Head and Shoulders Gestures: Exploring User-Defined Gestures with Upper Body

Jean Vanderdonckt[1(✉)], Nathan Magrofuoco[1], Suzanne Kieffer[1],
Jorge Pérez[1,2], Ysabelle Rase[1], Paolo Roselli[1,3], and Santiago Villarreal[1]

[1] Université Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium
{jean.vanderdonckt,nathan.magrofuoco,suzanne.kieffer,jorge.perezmedina,
ysabelle.rase,paolo.roselli,santiago.villarreal}@uclouvain.be
[2] Universidad de las Américas, Intelligent & Interactive Systems Lab (SI^2 Lab),
Sede Queri, Av. De los Granados, Quito 170504, Ecuador
jorge.perez.medina@udla.edu.ec
[3] Matematica Dipartimento, Università degli Studi di Roma "Tor Vergata",
Via Orazio Raimondo, 18, 00173 Rome, Italy
roselli@mat.uniroma2.it

Abstract. This paper presents empirical results about user-defined gestures for head and shoulders by analyzing 308 gestures elicited from 22 participants for 14 referents materializing 14 different types of tasks in IoT context of use. We report an overall medium consensus but with medium variance (mean: .263, min: .138, max: .390 on the unit scale) between participants gesture proposals, while their thinking time were less similar (min: 2.45 s, max: 22.50 s), which suggests that head and shoulders gestures are not all equally easy to imagine and to produce. We point to the challenges of deciding which head and shoulders gestures will become the consensus set based on four criteria: the agreement rate, their individual frequency, their associative frequency, and their unicity.

Keywords: Gesture elicitation study · Gesture interaction

1 Introduction

From the human point of view, a gesture is often defined as a communicative movement of the hands and arms which express, just as language, speakers attitudes, ideas, feelings and intentions [2,9]. This early definition focuses on gestures issued by hands and arms, which are certainly among the most mobile human limbs in terms of planes (i.e., the frontal plane along the X axis, the sagittal plane along the Y axis, and the transverse plane along the Z axis - Fig. 1), range of motion (e.g., angle with respect to a standing body), and therefore in terms of expressiveness. It also emphasizes gestures as a mean to support verbal communication (hence, the speaker). Actually, a gesture can be issued theoretically by any human limb, not just the most mobile ones [15]. And a gesture can be typically involved in any verbal or non-verbal mode of communication [18]. This is partially reflected in the
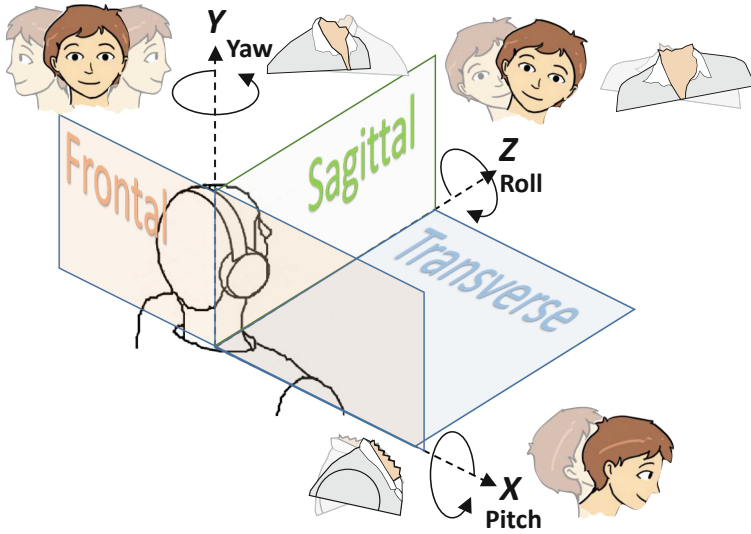
Fig. 1. Transverse, frontal, and sagittal planes for head movements (based on [8]: images by courtesy of T. Jacob, G. Bailly, and E. Lecolinet).

system point of view for gestures: a gesture is considered as any physical movement that a digital system can sense and respond to without the aid of a pointing device such as a mouse or stylus [21]. This definition does not specify what type of response should be given by the system: an object (e.g., a deictic gesture expresses a reference to an object simply by pointing to it), an action (e.g., a gesture translates a human command into an executable function like "turn a TV on"), an attribute of an object or a parameter of an action, a non-verbal information, or any combination of those (e.g., "turn this TV on my favorite channel"). By combining these definitions, we hereby refer to a *gesture* as any movement of one or many human limbs that actually convey a meaning that can be acquired, and hopefully interpreted by an agent, which can be human, software, and/or hardware.

Although our whole body can conduct gestures, they are preferably and frequently issued with our most mobile human limbs, such as fingers (especially for micro-gestures) [3], hands (especially for mid-air gestures) [1], forearms and arms (especially for body-based gestures) [15] because they belong to the most mobile limbs. Gestures issued by the human head and/or the shoulders are a particular category of mid-air gestures that are particularly appropriate in contexts of use where the other human limbs (e.g., fingers, hands, arms, legs) are already busy or cannot be used for other non-physical reasons (e.g., hygienic, social, psychological, cultural interpretations) and prevent from using them for issuing gestures. These situations include: eye-free situations [17] (e.g., driving a car, checking a machine usually require that the driver or operator does not change the locus of attention in fear of loosing control), busy-hands situations

(e.g., in a freezing atmosphere, in an industrial context), stationary situations (e.g., the human body is forced to stay in a fixed position). Head and shoulders gestures offer some movement capabilities below those offered by other gestures, but have a real potential as they occur naturally and may prove less distracting or less demanding than other types of gestures, even if their repertoire of physically possible gestures is narrower than those offered by hands for example.

In order to identify the sub-set of preferred gestures from the set of physiologically possible head and shoulders gestures, we chose to conduct a Gesture Elicitation Study (GES) as a method. This paper reports on the results of conducting this method. The remainder of this paper is structured as follows: Sect. 2 reports work related to head and shoulders from an anatomic and interaction point of views and on major gesture elicitation studies, Sect. 3 defines the experiment conducted, Sect. 4 discusses the results obtained, and Sect. 5 concludes the paper and provides some future avenues for this work.

2   Related Work

This section is divided into three parts: an introduction to the anatomy of head and shoulders, a review of previous work conducted with this mode of interaction, and an overview of existing elicitation studies performed on the human body.

2.1   Anatomy of the Head and Shoulders Movements

Shoulders. According to the field of osteokinematics [4], the shoulder joints offer the following repertoire of possible movements: flexion, extension, hyperextension, abduction, adduction, medial rotation (internal rotation), lateral rotation (external rotation), horizontal abduction, horizontal adduction, and circumduction. For example, flexion occurs in the sagittal plane of motion with respect to the human body, exploits the transverse axis through the center of the humeral head, and have a range of motion between $0°$ and $90°$. Conversely, an extension share the same plane and axis of motion than flexion, but have a more restricted range of motion, situated between $0°$ and $45°$ up to $60°$. Abduction occurs in the frontal plane of motion, along the sagittal axis through the center of the humeral head, and benefits from an extraordinary range of motion: from $0°$ to $175°$ ($0°$ to $60°$ in internal rotation and $0°$ to $90°$ in external rotation). Internal rotation occurs in the transverse plane, along the vertical axis, $0 - 70°$ as the arm at $90°$ of shoulder abduction and $90°$ elbow flexion. External rotation differs from internal rotation only in that it displays a range of motion of $0° - 90°$ as the arm at $90°$ of shoulder abduction and $90°$ elbow flexion. Adduction occurs in the frontal plane with respect to the human body, still along the sagittal axis, but is rapidly constrained by the trunk in its range. Circumduction combines flexion, abduction, extension, and adduction or in the reversed sequence. Consequently, movements at the shoulder joints are interesting as they can occur in every direction (flexion, extension, abduction, adduction, rotation, circumduction), they are considered as highly mobile due to the large size of head of humerus and the

looseness of the capsule of the joint. But arm movements are arrested by contact of the bony surface. This repertoire reveals possible movements of the arm based on the shoulder joints, but does not identify the movements of the shoulder itself. We therefore define a *shoulder gesture* as any movement of the shoulder joint that leaves the rest of the arm unaffected (stationary). A shoulder gesture occurs in any plane of motion (sagittal, transverse, frontal) or direction (forward, backward, or circular) (Fig. 1). *Shrugging* consists in a gestural condition whereby the participant moves one or both shoulders up and/or down.

