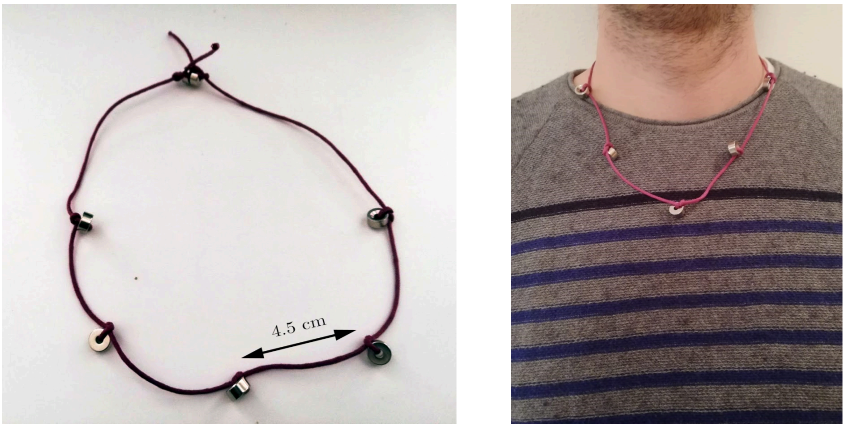
Safe and Unsafe Orientations

As a preparatory phase, we identified a range of admissible wrist orientations for face-touch events with the aim of establishing discriminatory conditions for the face-touch detection. Indeed, not all wrist orientations are compatible with natural touches of the face. Anatomical constraints and articular control strategies suggest that, in order to detect the hand approaching the face, we can consider a narrow subset of all the possible wrist orientations [148]. Therefore, starting from theoretical values found in literature, we recruited participants to define the boundaries of the above mentioned subset. The experiment aimed at classifying wrist orientations into two categories:

safe: wrist orientations that are not compatible with natural face-touch movement;

unsafe: wrist orientations adopted while the hand approaches the face.

To this end, subjects' hand movements were measured by means of an *ad-hoc* app running on the smartwatch. The application implements the algorithm described in [149] which, on the basis of the Multiplicative Extended Kalman Filter (MEKF), accurately estimates the body posture with a low-cost wearable setup. In particular, the MEKF estimation proposed in [149] performs a correction step only when the measurements are sufficiently reliable. The resulting system does



(a) (b)

Figure 4.2: In (a) a handmade magnetic necklace prototype. It contains 5 neodymium magnets, 4.5 cm far from each other. In (b) a user wearing the magnetic necklace.

not suffer from occlusion problems and lightening conditions, and it can be used in indoor and outdoor environments. Moreover, since only accelerometer and gyroscope are used to estimate the orientation, the system can be used in the presence of hard and soft iron and magnetic disturbances, common in smartwatches. The interested reader is referred to [149] for further details.

Boundary values for *safe* and *unsafe* wrist orientations obtained in this experimental phase are adopted in both detection algorithms.

Detection with magnetometer

The proximity between hand and face is estimated thanks to a virtual magnetic barrier generated by the worn magnets (see Subsect. 4.1.2). We used 5 tiny and low-cost neodymium (N42) magnets (10 mm external diameter, 5 mm thick, 2 kg pull force). We experimentally verified that each magnet can generate a magnetic field of $\sim$420 µT at 5 cm distance, *i.e.* a substantial variation with respect to the magnitude of the Earth's magnetic field which at its surface ranges from 25 to 60 µT [150]. In a typical scenario the user wears a necklace with 5 magnets 4.5 cm far from each other, as shown in Fig. 4.2. Such arrangement is adequate to detect the smartwatch proximity around the necklace, as a result of the significant magnetic field generated by the high performance magnets. Thanks to the calibration procedure, a different number of magnets with different technical specifications can be employed, allowing each user to build his own wearable accessory with great flexibility.

