A Method for Evaluating Animal Usability (MEAU)

Luisa Ruge
The Open University
Milton Keynes, UK
luisa.ruge@open.ac.uk

Clara Mancini
The Open University
Milton Keynes, UK
clara.mancini@open.ac.uk

ABSTRACT

Animal Computer Interaction, aims to design user centered interactions that result in good user experiences (UX). During evaluation, the quality of the UX is assessed by measuring the degree to which the interaction between the user and the artefact meets the users' needs and preferences, as evidenced by their behavior. A key measure of the UX is usability. When evaluating usability for different species, ACI researchers face two major challenges: the differences in cognitive, physical and sensory capabilities between human evaluators and animal users, with the implications these differences have for assessing the users' behavior; and the human-centric focus of most usability evaluation methods currently available. To address these challenges, this paper proposes a Method for Evaluating Animal Usability (MEAU), here tailored to Mobility Assistance Dogs as the users, and illustrates its application during a study that compared the canine usability of different access controls.

Author Keywords

User centered design; canine usability; mobility assistance dogs; animal-computer interaction; animal-centered design.

ACM Classification Keywords

H.5.2. User-centered design

INTRODUCTION

Since calls for the systematic development of Animal-Computer Interaction (ACI) as a discipline [1], various approaches have been proposed to address the core aims of ACI [2], including the development of animal-centric design tools for the investigation and evaluation of animals' interactions with technological artefacts. However, when measuring the usability of interactions between animals and technology [3-9], ACI research is still limited and unsystematic, with most measures of usability focusing on single principles, goals or metrics [4, 5, 7]; or relying heavily on anecdotal data [8, 10, 11]. In consequence, there is still a need to adapt and develop methods that can systematically and reliably measure and evaluate animal usability, and effectively address the major challenges faced by ACI researchers. These challenges stem from differences in sensory, cognitive, and physical capabilities between human evaluators and animal users, and their implications for assessing usability; and from the historically human-centric focus of usability evaluation methods [12-14], and the resulting human-centric measures of usability.

To help address these challenges, the research reported here is concerned with the adaptation and development of methods to assess animals' UX, i.e. the quality of the interaction between a user and a technological interface within a given context and the extent to which the users' needs, and preferences are met during said interaction. Here we focus on evaluating a key aspect of UX, that is usability, *"the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use."* [14]. Evaluation enables designers to assess a design to find whether this meets requirements that have already been identified, to identify requirements that have not yet emerged and to establish what changes need to be made so that requirements are met [15]. Evaluation is therefore an essential activity in the interaction design cycle.

While aiming to design and evaluate user-centered interactions more broadly, this research focuses on studying and designing for the interaction of Mobility Assistance Dogs (MADs) with the technological interfaces that constitute part of their working environment. MADs are especially trained to perform some of the functions and tasks that individual humans cannot perform as a result of some disability. Tasks may range between assisting with self-care, mobility and other physical tasks, including opening doors, retrieving objects and switching lights [16, 17]. The study of MADs' interaction with technological artefacts is of particular interest within ACI due to usability and UX issues that typically result from the mismatch between MADs' user characteristics and the anthropocentric environments in which they have to operate. This mismatch affects the dogs' learning during training, their ability to consistently and successfully assist their human partners once deployed and their overall performance and welfare [11].

In the remainder of this paper, we discuss related work within ACI which has informed the development of the Method for Evaluating Animal Usability (MEAU). We introduce the method in phases, including: the systematic definition of the

ACI'19, November 12–14, 2019, Haifa, Israel © 2019 Copyright is held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 978-1-4503-7693-8/19/11…$15.00 https://doi.org/10.1145/3371049.3371060

interaction under evaluation and the identification of relevant user and interface characteristics; the reinterpretation of interaction design principles and goals from a canine perspective and their systematic application to measure canine usability; and resulting measures of usability as a means to assess the interaction and inform future design iterations. We demonstrate the application of the proposed method to the analysis of data collected during a study that took place at the MAD training facilities of UK Charity Dogs for Good [18] and that compared the usability of different access controls tested by MADs. We show how it is useful to measure canine usability, thus supporting the assessment of canine UX. Finally, we discuss the scope and limitations of the method and how it could be applied to usability studies involving the use of interactive technologies by other species.