Head. Similarly, we define a *head gesture* as any movement of the head leaving the rest of the body unaffected (stationary). A head gesture could occur in any plane (sagittal, transverse, frontal) [8]. For instance, a *downhead* gesture, respectively a *uphead* gesture, occurs when a downward, resp. upward, head movement is produced. Head movements are studied in many domains, such as linguistics [18] and body language. Indeed, since our head usually turns towards a scene of interest, it indicates that some object belonging to this scene becomes the primary focus of attention for a number of reasons: we like or dislike something, we feel good or bad about something. The *nod* gesture is often considered as an approval (it means "yes", "I concur", "I agree") or a positive expression of interest for something (it means "I like it", "I am enthusiastic") while a *shaking* gesture is considered as a denial ("I disagree") or a negative expression of interest for something ("I do not like it"). We can recognize other subtle movements subconsciously because they express some feedback. A *slow shaking* gesture reveals disbelief or an expression of uncertainly about a scene being looking at. A *fast shaking* gesture reinforces the message by saying that the negative expression of interest is definitive. Non-command head gestures are different from command gestures in that they are intended to convey some idea, mood, but not an object, an action, or any combination of them. Non-command gestures are often studied in the area of body language. For example, erratic head gestures with frequent eye glances to the sides of the field of view can reveal some discomfort, some tension. A *thrust* gesture consists of a downward gesture in a fixed position expressing readiness for attacking something (like tackling a problem) or somebody (like being confronted with someone). Conversely, a *retreat* gesture consists of a head backward gesture performed in the frontal plane expressing a defense position that is opposite to the thrust. A *head tilt* is performed left or right in the transverse plane: if the head tilts to the right side of a person, the body language interprets this as a person being smart, if the head tilts to the left side, it is interpreted as a person being or willing to be more attractive.

Head and Shoulders. When combined, *head and shoulders gestures* offer the capability to produce gestures that either share the same plane of motion (for example, the head and the shoulders all move in the sagittal plane) or not (for example, the head moves in the frontal plane while the shoulders are moving in the transverse plane). Same for their directions or other parameters. Some types of gestures often occur simultaneously because the human being naturally

associate them: for instance, a shrugging gesture is produced on the shoulders while a downward head gesture is simultaneously issued.

2.2   Interaction Techniques

Head gestures have been mainly employed in combination with eye gaze interaction to designate objects of reference in a scene: head gesture recognition by combining gaze and eye movement [17], head to face input [22], and gaze with head gestures [23]. The only GES dedicated to head and shoulders that we know consisted in eliciting gestures for changing the view of a 3D scene while creating objects in this scene [8]. The consensus gestures resulting from their study were ranked in decreasing order of agreement score [28]: downward and upward head and shoulders gestures for zooming in/out ($A(r) = .8$, very high agreement), downward and upward head gestures for horizontal control ($A(r) = .5$, very high), up/down head gestures for vertical control ($A(r) = .4$, high) and for horizontal orbit ($A(r) = .3$, high), head and shoulders nodding for horizontal panning ($A(r) = .3$, medium), up/down gestures for vertical orbit ($A(r) = .2$, medium) and panning ($A(r) = .19$, medium).

2.3   Overview of Gesture Elicitation Studies

Understanding users' preferences and behavior with new interactive technology right from the early stages of design empowers designers with valuable information to shape a product's characteristics for more effective and efficient use. This process is known as *Gesture Elicitation Studies* (GES) [28–30], which have been popular to understand users' preferences for gesture input for a variety of conditions studied along the three dimensions of the *context of use*:

– **On various platforms and devices**. Since their inception, GES primarily focused on some particular *platform* or *device.* For instance, Wobbrock et al. [30] reported users' preferences for multi-touch input on interactive tabletops. Vatavu [27] and Zaiţi et al. [31] addressed mid-air gesture input to control a TV set. Ruiz et al. [20] investigated users' preferences for motion gestures with smartphones.
– **In different environments**. Gestures are typically elicited in a particular physical and/or psychological *environment* in which devices are determined, such as the steering wheel in a car. Gestures can be also constrained by type, such as hand gestures [1], micro-gestures with one hand only [3] or not.
– **For diverse users**. Some studies are *user-independent* when no particular profile is involved, while some others are *user-dependent*: whole-body gestures [12] are dedicated to a particular type of *users*, e.g., children, thus underlining that the elicitation study can target any particular population of end users instead of platform or environment. Hand gestures [1] , while [6] compared freehand gestures with gestures issued on the skin, thus demonstrating that any particular human ability or physical capability or the deficiency thereof could also become the central subject of a GES.

The GES outcome consists of a characterization of users' gesture input behavior with valuable information for designers, practitioners, and end users regarding the consensus levels between participants (computed as *agreement* [28,30] or *coagreement rates* [29]), the most frequent (thus, generalizable across users) gesture proposals for a given task, and insights into users' conceptual models for performing tasks. The most recent formalization of the elicitation methodology proposed both repeated measures [28] and between-subjects [29] designs.

Virtually, any human limb capable of some mobility can theoretically be the source of a gesture. As a matter of fact, several studies have concentrated their efforts on some human limb in isolation (e.g., the legs), or combined with subsequent limbs (e.g., the legs with the feet), while others considered the human body as a whole, which is of utmost importance for full-body gesture interaction. Hence, the range of investigation starts from any limb in particular until the full body is reached. In the *human gesture continuum*, any gesture starts from any individual limb and evolves to several limbs captured together until the full-body is attained. Based on the human gesture continuum, the human body can be decomposed into one to many gesture types. For instance, the upper-body [16] gesture interaction is decomposed into several limbs that have been subject to GES: the face [22], the head [23], eye-based head gestures [17], the nose [13], the shoulders [8]. Belonging to the upper-body, the human arms are themselves subject to a gesture continuum: fingers [3], wrists [20], hands [1,6,19,30,31], arms [15], and skin-based gestures [6] in general, and from hands to other parts of the body [3]. Lower-body gesture interaction is decomposed into sub-limbs: feet [5], legs until the whole-body gesture interaction [12] is attained.

In conclusion, we motivate a GES on the head and shoulders by the following reasons: it has never been subject to any GES (apart from [8] for a 3D navigation), the gesture set explored insofar is limited to 3D movements in the 3 planes [13]), no qualitative or quantitative analysis has been carried out about the gestures preferred by end users in this case, these gestures are still in their infancy, especially in eyes-free conditions. When combined, head and shoulders gestures offer the capability to produce gestures that either share the same plane of motion (for example, the head and the shoulders all move in the sagittal plane) or not (for example, the head moves in the frontal plane while the shoulders are moving in the transverse plane). Same for their directions.