As a matter of fact, devices and objects that interfere with the magnetometer populate almost every daily environment. The Earth magnetic flux is remarkably deflected and modified by ferromagnetic materials, while electronic devices such as computers, mobile phones, and general household appliances generate electromagnetic fields (EMF) that may result in artifacts or relevant fluctuation of the baseline magnetic field noise detected. From this observation arises the need to reinforce the estimation based on magnetometer data with wrist orientation measurements. Results of preliminary experiments (Subsect. 4.1.4) revealed that exploiting roll and pitch angles and neglecting yaw orientation is enough to predict if the hand is reaching an unsafe position. Moreover, the robustness granted by the MEKF algorithm in Subsect. 4.1.3 comes at the cost of heavy computations. Given the target devices and the related battery consumption issues (the app might work as a background process for several hours), we opted for a simplified estimation procedure exploiting the gravity vector components sampled by the triaxial accelerometer.

The algorithm continuously estimates the orientation of the wrist and checks if it is within the boundary limitations previously defined, *i.e.* if the hand orientation is safe or unsafe. The *safe orientation* condition (cfr. Subsect. 4.1.3) relies on biomechanical constraints defined in the preparatory experiments, thus we can assume that the hand is not approaching the face whenever this condition is met. In this case, the app retrieves information about the baseline magnetic field around the user, which is an essential step for discriminating variations due to sensor proximity to the magnets when the wrist orientation is unsafe. Magnetic field measurements occur at 100 Hz and are used to update a 50-elements buffer that stores information about the environment. This method allows to always keep track of the properties (average value and variance) of the magnetic field around the user, so that a robust threshold approach can be applied. On the contrary, if the current wrist orientation is considered *unsafe*, the magnetic field measurements are compared with the current model of the magnetic field around the user. If the value of the sensed magnetic field exceeds the standard deviation of the baseline recorded (std) multiplied by a constant factor α, then the smartwatch is considered too close to the face and a vibratory alert is triggered to stop the current hand movement. The vibration is interrupted after the face-touch conditions are no longer met. The quantity $\alpha \cdot std$ represents the maximum deviation from the baseline explained by background noise in the algorithm model. By default settings, the α value is set to 3. Indeed, from the theoretical point of view, given mean and standard deviation of a Gaussian distribution, almost 99.7% of values would fall in the $mean \pm 3 \cdot std$ interval. On the other hand, because of the irregularity of the magnetic noise, this value has to be increased to efficaciously detect the presence of the magnets. Therefore, the multiplication factor α used in the real application is defined after a calibration procedure (lasting 5 seconds). During this phase the user is asked to extend and move his arm in front of him for 2 seconds and slowly move it toward the necklace, reaching a distance of about 20 cm far from the latter in the remaining 3 seconds. In the first 2 seconds, the procedure records the magnetic field in the environment (far from the magnets), then computes the mean and standard deviation to construct the baseline magnetic field model. During the following 3 seconds, the maximum value of the magnetic field is measured. As a result, α is calculated as the fraction between the maximum value stored in the last 3 seconds and the standard deviation computed in the first part of the calibration. The distance adopted in the calibration phase is recommended as an indicative value for ensuring that the collected data do not include relevant variations of the magnetic field caused by the presence of the permanent magnets. Indeed, considering that the magnetic field intensity decreases proportionally to the squared distance from the source, the influence of the magnetic accessory is considered negligible when the sensor is at a distance of 20 cm from the necklace. In practice, the user can estimate this distance by touching the central magnet with the index fingertip.

A pseudo-code implementation of the method is reported in Algorithm 1.

Detection without magnetometer

A different version of the face-touch detection algorithm has been developed to be functional also in simpler wearable devices that do not embed a magnetic field sensor. Examples of these devices are the common and widespread fitness smartbands that can be exploited with a companion smartphone. Similarly to [151], the proposed algorithm leverages inertial measurements only. The developed policy aims at recognizing wrist motions that could be associated with a face-touch. The main assumption is that only a subset of all the possible hand motions terminates in a contact with the face. By taking advantage of biomechanical constraints of the human body, we can predict if the current combination of wrist orientation and acceleration profile leads to a touch. A pseudo-code implementation of the algorithm is provided in Algorithm 2, while its functioning is detailed in the next lines.