RELATED WORK

ACI researchers have been adapting a range of methods grounded in disciplines related to the study of animals, design and technology as a means to investigate, theorize and design animal-centered interactions, including those that aim to assess the animals' UX. However, when measuring usability, most of these methods do not support the comprehensive analysis of user-centered criteria that could affect an animal's UX when interacting with technology. For example, in his early work Resner [10], applied interaction design principles and usability goals to inform the design of a human-canine remote communication system. However, the evaluation of the system was based on dog owners' feedback rather than driven by a systematic analysis of how the design conformed to those principles and goals. Robinson *et al.* [8] conducted a broader usability evaluation, considering a range of criteria when designing a canine alarm that would allow medical alert dogs to call for help on behalf of their incapacitated assisted humans. However, the authors' analysis was based on on qualitative data, presented in an anecdotal style, which limited the validity of their findings. Similarly, the work of Mancini *et al.* [11, 19] dealt with the implementation of interactions design principles relevant to dogs as a means to inform canine working environments. However, the authors' assessment of their dog friendly controls' usability relied on reports from the dogs' handlers and partners, rather than on more objective measures of the dogs' behavior as they were interacting with the controls.

Baskin and Zamansky [4]'s study of canine UX during an interaction with two digital games presented in a tablet, the authors assessed the behaviors exhibited by their participants against the canine ethogram (a description of a species' behavioral repertoire). While their aim was to gauge the dog's emotional response to the audio-visual stimuli coming from the tablet, specific usability factors that might have influenced the dogs' experience were not accounted for in the study. Similarly, Westerlaken and Gualeni [20] used a feline ethogram to measure and interpret cats behavior during their interaction with digital games. The behaviors were measured to establish user requirements and inform the design of good animal UX, but the ethogram was used to identify general preference behaviors (i.e. rejection, interest, extra interest) with no reference to specific usability criteria. A similar focus on preference behaviors was also in the work of Pons *et al.* [6], in which the authors aimed to recognize and automatically classify postures exhibited by feline users during their interaction with robots, in order to inform the design of adaptive playful systems. The authors' system identified a number of user behaviors, characterizing the cats' states during the interaction, but again the identification of usability criteria against which to interpret the cats' behaviors was outside the scope of the study.

In the work of Zeagler *et al.* [7] there was a clear focus on usability criteria, as the authors systematically investigated the interaction design principles of *perceivability* and *affordance*, in relation to canine users' characteristics and their interactions with touchscreen interfaces. However, the narrow focus on only two interaction design principles limited the scope of the authors' evaluation. A more systematic analysis of canine behavior with specific reference to usability was conducted by Jackson *et al.* [9] during an evaluation of wearable communication interfaces for working dogs. Here, the behaviors of the canine participants during their interactions with the devices were systematically recorded and quantitatively analyzed. However, the authors' analysis is almost exclusively based on quantitative data, with little attention to the qualitative assessment of the impact of the devices on the overall canine experience when communicating with wearable devices. In contrast, Byrne *et al.* [21]'s evaluation of haptic interfaces for dogs took a more comprehensive approach to measuring canine UX. The authors systematically took into account canine users' physical, cognitive and sensory characteristics, providing design guidelines that were directly relevant to usability, although they did not explicitly set out to measure this aspect of canine UX

Overall, while the above research tends to take into account a range of aspects related to usability and UX, more holistic work tends to yield anecdotal findings; while systematic work tends to focus on a limited number of usability metrics. In order to support the design of interactive systems that measurably afford good usability and UX for animal users, we propose an evaluation method that reframes the evaluative process to specifically account for the user's physical, cognitive and sensory characteristics. The method supports the systemic analysis of experimental data as a comprehensive measure of canine usability.