3   Design Space

Before conducting an experiment, we built a design space of all physiologically possible gestures based on the field of osteokinematics [4] and linguistics [18] (see Sect. 2.1) and the literature about head and/or shoulders gestures [7,8,17,23]. Table 1 defines these gestures based on which plane is maintained constant or left variable. For quick reference, the column 'Alias' gives a unique short name. The first row of Fig. 2 gathers the three first gestures of Table 1: for example, the 'Face left' gesture occurs when the face is maintained in the same plane, while the neck is moving left. The second, resp. third, row of Fig. 2 gathers the

three possible types for tilting, resp. for rotation about each axis. The fourth
row consists of the three possible shoulders gestures occurring when a translation
occurs about each axis. Simple gestures appearing in the four first rows could
form a compound gesture, such as rows five and six: *shrug* (raise left, right, both
shoulders, then lower left, right, both shoulders quickly), *clog left, right* (raise
the right, left shoulder and tilt the head to the left, right), *nod horizontally* (do a
left head, then a right head quickly, possibly repeatedly, so as to express a 'no'),
nod vertically (bend up, then down quickly, possibly repeatedly, so as to express
a 'yes'), *rotate clockwise* (bend up, then right, then down, then left, then up so
as to draw a circle in mid-air), *rotate counterclockwise* (bend up, then left, then
down, then right, then up so as to draw a reverse circle in mid-air), *balance left*
(raise left and lower right), *balance right* (raise right and lower left).

Table 1. Definition of head and shoulders gestures with their physiological movement
(c=constant plane, v=variable plane).

| Head	Label	Alias	Movement (frontal, trans., sagit.)
X translation	Move the head left, right	Face left, face right	Lateral translation (v,c,c)
Y translation	Move the head up, down	Face up, face down	Neck elevation, depression (c,v,c)
Z translation	Move the head forward, backward	Thrust, retreat	Protraction, retraction (c,c,v)
Frontal tilting	Tilt the head to the left, right	Bend left, right	Lateral flexion (v,v,c)
Trans. tilting	Tilt the head up, down	Bend up, down	Extension, flexion (v,c,v)
Saggital tilting	Tilt the head forward, backward	Bend forward, backward	Extension, flexion (c,v,v)
X rotation	Turn the head up, down	Uphead, downhead	Horizontal rotation (c,v,v)
Y rotation	Turn the head left, right	Lefthead, righthead	Vertical rotation (v,c,v)
Z rotation	Turn the head forward, backward	Forehead, backhead	Facial rotation (v,v,c)
Shoulders	Label	Alias	Movement
X translation	Move shoulder horiz. to left, right	Decontract, contract	Extension, flexion (v,c,c)
Y translation	Raise shoulder, lower shoulder	Raise, lower	Shoulder elevation, depression (c,v,c)
Z translation	Move shoulder forward/backward	Protract, retract	Shoulder protraction, retraction (c,c,v)
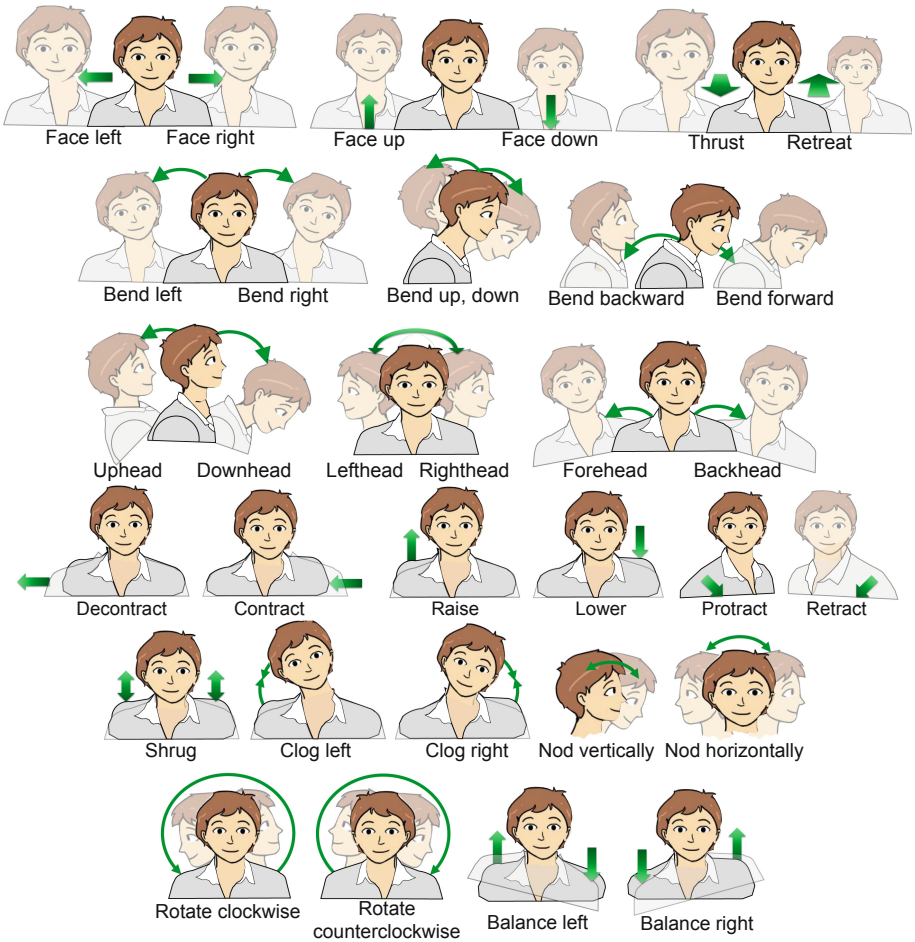
Fig. 2. Design space of head and shoulders gestures (images based on [8]).

4   Experiment

While physiologically possible gestures were identified n the previous section, we
do not know which ones would be naturally suggested by people to issue ges-
tures attached to commands or non-command interfaces. Human preference for
some gestures may be fueled by various factors such as: physical difficulty of the
gesture (all these gestures are submitted to constraints: e.g., the shoulder abduc-
tion is limited by different physical factors such as ligament position, elasticity,
and tightness of the joint), physical ability or disability [16] (e.g., a capsulitis
decrease movements of the shoulder joint), spontaneity to produce a gesture
(some gestures come more naturally than others not because we are less capable
of producing them, but simply because we are more akin to produce them when

thinking about them), fatigue (when the gesture should be repeated), differen-
tiation (how people can easily differentiate one gesture from another), cognitive
load (whether a gesture belongs to the acceptable range of gestures for a user
depending on her cognitive style, traits, or maximal load), memorability (when
the gesture should be remembered after some period of time), reproducibility
(whether we are able to reproduce more or less the same gesture even if me
remember it properly). To identify the preferred gestures from the set of pos-
sible ones (Table 1, we conducted a GES following the methodology originally
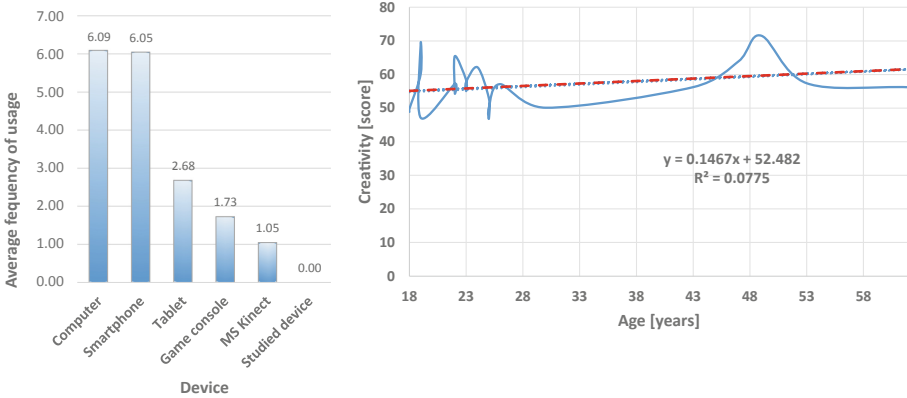defined from the literature [28,30] to collect users' preferences for our gestures.