As in the first part of the previous algorithm, gravity vector components (a_x, a_y, a_z) sensed by the smartwatch are used to reconstruct roll and pitch angles of the wrist in world reference frame. In the initialization phase, the user is asked to maintain the hand in resting position for 2 seconds while a calibration procedure acquires inertial data and computes the correspondent wrist orientation. This process provides an estimate of the sensor's noise, which is necessary to discriminate angle variations due to slow movements from random processes. Thus, a threshold (β) is obtained from the standard deviation of the variation

Smartwatch Model	**Magnetometer**	**Accelerometer**	**Weight**	**Operative System**	**Battery**
LG Watch Urbane 2	Yes	Yes	92 g	Wear OS 2.1	Li-Ion 570 mAh
Asus Zenwatch 3	Yes	Yes	48 g	Wear OS 2.1	Li-Ion 341 mAh
Huawei Watch 2	Yes	Yes	57 g	Wear OS 2.1	Li-Ion 420 mAh
Apple Watch Series 4	No	Yes	48 g	WatchOS 6.1	Li-Ion 292 mAh

Table 4.1: List of the devices used for the experimental evaluation.

of the orientation and it is used to distinguish angle variations due to movement from noise.

Once the calibration is done and the threshold is defined, the algorithm monitors both the wrist orientation and the variation of the wrist pitch. More in detail, at each iteration, the difference between current pitch value and previous one is calculated and stored in an array named *slope* as follows: if the last variation is greater than β, then *'1'* is appended, otherwise *'-1'*. The average value of the last *50* samples of *slope* reveals whether the hand is moving upwards. If the hand was rising and the wrist orientation is within the unsafe range, then the user receives a vibratory alert. The vibration is interrupted after these conditions are no longer met. Clearly, there are many gestures that correspond to the subset of movements identified by the conditions defined. Hence, in this algorithm the number of false positive is expected to be higher with respect to the algorithm exploiting the magnetometer sensor.

4.1.4 Experimental validation

In this subsection we detail the stepwise validation undertaken to assess No Face-Touch functioning. A list of the adopted devices and their specifications is reported in Table 4.1. For each step, experimental protocol, setup, and results are described in what follows. The study was approved by the Local Institutional Ethics Committee. Each subject gave her/his written informed consent to participate and was able to discontinue participation at any time during experiments. The experimental evaluation protocols followed the declaration of Helsinki, and there was no risk of harmful effects on subjects' health. Data were recorded in conformity with the european General Data Protection Regulation 2016/679, stored locally on the smartwatch with anonymized identities (*i.e.*, Subject 1, Subject 2), and used only for the post processing evaluation procedure. Please note that no sensible data were recorded (only date and time of face-touch events detected). A detailed summary of the carried out experimental sessions is reported in Table 4.2.

Experiment	**Pattern set (per participant)**	**Location**	**Number of participants**
Preparatory (E1)	20 trials	Lab setting	10
Algorithm Comparison (E2)	2 conditions x 2 set of 30 trials each	Lab setting	10 (6 from E1)
Real Scenario (E3)	2 conditions x 6 trials lasting 4 hours	Home setting	10 (5 from E2)

Table 4.2: Summary of the experiments. The pattern set describes the number of trials performed by each participant. The 2 conditions referred to in E2 are the magnetometer-based and inertial-based algorithms, while in E3 the 2 conditions differ by the enabling of haptic alerts.

Preliminary assessment

As a first step, we investigated wrist orientations leading to face-touches. As anticipated in Subsect. 4.1.3, we consider *safe* an orientation that is not compatible with a face-touch event. Conversely, we indicate as *unsafe* the wrist orientations compatible with a contact between the hand and the face. The experiment was held in laboratory settings. Ten healthy subjects (7 males and 3 females, aged 24-59, 6 right-handed and 4 left-handed) were recruited. None of them reported any known deficiency in perception abilities or physical impairments. Participants were informed about the procedure and trained on the experimental system handling.