METHOD FOR EVALUATING ANIMAL USABILITY (MEAU)

The proposed Method for Evaluating Animal Usability (MEAU) aims to: 1) develop a protocol to evaluate animal usability for interactions with technological artefacts informed by relevant animal user and interface characteristics; 2) (re)interpret existing interaction design principles and goals to identify animal-centric requirements against which to evaluate the usability of technological artefacts for animal users (in this case canines); and 3) identify a process for evaluating animal usability that explicitly acknowledges existing differences between human evaluators and animal users.

The MEAU's development was informed by interaction design and animal computer interaction theory. Its stages mirror a typical interaction design evaluation process to maximize the robustness of the development and application process, and to facilitate future iterations and revisions of the method against a well-established blueprint. In this paper, we demonstrate the systematic application of the method, which produced a detailed record of the grounded decisions made throughout the evaluative process. The MEAU involves seven distinct stages; a schematic of these stages and their key aims are shown in Figure 1.

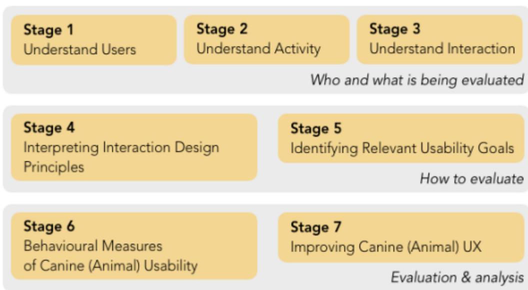

Figure 1: MEAU stages and key aims

1 - Understanding the user

When evaluating interactive products, taking the user as the central unit of analysis is arguably essential to ensure that users' needs and preferences are met [13, 14, 22]. This means taking into account the their physical, cognitive and sensory characteristics to try and comprehend the users' perspective and possible experience during an interaction. To raise awareness of the sensory, physical and cognitive differences between human evaluators and animal users, this phase entails carrying out a comprehensive review of the intended users' (in this case dogs') physical, cognitive and sensory characteristics (See Figure 2).

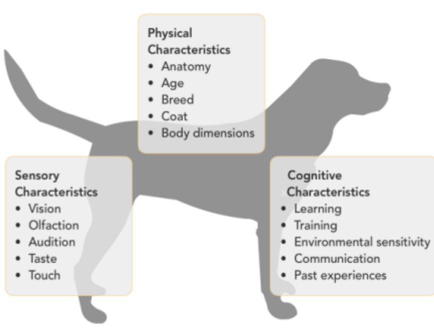

Figure 2: Understanding the user: a canine-centric approach

2 - Understanding the activity

This phase involves gaining an understanding of the interaction to be evaluated. In our case this was the use of access controls by MADs, specifically the use of a standard issue access control (commonly used to operate motorized doors) and two canine-friendly prototype controls (modeled on Mancini *et al.* [11]'s design and described later on).

In order to model the interaction in question, our method involves the development of *use cases*. Commonly used by interaction designers to define the single use of a system from the perspective of a user [14, 23], use cases help designers to model the tasks that need to be supported by the system being designed. Among them, *essential use cases* are simplified, abstract and structured use cases that capture the intentions of a user in relation to the system, whose function is to support a given task or interaction [24]. Essential use cases can capture the essential aspects of an interaction, thus mitigating human researchers' assumptions regarding the animal user (in our case, MADs) and their interaction with the system (in our case, access controls), and can enable them to focus on the essential aspects of the interaction. They can also help to identify specific user behaviors, which the system would need to support in order for the interaction to be successful and which researchers need to focus on during evaluation. Figure 3 shows the use case related to our study, highlighting the essential components of the interaction (user intention and system responsibility), the actions that the user carries out to fulfil their intention and the way in which the system supports them; and the relevant behaviors and activities exhibited by the users and by the system, which should be focused on during data collection.