Fig. 3. Device frequency of usage (a) and Creativity scores (b).

4.1   Participants

Twenty-two voluntary participants (10 Females, 12 Males; aged from 18 to
62 years, $M = 28.95$, $SD = 12.55$) were recruited for the study via a contact
lists in different organizations. Their occupations included secretary, teacher,
psychologist, employee, retirees, and students in domains such as economics,
nutrition, chemistry, history, and transportation. Various usage frequencies were
captured: computer, smartphone, tablet, game console, and Kinect-like device.
All participants reported frequent use of computers and smartphones in daily
life (Fig. 3a). All participants reported that they never saw any head and shoul-
ders interaction before and, therefore, they were not familiar with this kind of
technology.

4.2   Apparatus

The experiment took place in a usability laboratory to keep the control over the
experiment. A simple computer screen was used as a display for showing the
referents to the participants. All the gestures were recorded by a camera placed
in front of the participants to capture their head and shoulders.

4.3   Procedure

Pre-test Phase. The participants were welcomed to the setting by the researchers and were first asked to sign an informed consent form compatible with GDPR regulation. Then, they were given information about the study and the general process of the experiment. They were also asked to fill a sociodemographic questionnaire and to perform a creativity test and a motor-skill test. The researchers collected the sociodemographic data about each participant in order to use some of these parameters in the study. The questionnaire gives general information about the participants (e.g., age, gender, handedness) and asks a series of questions about their use of technologies (based on a 7-point Likert scale ranging from 1 = strongly disagree to 7 = strongly agree). We tested the participants' creativity via http://www.testmycreativity.com/: they were asked to answer a series of questions and received at the end an assessment of their level of creativity. The Motor-skill test [10] was applied to check dexterity.

Test Phase. During this phase, the experimenter explained to participants what nose interaction is all about, the following tasks that they had to perform, and the allowed types of gestures (they should be compliant with the aforementioned definition). Participants operated with the belief that no technological constraint was imposed in order to preserve the natural and intuitive character of the elicitation, such as no restriction on gesture recognition. Each session implemented the original protocol for a GES [30]: participants were presented with referents, i.e., actions to control various objects in an Internet-of-Things (IoT) environment, for which they elicited one gesture to execute those referents, i.e., gestures that fit referents well, are easy to produce and remember. Participants were instructed to remain as natural as possible. The order of the referents was globally randomized per participant based on a pseudo-random number generator (www.random.org). The thinking time between the first showing of the referent and the moment when the participant knew which gesture she would perform was timed by the experimenters. It was measured in seconds with a stopwatch. After eliciting each gesture, the experimenters asked the participant to rate it from 1 to 10 to express to what extent she thought her gesture was appropriate to the presented referent. Each session took approximately 45 min.

Post-test Phase. At the end of each session, the participants were asked to fill in the IBM CSUQ (Post-Study System Usability Questionnaire) [14], which enables participants to express their level of satisfaction with the usability of the setup and the testing process. This 16-question questionnaire is preferred because it has been empirically validated with a large number of participants on a significant set of stimuli, it is widely applicable for any system, and it benefits from a proved $\alpha = 0.89$ reliability coefficient between its results and the perceived system usability [14]. Each closed question is measured using a 7-point Likert scale (1 = strongly disagree, 2 = largely disagree, 3 = disagree, 4 = neutral, 5 = agree, 6 = largely agree, 7 = strongly agree) and summed up

in: system usefulness (SysUse: Items 1-8), quality of the information (InfoQual: 9-15), quality of the interaction (InterQual: 16-18), system quality (Overall: 19).

4.4   Design

Our study was within-subjects design one independent variable: REFERENT, a nominal variable with 14 conditions, representing common tasks to execute in a home environment [27]: (1) Turn the TV On/Off, (2) Start Player, (3) Turn the Volume up, (4) Turn the volume down, (5) Go to the next channel, (6) Go to the previous channel, (7) Turn Air Conditioning On/Off, (8) Turn Lights On/Off, (9) Brighten Lights, (10) Dim Lights, (11) Turn Heating system On/Off, (12) Turn Alarm On/Off, (13) Answer a phone call, and (14) End Phone Call.

4.5   Measures

We employed the following measures to understand users' preferences and cognitive and motor performance for nose gestures:

1. We computed Agreement Scores $A(r)$ [28,30] and Agreement Rates $AR(r)$ [29] for each REFERENT r condition using the formula:

$$A(r) = \sum_{P_i \subseteq P} \left(\frac{|P_i|}{|P|}\right)^2 \geq AR(r) = \frac{|P|}{|P|-1} \sum_{P_i \subseteq P} \left(\frac{|P_i|}{|P|}\right)^2 - \frac{1}{|P|-1} \quad (1)$$

   where r denotes the referent for which a gesture will be elicited, $|P|$ denotes the number of elicited gestures, and $|P_i|$ denotes the number of gestures elicited for the ith subgroup of P.
2. Participants' CREATIVITY was evaluated using an on-line creativity test returning a score between 0 and 100 (higher scores denote more creativity) computed from answers to a set of questions which cover several factors: *abstraction* (of concepts from ideas), *connection* (between things without an apparent link), *perspective* (shift in terms of space, time, and other people), *curiosity* (to change and improve things accepted as the norm), *boldness* (to push boundaries beyond accepted conventions), *paradox* (the ability to accept and work with contradictory concepts), *complexity* (the ability to operate with a large quantity of information), and *persistence* (to derive stronger solutions).
3. Participants' fine motor skills was measured with a standard motor test of the NEPSY test batteries (a developmental NEuroPSYchological assessment) [10]. The test consists in touching each fingertip with the thumb of the same hand for eight times in a row. Higher motor skills are reflected in smaller times.
4. THINKING-TIME measures the time, in seconds, needed by participants to elicit a gesture for a given referent.
5. GOODNESS-OF-FIT represents participants' subjective assessment, as a rating between 1 and 10, of their confidence about how well the proposed gesture fits the referent.

5   Results and Discussion

A total amount of 308 gestures were elicited from 22 participants × 14 referents, which we clustered/classified into groups of similar types according to the following criteria inspired and/or adapted from various sources [19,26,27,30]:

– *Body part*: expresses which human limb is involved (head and/or dominant or non-dominant shoulders).
– *Laterality*: specifies the side(s) involved in the gesture (central, unilateral dominant, unilateral non-dominant, bilateral or a combination).
– *Range of motion*: relates the distance between the position of the human body and the location of the gesture (small, medium, or large).
– *Plane of motion*: specifies which axis/axes are concerned (transverse, frontal, and/or sagittal).
– *Composition*: expresses whether a gesture is simple (only one occurrence is produced) or compound (two or more simple gestures compose the new one).
– *Amount of strokes*: states how many strokes were involved (1, 2, 3 or more).
– *Gesture synchronization*: expresses whether a compound gesture is sequential (simple gestures are produced one after another) or concurrent (simple gestures are produced concurrently).
– *Nature*: describes the underlying meaning of a gesture (a symbolic gesture depicts commonly accepted symbols employed to convey information, such as emblems and cultural gestures; a metaphorical gesture is employed to shape an idea or concept, such as turning an invisible knob; a physical gesture is made when the gesture is produced as if it is physically acting on a real object; an abstract gesture does not convey any particular meaning).
– *Form*: specifies which form of gesture is elicited (stroke when the gesture only consists of taps and flicks, static when the gesture is performed in only one location, static with motion (when the gesture is performed with a static pose while the rest is moving, dynamic when the gesture does capture any change or motion).