The smartwatch adopted was a LG Urbane 2. Both right-handed and left-handed users wore the smartwatch on their non dominant arm, in accordance with [152]. Then, they were tasked to freely touch their face 20 times for at least 5 s with the hand wearing the smartwatch. Hand motions were recorded using a RGB camera (12 Mp, 4608 x 2592 pixel, F 2.2, 30 fps) and tracked by means of an *ad-hoc* application implementing a customized version of the algorithm presented in [149] and running on the smartwatch.

Results In the post-processing phase, video recordings were synchronized with inertial data recorded by the app and used as reference to identify angular patterns measured during the face-touches. Wrist angles were estimated for each timeframe and selected by visual inspection of the video. The set of angles compatible with face-touches were used to classify *safe* and *unsafe* orientations. Considering the Earth as a reference system, we used a common aeronautical inertial frame where the x-axis points north, the y-axis points east, and the z-axis points down, aligned with the gravity. We will call this as North-East-Down (NED) reference frame. The terms used to represent a given orientation are roll (θ), pitch (ϕ), and yaw (ψ) for rotations around x-, y-, and z- axes, respectively. The sensor reference system for the smartwatch employed is depicted in Fig. 4.3. Results of this preparatory phase assessed that any values of yaw is compatible

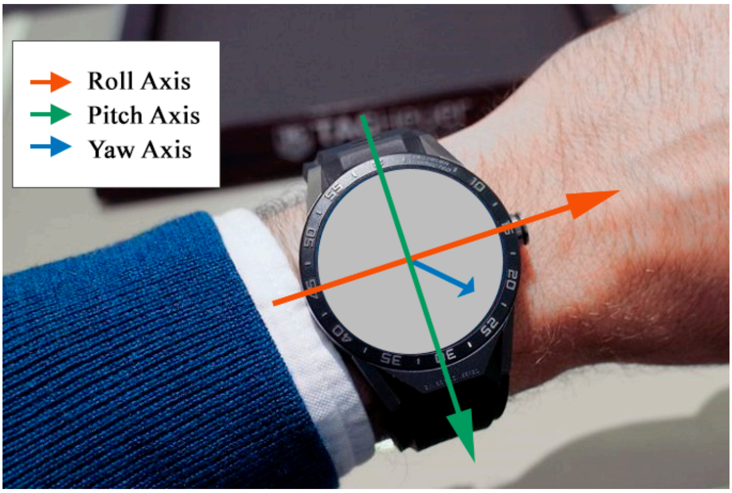


Figure 4.3: The smartwatch local reference system is defined by sensors axes: the longitudinal axis (roll), transverse axis (pitch), and vertical axis (yaw) are depicted in red, green, and blue, respectively.

with a face-touch. This is an obvious result, since yaw rotations describe orientation changes around the z-axis (*i.e.*, the Earth gravity vector). For what concerns roll and pitch, inertial data recorded during the experimental trials were analysed to retrieve the set of angles compatible with a face-touch.

The following values are considered compatible with unsafe orientations for right-handed users:

$$\begin{aligned} -90^\circ = \theta_{min} < \theta < \theta_{max} = 70^\circ \\ 30^\circ = \phi_{min} < \phi < \phi_{max} = 100^\circ \end{aligned} \tag{4.1}$$

For what concerns left-handed users, obtained values were very close to the ones identified in (4.1), so for the sake of simplicity, we decided to exploit the same ranges. Due to the opposite orientation in wearing the watch, the values are symmetrical:

$$\begin{aligned} -70^\circ = \theta_{min} < \theta < \theta_{max} = 90^\circ \\ -100^\circ = \phi_{min} < \phi < \phi_{max} = -30^\circ \end{aligned}$$

Algorithms comparison

As a further step towards the goal of evaluating the system, we compared the performance of the two proposed algorithms. An experimental validation was conducted in laboratory settings to measure the accuracy in detecting potential face-touches.