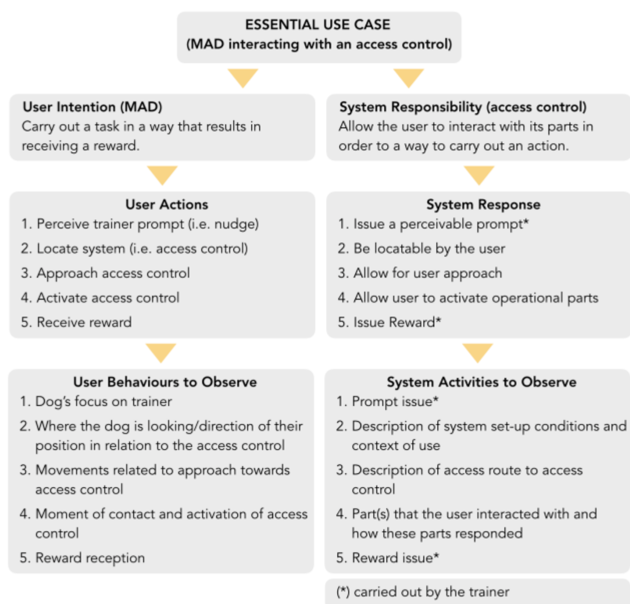

Figure 3: Essential use case of MAD's interaction with an access control

3 –Understanding the interaction

Albeit focused on humans, Hekkert and Schiffersten [25]'s model of user-product interaction (Figure 4) provides a detailed breakdown of the systems and properties that support a users' interaction with a product. When designing for a nonhuman user group, whose characteristics are very different from those of human researchers, this model can help evaluators to systematically identify the users' physical, sensory and cognitive characteristics, and the products' properties, functionalities and features that are relevant to the interaction under evaluation.

Figure 4. Hekkert and Schifferstein [25]'s model of user-product interaction

According to the model, the user is supported by the systems that allow them to interact with a product or environment (motor system, sensory system, cognitive system and instincts), and the capabilities used to make sense of the interactions (motor skills, sensitivities, cognitive skills and concerns). Meanwhile, products present formal properties (structural properties, materials, composition, technology and labels), which during the interaction acquire meaning (sensory properties, possibilities for behavior, and functionality) informed by the users' support systems [25]. In our case, when considering a product such as an access control intended for canine use, working through Hekkert and Schiffersten [25]'s model helped us to focus on the fact that, in order to use the control, MADs would need to make use of their sensory system to perceive the interface's parts; their motor system to activate the control; and their cognitive system to make sense of and carry out the task of activating the control. In turn the access control would have to provide dogs with a way of physically operating it; to be made of appropriate materials for its intended use (i.e. dog snout or paw); to be located at the appropriate distance from the door to ensure safe door opening and closing times; and to perform the function of opening the door.

4 – Interpreting interaction design principles

Interaction designers commonly make use of a set of interaction design principles to help them identify the essential requirements a product should meet in order to ensure that a user can make sense of, and interact with it [12, 13]. If a product fails to adhere to said principles, it will not support its intended use and result in a poor UX.

In their work, Mancini *et al.* [11] discuss the applicability of the most common interaction design principles when designing for dogs. Among the ones they regard as most fundamental, the authors' include: *perceivability* (how the elements of an interface are detectable by the sensory capabilities of the user); *consistency* (how the function, organization and appearance of the elements of an interface present regularities and similarities, both within its parts and in relation to similar products); *affordance* (how the form of an interface's elements suggest the way it could be interacted with); and *feedback* (how a system lets the user know that the effect of their interaction with it was successful, particularly important where there is a space-temporal distance between the user's input and the interaction's outcome). The authors also discuss *mapping* (a form of consistency between the presentation of a function on an interface and its outcome) and *constraints* (whereby a system prevents a user from engaging with an object in ways that are fruitless); however, they hypothesize that these two principles might play a less fundamental role when designing for animals.