Based on the aforementioned measures, the 308 elicited gestures were classified into 10 categories clustered into 3 groups (e.g., 1–4: simple gestures, 5–8: repeated simple gestures, 9–10: combined gestures) (Table 2). Instead of classifying them based on a single property, we preferred to classify them according to three levels of complexity because it enables us to quickly identify which body part is involved and to check whether combined gestures, potentially more complex than simple gestures, are viable alternatives to simple gestures, which are more intuitive in principle. For instance, a repeated gesture avoids introducing another gesture type and a combined gesture builds on previously elicited gestures, thus reducing the amount of simple gestures to remember. These results suggest that central gestures (which do not differentiate the laterality) are more frequently selected since they characterize the 4 most frequent categories covering $245/308 = 80\%$ of gestures, the rest being considered as insignificant. It is also worth to notice that the laterality is also postponed as far as possible: dominance only appears for the fifth category, and only for one shoulder, dominance first ($19/308 = 6\%$), non-dominance afterwards ($14/308 = 5\%$).

Table 2. Definition of gesture categories after classification.

Cat.	Name	Body part	Laterality	Range motion	Comp.	Str.	Nature	Form
1	Single head gesture	Head	Central	Medium	Simple	1	Physical	Stroke
2	Single dominant shoul-der gesture	Shoulder	Unilateral dominant	Small	Simple	1	Symbolic	Static
3	Single non-dominant shoulder gesture	Shoulder	Unilateral non-dom	Small	Simple	1	Symbolic	Static
4	Single bilateral shoulder gesture	Shoulders	Bilateral	Small	Simple	2	Symbolic	Static
5	Repeated similar head gesture	Head	Central	Medium	Simple	≥ 2	Physical	Stroke
6	Repeated similar dominant shoulder gesture	Shoulder	Unilateral dominant	Small	Simple	≥ 2	Symbolic	Static
7	Repeated similar non-dominant shoulder gesture	Shoulder	Unilateral non-dominant	Small	Simple	≥ 2	Symbolic	Static
8	Repeated similar bilateral shoulder gesture	Shoulders	Bilateral	Small	Simple	≥ 4	Symbolic	Static
9	Concurrent compound head and shoulders movement	Head and shoulders	Combined	Medium	Com-pound	≥ 3	Abstract	Dynamic with motion
10	Sequential compound head and shoulders movement	Head and shoulders	Combined	Medium	Com-pound	≥ 3	Abstract	Dynamic with motion

Although repetition appears appealing as it reduces the amount of gestures to remember, it still concerns the least preferred gestures grouped into the "Others" pie $(26/308 = 8\%)$. Figure 4 graphically represents the distribution of elicited gestures across these 10 categories. Single head gestures are the most frequently used $(102/308 = 33\%)$, followed by compound gestures, respectively concurrent $(70/308 = 23\%)$ and sequential $(44/308 = 14\%)$. The second most frequent elicitation concerns gestures involving the head or a combination of the head with both shoulders $(172/308 = 56\%)$. Consequently, the head is reported as the principal source for eliciting head and shoulders gestures. This is confirmed by Fig. 5b,c: the
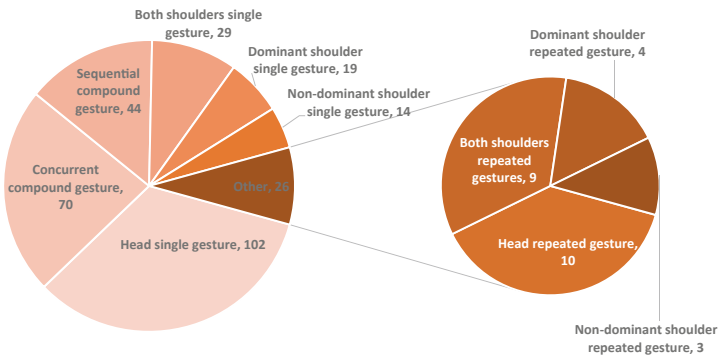

Fig. 4. Distribution of gestures per category.

head alone is involved in 51% gestures, the shoulders alone in 31%, and both in 18%; participants tend to prefer gestures minimizing the amount of strokes with one stroke in 69% of cases, two strokes for 24%, and three or more stroke in the remaining 7% of cases. The lower the physical articulation of gestures is, the more frequent it is. Figure 5a decomposes gestures based on Table 1.
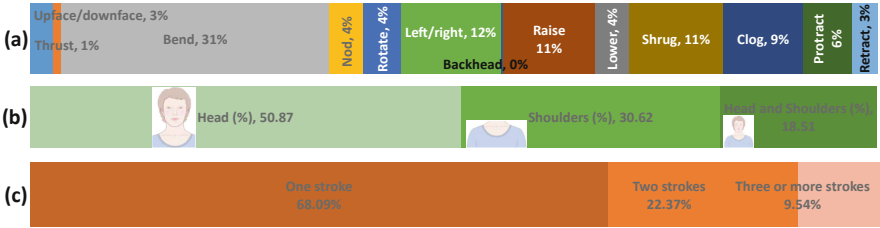

Fig. 5. Breakdown of gestures per criteria.

5.1 Agreement Scores and Rates

Figure 6 shows the agreement scores and rates (Eq. 1) obtained for each REFERENT condition sorted in decreasing order of their rates, along with the final consensus gesture. Several global observations can be made. Firstly, in both agreements, referents often appear in symmetric pairs (e.g., "Go to next and previous channel", "Answer and End Phone Call") or in semantically related ones (e.g., "Play/pause" with "Turn TV On/Off"), which suggests that participants had a higher level of familiarity with some types of referents (after all, changing a channel is a very frequent task) than with others (turning the air conditioning or heating system on/off is considered less frequent or familiar). Secondly, the least agreed referents appeared in these positions because participants were less familiar with physical commands than with popular devices like a television). Thirdly, the ordering of agreement scores and rates remains consistent from one computation to another, except for one pair of referents: "Decrease Volume" was ranked higher according to its score (#7) than for its rate (#9), which suggests that the metrics preserve the ordering apart some particular case. Overall, agreement scores and rates are medium in average magnitude, in particular for rates (which are the most demanding ones) between .104 and .368 for the global sampling ($M = .232$, $SD = .066$). Apart for the "Go to Next/Previous Channel" referents which are ranked with a *high* magnitude, agreement rates belong to the *medium* range according to Vatavu and Wobbrock's method [28] to interpret the magnitudes of agreement rates. These results are very similar to the other rates reported in the GES literature [28]. Hence, our results fall inside medium consensus ($< .3$) category with their average in the same interval (highlighted bars in Fig. 6). To decide the consensus gesture depicted for each REFERENT at the bottom of each bar in Fig. 6, four criteria were successively considered: the agreement rate (blue bars), the individual frequency of occurrence (represented
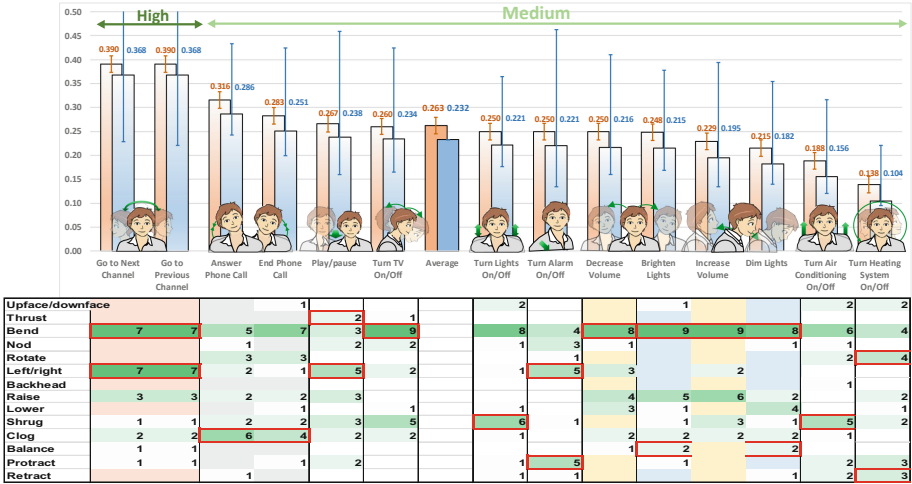
Fig. 6. Agreements Scores and Rates with error bars showing standard error (scores) and $\alpha = .95$ confidence intervals (rates), the consensus gesture, and the frequency table. (color figure online)

by numbers in the bottom part of Fig. 6) for each REFERENT, the associative frequency when two referents are symmetric (e.g., "Go to next channel" and "Go to previous channel" to take into account the consensus by pair, and the unicity of each gesture (whether a gesture was elicited only for one REFERENT). By applying thee criteria, the consensus gestures for each REFERENT are (surprisingly, some common gestures have been suggested by participants, such as nod, but not in a fashion significant enough to warrant any consensus):