Ten subjects (7 males and 3 females, aged 21-61, all right-handed) were involved in this experiment. Three different smartwatches running the app in background were used: a LG Urbane 2, an Apple Watch series 4, and an Asus ZenWatch 3. Participants were tasked to perform two trials: *i)* attempt to touch their face 30 times, and *ii)* simulate 30 common gestures of Activity of Daily Living (ADL). The set of ADL was previously selected from the list proposed in [153], according to the criteria of choosing gestures similar to a face-touch (e.g., eating with a spoon, drinking from a mug, hair-combing, putting on a t-shirt). Subjects were asked to wear the smartwatch on their non dominant arm and perform twice both trials (face-touch and ADLs gestures, respectively), once with the magnetometer-based algorithm and once with the accelerometer-based algorithm. The order of gestures and conditions was pseudorandomly selected at the beginning of each experiment. Participants were recorded using RGB camera and data were post-processed to evaluate and compare the accuracy of the algorithms. *Correct detections* and *false positives* were used as metrics. To estimate the number of correctly detected face-touches, we measured the number of alerts generated by the system and we compared them with the total number of face-touch gestures performed by the user. Similarly, the number of notifications displayed by the device while performing other motions is reported as false positive *i.e.*, number of undesired vibratory alerts.

Results In Table 4.3 we reported the average percentages of correct detections and false positives computed among the users. Results confirmed the hypothesis, *i.e.* the algorithm relying on both accelerometer and magnetometer sensors is more robust and more accurate than the one exploiting only accelerometer measurements. It is worth pointing out that the higher percentage of correctly detected scored by Algorithm 2 is biased by the large amount of false positives. In fact, this algorithm is prone to exceed in alerting the user, regardless the motion.

On the basis of the obtained results, we decided to continue the experimental validation using only the algorithm comprising the magnetometer sensor.

Real scenario

Once the algorithm accuracy had been assessed, we performed an experimental campaign to evaluate the effectiveness of the app. We formulated two hypotheses:

i) the system reduces the duration of face-touches (immediate effect due to notification of gesture detection);

ii) the system reduces the amount of face-touches (medium-to-long term effect).

	Correctly Detected	**False Positives**
Algorithm 1	91.3%	3.2%
Algorithm 2	92.6%	38.1%

Table 4.3: Algorithms accuracy comparison. Correctly Detected expresses the percentage of alerts provided over the total face-touch gestures, whereas False Positives indicates the ratio of generated vibrations with respect to the total number of Activity of Daily Living executions.

In order to test No Face-Touch for both hypotheses, experiments were carried out using two conditions: detection notified with a vibration (V) and detection not notified (N). The latter has been used as a control condition for the statistical analysis.

Ten subjects (6 males and 4 females, aged 23-65, all right-handed) were involved in this phase. Three different smartwatches running the application in background were used: a LG Urbane 2, a Huawei Watch 2, and an Asus ZenWatch 3. All the considered devices embed magnetometer sensors with comparable accuracy. Participants were asked to wear the smartwatch on their non dominant arm for 6 days, 8 hours per day: from 9 to 13 (A.M. time interval) and from 15 to 19 (P.M. time interval). Users performed the same actions (working or activities of daily living) in all the considered days. The experiment was carried out mainly in home settings, although participants were free to go outside for shopping or working purposes. All the events in which the users crossed the virtual magnetic barrier and the related duration were recorded in a textual log file.

The two monitoring modalities (with and without haptic feedback) were used once per day by each user, following a pseudo-randomly generated sequence. During the post processing phase, potential face-touch events recorded by the app were classified depending on their duration. Indeed, since we could not measure the hand position, we exploited a time threshold to distinguish a Touch Attempt (TA) from a Happened Contact (HC). We experimentally assessed that face-touch attempts last less than a second, then we considered as TA all the recorded events in line with aforementioned short-lasting time. On the contrary, whenever the hand remained in the alerting state for more than a second, the event was classified as a HC, and the exceeding time was recorded as the HC duration. This classification was useful to analyse the effects of the alert (vibration) on subjects' behaviour. In addition, since it is of interest of this work preventing unsafe behaviour and training people to develop good habits, we considered worthy of analysis both number of TAs and number of HCs.