When designing for animal users, the principles need to be reinterpreted based on the targeted species' (i.e. canines) characteristics, to ensure that species-specific requirements are identified. For example, the principle of *perceivability*, which among other things relates to vision, is relevant to the interaction with access controls since MADs are expected to be able to see the control and identify its parts when they approach it. However, as human evaluators, we need to take into account the visual capabilities of dogs (which differ from ours) to ensure that they can actually perceive the control. Vision comprises several distinct aspects, such as light, motion and color sensitivity as well as visual acuity and perspective [26]. When it comes to visual perspective, a dog's field of vision is directly related to the height and location of their eyes, and their overall facial bone structure [26]. Most dogs have an average field of vision of 240°, against a human field of vision of 180° [26]; while this allows dogs to better scan the horizon, it reduces their visual acuity, estimated at 20/75; so, a static object that a human with normal vision can see from 75 feet, a dog can only see from 20 feet [27]; at the same time, dogs can focus on objects within 50-33cm of their eyes, but closer objects appear blurred to them [27]. Dogs have dichromatic vision, meaning that they only have two types of cone photoreceptors (blue and yellow) as opposed to the three found in humans (blue, red and yellow), making reds and greens harder to detect [27]. Once the principle of *perceivability* is considered in relation to the canine user's characteristic of visual perspective acuity and color sensitivity, the resulting requirement for our intended user is as follows: *in order for dogs to perceive the elements of the control they need to interact with, these should be clearly detectable at a distance of 50-33cm or, if closer, they should be detectable even if blurry, and should feature colors in the yellow-blue spectrum to support discrimination against other background objects.*

Perceivability also relates to auditory capabilities, which depend on aspects such as the shape and size of the auditory organ, the frequency of vibrations and the capacity to differentiate sounds [28]. In dogs, although differences in hearing between breeds exist, most dogs have an upper frequency limit of 41–47 kHz, much higher than that of humans, which is 15–20 kHz. Dogs peak auditory sensitivity is in the range of 4-8 kHz, and their low frequency limit is 67Hz [29]. This is relevant to the evaluation of our controls, which produce a clicking sound upon activation. Thus, the resulting audibility requirement is as follows: *in order for dogs to be able to comfortably perceive the elements of the control they need to interact with, any sounds emitted by the controls should be within the 8 kHz-67Hz acoustic range with a frequency no higher than 47kHz.*