– "Go to next and previous channel": with the highest agreement independently from their symmetry, Bend and left/righthead were the most frequent and shared gestures between the two (both were elicited 7 times in Fig. 6). Hence, we decided to assign the **Left/righthead** gesture as symmetric gestures for this referent.
– "Answer Phone Call" and "End Phone Call": they receive the next two highest rates, above the average, and they totalize $6 + 4 = 10$ **Clog** gestures, which has been the most elicited for this pair of referents.
– "Play/pause": **Thrust** has been only elicited for this mode-switching referent and has been preferred as soon as Left/righthead have been already assigned.
– "Turn TV On/Off": **Bend up/down** was by far the most frequently selected gesture for this referent (9).
– "Turn Lights On/Off": **Shrug** was the second most frequently elicited gesture (6) after bending, already assigned (8).
– "Turn Alarms On/Off": **Protract** was the most frequent (6) and uniquely assigned gesture for this referent.
– "Decrease and Increase Volume" (yellow area in Fig. 6): **Bend left/right** totalize 7 elicitations of the same type together.

– "Brighten/dim Lights" (grey area in Fig. 6): **Bend forward/backward** count 7 elicitations of the same type together.
– "Turn AC On/off": **Shrug** was also the most elicited gesture (6) but with maintaining shoulder up as opposed to the complete movement for "Turn Lights On/Off".
– "Turn Heating System On/off": **Rotate clockwise** was the first elicited gesture for this referent.

5.2   Other Measures

Goodness of Fit. Figure 7 distributes the GOODNESS-OF-FIT into six regions depending on its respective value and current interpretation [29]. Overall, the value collected for GOODNESS-OF-FIT for most gestures belong to the "excellent" region ($v > 7$, 12/22=55%) or the "good" region ($v \in [5.5, 7]$, 9/22 = 41%) between 3.36 and 8.14 for the global sampling ($M = 6.78$, $SD = 1.63$). These results are quite above the average values: participants were particularly happy with the gestures they chose and reinforces the acceptability of the elicited gestures. Participant #17 gave the maximum (8.14) and participant #21 was the most severe (3.36). All elicited gestures received an average value between 6.14 and 7.41 (good to excellent range). If we consider the order according to which referents were presented, the GOODNESS-OF-FIT turns out to be usually more positive during the first half of the experiment than during the second part, probably revealing a progressive status of fatigue or boredom. Once could imagine that the most instinctive, spontaneous gestures bring a more important satisfaction among participants. Figure 8 compares the evolution of the GOODNESS-OF-FIT for two randomly selected participants, one with values progressively increasing while the other progressively decreasing. The values do not
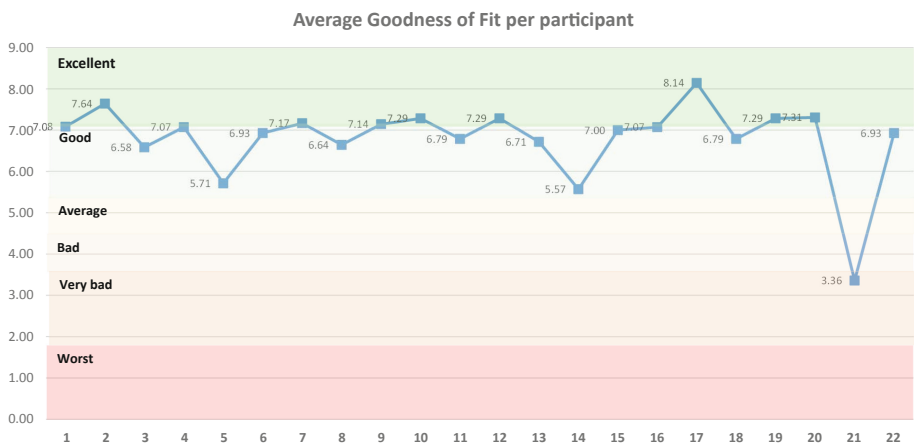

Fig. 7. Average Goodness of fit for all referents per participant.

really depend on the referent, but the order according to which they were presented. Participants were able to quickly find out a fit gesture, but when their source of inspiration was running dry, they elicited less spontaneous, adapted and satisfying gestures.

Thinking Time. Figure 9 compares the average thinking time for all referents with its corresponding agreement rate. Since referents were randomly presented, there is no particular correlation between the thinking time for pairs of related referents. For instance, "Go to previous channel" received the smallest thinking time (2.45 s) while its symmetric referent "Go to next channel" received an average time (9.64 s). Thinking times range between 2.45 s and 22.50 s for "Answer phone call". Contrarily to agreement rate which seems to be linked with the familiarity, the thinking time is apparently not correlated with the referent type. Non-familiar or non-frequent referents do not necessarily receive high times. We did not find any correlation between THINKING-TIME and GOODNESS-OF-FIT. But apparently, the agreement rate decreases when the thinking time increases: the more time a participant may need to appropriately identify a gesture, the lower the agreement rate becomes. We point out in Fig. 9 three referents for which the thinking time was significantly high than the others: "Turn AC on/off" (20.68 s), "Turn heating system on/off" (21.81 s), and "Answer Phone call" (22.50 s). While these three tasks are less frequent than others such as "Play/pause" (15.14 s), it is more their lack of physical reference that harms the time more than their familiarity. The referent "Answer Phone call" is often associated to a physical movement bringing the phone to the ears, which is impossible to achieve in this case. Figure 10 sorts referents in decreasing order of its GOODNESS-OF-FIT along with its correspond thinking time.
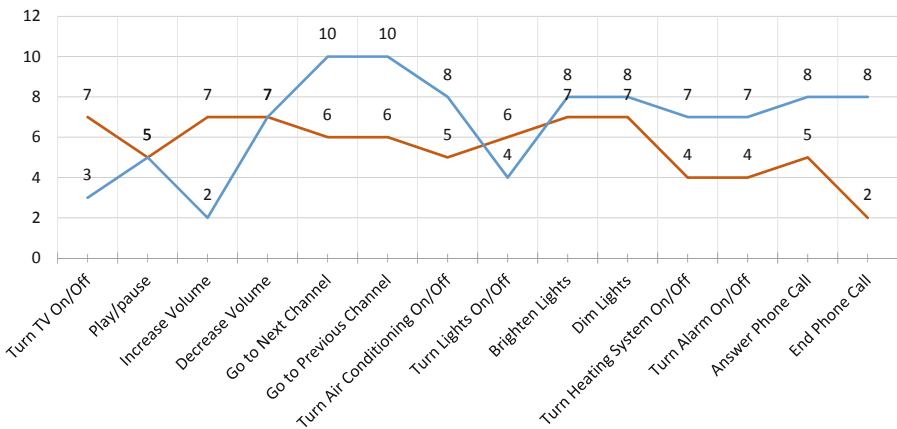

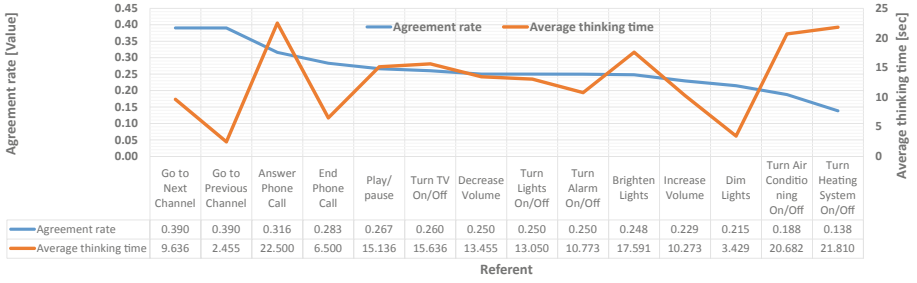**Fig. 8.** Goodness of fit of referents for two participants with different evolutions.