After the conclusion of the experiment, an online anonymous survey was requested to each subject. The survey collected opinions and feedback with a single open-end question proposed within a text box, in which respondents could formulate their own answers in less than 100 words. To highlight the system efficacy in preventing and reducing face-touches, a comparison among the two modalities was carried out by means of a statistical analysis of the data. Multiple paired-samples t-tests were conducted to determine whether there were statistically significant differences between the metrics of interest over the two feedback modalities (V and N). Then, two paired-samples t-tests were performed to determine whether there was a statistically significant effect of the condition on the number of HC/hour and HC duration in the two modalities.

Results In what follows, data are mean $\pm$ standard deviation, unless otherwise stated.

Firstly, *a-posteriori* analysis with paired samples t-test determined that there is a statistically significant mean reduction in number of TA per hour. The assumption of normality was not violated, as assessed by Shapiro-Wilk's test ($p = 0.682$). Unsafe gestures were observed more often when no vibration was provided (25.18 ± 9.03 TA/hour) than in case of vibrotactile notification (17.53 ± 5.10 TA/hour). We verified a statistically significant reduction of 7.65 TA/hour, $t(9) = 2.55$, $p = 0.031 < 0.05$. Fig. 4.4a reports the comparison of mean and standard deviation of Touch Attempts per hour.

For what concerns the HC/hour ratio, no outliers were detected and the assumption of normality was not violated, as assessed by Shapiro-Wilk's test ($p = 0.303$). The number of face-touches was smaller when the haptic feedback was enabled (2.56 ± 1.76 HC/hour) compared to the condition without notification (11.59 ± 6.68 HC/hour). A statistically significant increase of 9.02 HC/hour was confirmed by the test, $t(9) = 4.77$, $p = 0.001 < 0.05$. Such results are also depicted in Fig. 4.4b.

Regarding the analysis of touch duration, data did not pass ShapiroWilk normality test in both condition, as visible in Fig. 4.4c. After squareroot transformation, the normality condition was satisfied (ShapiroWilk normality test $p > 0.05$). The paired-samples t-test assessed a statistically significant difference of face-touch durations between the conditions with the haptic signal enabled and disabled. The duration of face-touches was significantly reduced, from a median value of 5.21 seconds when participants were not notified to 2.03 seconds if the smartwatch vibration was enabled, $t(9) = 2.795$, $p = 0.021 < 0.05$.

In addition, we estimated the efficacy of the system in preventing happened contacts considering the number of gestures that ended with a face-touch (both with and without vibratory notification). Results demonstrated that the 86.3% of unsafe gestures were interrupted in time when the vibrating alert was active.