Fig. 9. Agreement rates vs Average thinking time.

User Subjective Satisfaction. Figure 11 reports the results from the IBM CSUQ questionnaire expressing the subjective user satisfaction regarding nose interaction as experienced in this study (error bars show a confidence interval of 95%). First of all, the four CSUQ measures are usually considered as good enough to support the correlation with the perceived usability since their value is superior or equal to 5. Interface quality (InterQual: $M = 5.60$, $SD = 1.20$) exceeds this threshold with the widest standard deviation. System usefulness (SysUse: $M = 5.65$, $SD = 1.08$), Information quality (InfoQual: $M = 5.53$, $SD = 1.17$), and Overall satisfaction (Overall: $M = 5.61$, $SD = 1.14$) all share a value above 5, which suggests that participants were quite subjectively satisfied with head and shoulders interaction, usually more than average. Two reasons could explain this: these gestures are straightforward to imagine (the body language is quite related to some gestures), they are easy to reproduce in a consistent way without
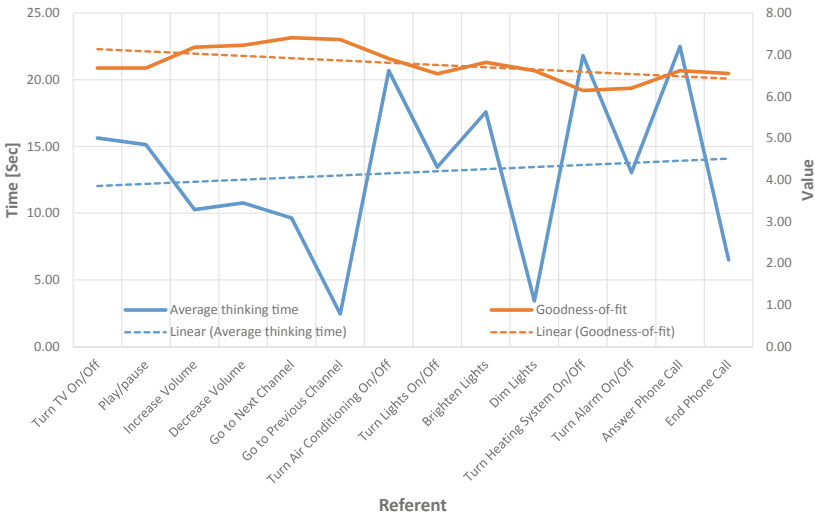

Fig. 10. Thinking time vs Goodness of fit.

endangering recognition. But participants mostly deplored that there is no guid-
ance, no user immediate feedback on how the gestures should be issued and how
they could be recognized and actually triggering an action. Some participants
confessed that they were torn out between the desire to have some guidance
or feedback and the recognition that only the resulting action being executed
should be the only feedback because of discretion. This is also partially reflected
in the individual questions. Questions related to information quality were either
considered as 'not appropriate' (hence, less values are reported in Fig. 11) are
considered positive because of the discretion goal. The questions related to the
other measures all received some agreement. On the other hand, efficiency in
achieving the tasks was recognized to be satisfying (Q3 and Q6 are the most
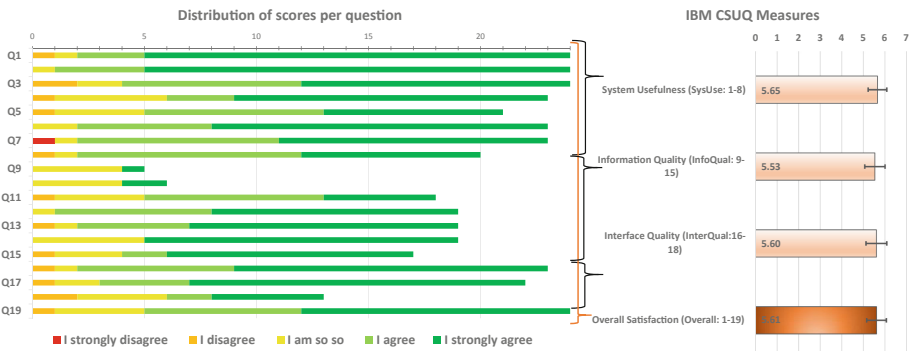positively answered questions).

Fig. 11. Results of the IBM CSUQ questionnaire and measures.

6   Conclusion and Future Work

A gesture elicitation study was conducted with one group of 22 participants who
elicited 308 head-and-shoulders gestures for 14 referents associated to frequent
IoT tasks. These initially elicited gestures were then classified according to sev-
eral criteria to come up with a classification of 10 categories. The final consensus
set consists of 14 hand-and-shoulders gestures reproduced in Fig. 6. Our results
can be summarized as design guidelines that can be easily accessed [24] and
incorporated into a model-based approach [25] to gesture user interfaces:

- Use bending gestures as a first-class citizen: bending gestures of multiple types
  have been elicited almost for every referent as they are probably the easiest
  gestures to (re)produce. Thus, they could be used everywhere, preferably for
  the most frequent tasks that do not involve precise configuration.
- Use Upface/downface for infrequent tasks: these gestures are easy to charac-
  terize, but require some flexion or extension of the neck, which is not desired
  over a long period of time. For example, these gestures were accepted for
  turning on/off the heating and alarm systems.

– Use thrust only for play/pause: this unique gesture works for a well designated task and should not be used for other tasks.
– Forehead and backhead gestures should not be used, apart for exceptional assignation, such as turning the AC on/off, the least frequent task.

This study is limited to its particular conditions (IoT tasks) and participants (random small sampling without representativity). Hence, results are not necessarily generalizable to other contexts of used. It may be hypothesized, however, that different gestures might be elicited by varying such elements as other types of tasks and referents, overall participant posture (e.g., standing vs sitting), repetition and rhythm (several gestures were simply repeated, sometime in a rhythmic way, to augment the vocabulary), or methods of measurement. However, this particular study is not concerned with such possible variables. Rather, its purpose was to come up with a first consensus set of head-and-shoulders gestures based on a design space. This design space could serve for further experiments as it is valid everywhere. Some studies are aimed at examining the musculo-skeletal constraints and the physical fatigue induced by these movements, which is not taken in to account here. Future research may explore other variables that may contribute to elicitation of possible correlations between tasks and gestures.