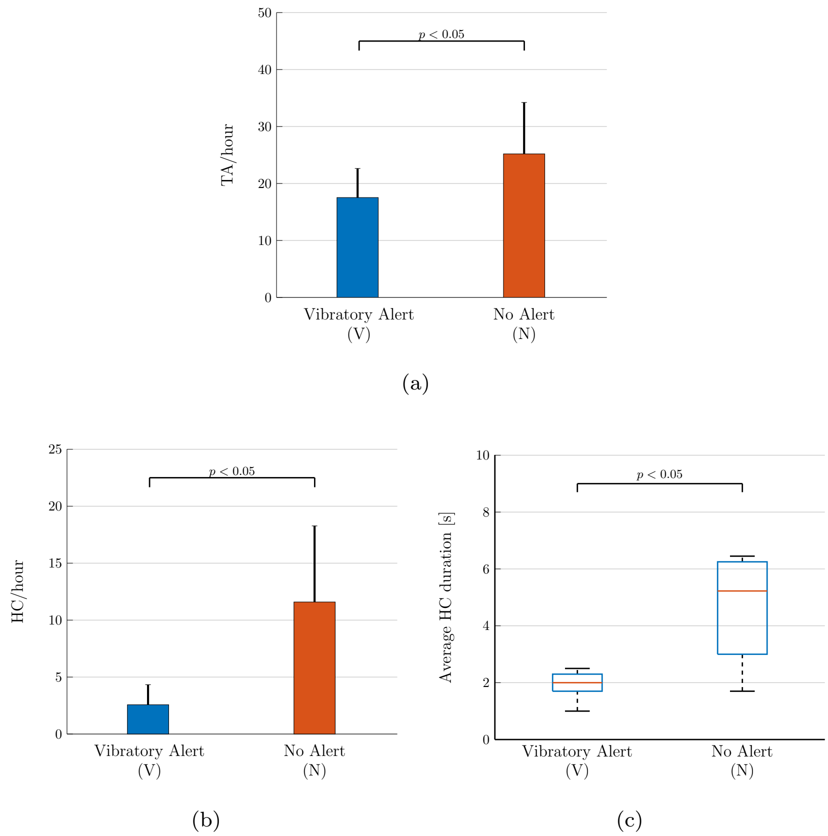


Figure 4.4: Experimental validation results. Mean and standard deviation of the two modalities outcomes are plotted. The p-values, computed with paired-sample t-test, are reported above the bar charts. Number of touch attempts (TA) per hour are reported in (a), while in (b) each bar represents the average amount of happened contacts (HC) per hour. Finally, (c) represents the distributions of HC duration for the two conditions.

This means that after the notification, subjects stopped their gesture in less than 1 s. Longer actions instead were classified as HC.

Qualitative results were derived by analysing partecipants' feedbacks. All the user reported that using the No Face-touch application promotes the sense of perceived safety. Most of them pointed out that, when notified, the gesture was continued only in case the face-touch was strictly necessary, *e.g.* in case of itches. Additionally, users reported that in such cases the gesture was conducted paying greater attention. A negative flaw of the system was the reduction of the smartwatch battery lifetime. Indeed, running the application in background rapidly discharges the device, that needs to be recharged once per day.

4.1.5 Conclusions

With this work we propose a ready-to-use solution to discourage people from touching their face, whose short/mid-term effectiveness has been proven by the experiment results. The proposed system can also help to improve face-touch awareness in patients undergoing habit reversal therapy. On the basis of the literature review presented in the introduction, we can hypothesise that the use of the application might induce a corrective behavior in the long term. We will examine in depth this aspect in a future work.

The presented results pave the way for numerous interesting research directions. For instance, the same approach can be exploited also to limit nail-biting behavioural disorder. In such specific case the wrist orientation range can be further reduced to have a more precise hand position estimation.

One of the great advantages of the proposed framework is that it does not require complex hardware equipment or software implementation. With few efforts in adapting, No Face-Touch can be integrated in a wide group of smart-bands embedding magnetometer sensors, or alternatively, a simple DIY bracelet can be built with off-the-shelf electronic components. Such cheap alternatives would enable the user to wear a pair of devices, one per arm.

4.1.6 Acknowledgments and software repositories

All the software mentioned in this manuscript was developed by the authors and is freely available in a public repository, released under GNU GPL software license. To reach as many people as possible, multiple versions for several platforms were developed. Source codes for Wear OS and watchOS are public[1]. The Tizen app release (for Samsung smartwatches) is still in a beta version and not yet ready for a public dissemination. The Wear OS app release of No Face-Touch is already available on Google Play Store[2]. In Fig. 4.5 a flow chart represents the final No Face-Touch application functioning, which exploits the two algorithms explained in this manuscript. As the app is launched, the user is asked to accept a privacy disclaimer for using the software and then declare which hand is wearing the smartwatch. The software checks if a magnetometer sensor is available, and enables the corresponding algorithm accordingly (see Algorithm 1 and

[1] `https://github.com/sirslab/COVID-19-DoNTYF-wear`
[2] `https://play.google.com/store/apps/details?id=it.unisi.sirslab.covidwear`