References

1. Bostan, I., et al.: Hands as a controller: user preferences for hand specific on-skin gestures. In: Proceedings of the ACM International Conference on Designing Interactive Systems (DIS 2018), pp. 1123–1134. ACM, New York (2017). https://doi.org/10.1145/3064663.3064766
2. Bressem, J., Ladewig, S.H.: Rethinking gesture phases: articulatory features of gestural movement? Semiotica **184**, 53–91 (2011). https://doi.org/10.1515/semi.2011.022
3. Chen, Z., et al.: User-defined gestures for gestural interaction: extending from hands to other body parts. Int. J. Hum. Comp. Interact. **34**(3), 238–250 (2018). https://doi.org/10.1080/10447318.2017.1342943
4. Drake, R., Vogl, W., Mitchel, A.W.M.: Gray's Anatomy for Students, 4th edn. Elsevier, Amsterdam (2019)
5. Felberbaum, Y., Lanir, J.: Step by step: investigating foot gesture interaction. In: Proceedings of the ACM International Working Conference on Advanced Visual Interfaces (AVI 2016), pp. 306–307. ACM, New York (2016)
6. Havlucu, H., Ergin, M.Y., Bostan, İ., Buruk, O.T., Göksun, T., Özcan, O.: It made more sense: comparison of user-elicited on-skin touch and freehand gesture sets. In: Streitz, N., Markopoulos, P. (eds.) DAPI 2017. LNCS, vol. 10291, pp. 159–171. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58697-7_11
7. Hirsch, M., Cheng, J., Reiss, A., Sundholm, M., Lukowicz, P., Amft, O.: Hands-free gesture control with a capacitive textile neckband. In: Proceedings of the ACM International Symposium on Wearable Computers (ISWC 2014), pp. 55–58. ACM, New York (2014). https://doi.org/10.1145/2634317.2634328

8. Jacob, T., Bailly, G., Lecolinet, E.: A study on 3D viewpoint control through head and shoulders motion. In: Proceedings of the 27th International Conference on Interaction Homme-Machine (IHM 2015), Article 25. ACM, New York (2015). https://doi.org/10.1145/2820619.2825005

9. Kendall, M.G., Smith, B.B.: The Problem of m Rankings. Ann. Math. Stat. **10**(3), 275–287 (1939). http://www.jstor.org/stable/2235668

10. Korkman, M., Kirk, U., Kemp, S.: NEPSY: A Developmental Neuropsychological Assessment. Psychological Corporation, San Antonio (1998)

11. Kühnel, C., Westermann, T., Hemmert, F., Kratz, S., Müller, A., Möller, S.: I'm home: defining and evaluating a gesture set for smart-home control. Int. J. Hum. Comput. Stud. **69**(11), 693–704 (2011). https://doi.org/10.1016/j.ijhcs.2011.04.005

12. Lee, D.Y., Oakley, I.R., Lee, Y.R.: Bodily input for wearables: an elicitation study. In: Extended Abstract of Proceedings of International Conference on Human-Computer Interaction Korea 2016 (HCI Korea 2016), pp. 283–285 (2016). https://www.dbpia.co.kr/Journal/ArticleDetail/NODE06645483

13. Lee, J., Yeo, H.-S., Starner, T., Quigley, A., Kunze, K., Woo, W.: Automated data gathering and training tool for personalized "Itchy Nose". In: Proceedings of the 9th International Conference on Augmented Human Conference (AH 2018), Article 43. ACM, New York (2018)

14. Lewis, J.R.: IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. Int. J. Hum. Comput. Interact. **7**(1), 57–78 (1995). https://doi.org/10.1080/10447319509526110

15. Liu, M., Nancel, M., Vogel, D.: Gunslinger: subtle armsdown mid-air interaction. In: Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST 2015), pp. 63–71. ACM, New York (2015). https://doi.org/10.1145/2807442.2807489

16. Malu, M., Chundury, P., Findlater, L.: Exploring accessible smartwatch interactions for people with upper body motor impairments. In: Proceedings of the ACM International Conference on Human Factors in Computing Systems (CHI 2018), Paper 488. ACM, New York (2018). https://doi.org/10.1145/3173574.3174062

17. Mardanbegi, D., Hansen, D.W., Pederson, T.: Eye-based head gestures. In: Proceedings of the ACM Symposium on Eye Tracking Research and Applications (ETRA 2012), pp. 139–146. ACM, New York (2012)

18. McClave, E.Z.: Linguistic functions of head movements in the context of speech. J. Pragmatics **32**(7), 855–878 (2000). https://doi.org/10.1016/S0378-2166(99)00079-X

19. Piumsomboon, T., Clark, A., Billinghurst, M., Cockburn, A.: User-defined gestures for augmented reality. In: Kotzé, P., Marsden, G., Lindgaard, G., Wesson, J., Winckler, M. (eds.) INTERACT 2013. LNCS, vol. 8118, pp. 282–299. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40480-1_18

20. Ruiz, J., Li, Y., Lank, E.: User-defined motion gestures for mobile interaction. In Proceedings of the ACM International Conference on Human Factors in Computing Systems (CHI 2011), pp. 197–206. ACM, New York (2011). https://doi.org/10.1145/1978942.1978971

21. Saffer, D.: Designing Gestural Interfaces. O'Reilly Media, Sebastopol (2008)

22. Serrano, M., Ens, B.M., Irani, P.P.: Exploring the use of hand-to-face input for interacting with head-worn displays. In: Proceedings of the 32nd ACM International Conference on Human Factors in Computing Systems (CHI 2014), pp. 3181–3190. ACM, New York (2014). https://doi.org/10.1145/2556288.2556984

23. Špakov, O., Majaranta, P.: Enhanced gaze interaction using simple head gestures. In: Proceedings of the ACM International Conference on Ubiquitous Computing (UbiComp 2012), pp. 705–710. ACM, New York (2012). https://doi.org/10.1145/2370216.2370369

24. Vanderdonckt, J.: Accessing guidelines information with Sierra. In: Proceedings of IFIP International Conference on Human-Computer Interaction (Interact 1995), pp. 311–316. IFIP (1995). https://doi.org/10.1007/978-1-5041-2896-4_52

25. Vanderdonckt, J.: A MDA-compliant environment for developing user interfaces of information systems. In: Pastor, O., Falcão e Cunha, J. (eds.) CAiSE 2005. LNCS, vol. 3520, pp. 16–31. Springer, Heidelberg (2005). https://doi.org/10.1007/11431855_2

26. Vanderdonckt, J., Roselli, P., Pérez-Medina, J.L.: !FTL, an articulation-invariant stroke gesture recognizer with controllable position, scale, and rotation invariances. In: Proceedings of the ACM International Conference on Multimodal Interaction (ICMI 2018), pp. 125–134. ACM, New York (2018)

27. Vatavu, R.-D.: User-defined gestures for free-hand TV control. In: Proceedings of the 10th European Conference on Interactive TV and Video (EuroITV 2012), pp. 45–48 (2012). https://doi.org/10.1145/2325616.2325626

28. Vatavu, R.-D., Wobbrock, J.O.: Formalizing agreement analysis for elicitation studies: new measures, significance test, and toolkit. In: Proceedings of the 33rd ACM Conference on Human Factors in Computing Systems (CHI 2015), pp. 1325–1334. ACM, New York (2015)

29. Vatavu, R.-D., Wobbrock,J.O.: Between-subjects elicitation studies: formalization and tool support. In: Proceedings of the 34th ACM Conference on Human Factors in Computing Systems (CHI 2016), pp. 3390–3402. ACM, New York (2016)

30. Wobbrock, J.O., Morris, M.R., Wilson, A.D.: User-defined gestures for surface computing. In: Proceedings of the ACM International Conference on Human Factors in Computing Systems (CHI 2009), pp. 1083–1092. ACM, New York (2009)

31. Zaiţi, I.-A., Pentiuc, Ş.-G., Vatavu, R.-D.: On free-hand TV control: experimental results on user-elicited gestures with leap motion. Personal Ubiquit. Comput. **19**(5–6), 821–838 (2015). https://doi.org/10.1007/s00779-015-0863-y