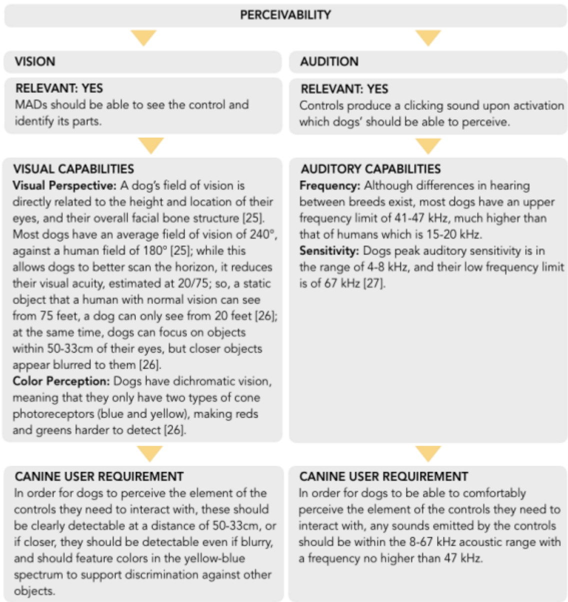

Figure 5. Example of the design principles model

In order to reinterpret interaction design principles from a canine perspective in a systematic and focused manner, we developed a Design Principles Model. The model illustrates the interaction design principles that we deemed relevant (*perceptibility*, *feedback* and *affordance)* to the system under evaluation (i.e. access controls); followed by, as per Hekkert and Schiffersten [25]'s model, the capabilities a dog needs to use to interact with the system (sensory system: vision, hearing, olfactive, touch; motor system: motor skills; cognitive system: learning, training, environmental sensitivity, communication, and past experiences), leading to the formulation of canine user requirements. Given the simplicity of the interaction we were evaluating, we did not deem the principles of *mapping* and *constraints* relevant; the former because the controls' input mechanism was not isomorphic to the output, and the latter because the controls allowed for only one input modality which was expected to be available at all times. Figure 5 provides an example of the Design Principles Model and how it helps reinterpret the principle of *perceivability* as it relates to a dog's visual and auditory capabilities, and to formulate the resulting canine user requirement.

5 – Identifying relevant usability goals

Alongside interaction design principles, designers commonly refer to interaction design goals. Unlike universal interaction design principles, goals address specific tasks within specific contexts of use. They can be formulated as questions relating to an interface's usability and user experience requirements [14]. In particular, the function of usability goals is to optimize the interaction between users and interfaces in order to best support the activities for which the systems have been designed [13, 14, 30]. While failing to meet relevant usability goals does not make a product unusable, it does significantly affect its usability and in consequence the UX. Usability goals most widely referred to include: *effectiveness* (the extent to which a product is good at doing what it is expected to do); *efficiency* (the extent to which a product enables users to complete a task quickly); *safety* (the extent to which a product prevents errors and enables error recovery); *utility* (the extent to which a product provides the required functions for a task); *learnability* (the extent to which a product is easy for the user to learn how to use); and *memorability* (the extent to which a product is easy for the user to remember how to use) [14].

Depending on the interaction being evaluated, and the characteristics of the user and the environment, different goals may need to be prioritized [14]. For example, with regards to the specific interaction evaluated as part of our study, we did not deem the goals of *learnability* and *memorability* relevant for the study. The former because our canine participants were already trained to operate access controls and the latter because of the short duration of the study. However, *efficiency*, *and safety* were deemed relevant usability goals against which to evaluate the controls, since MADs need to be able to operate them quickly and without error.

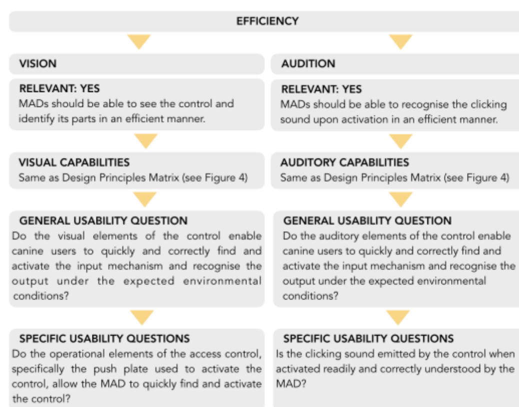

Figure 6. Example of the usability goals model

As with the application of interaction design principles, our method formulates questions relating to usability goals informed by the canine users' characteristics and the task for which the controls are designed. Thus, for our study, we developed the Usability Goals Model, in which the relevant interaction design goals (*efficiency*, and *safety)* were situated on the horizontal axis, followed by the relevant canine capabilities on the vertical axis, leading to the formulation of general and specific usability questions. For example, the goal of *efficiency* relates, among other things, to a dog's vision and hearing capacities. Thus, in regard to canine vision and hearing capabilities, the general usability question related to efficiency would be: *do the visual and auditory elements of the control enable canine users to quickly and correctly find and activate the input mechanism and recognize the output under the expected environmental conditions?* Further articulating the initial general questions into more specific ones, enables researchers to hypothesize possible answers to the questions and define specific criteria for the evaluation of the design in question [14]. In our case, more detailed questions included: *do the operational elements of the access control, specifically the push plate used to activate the control, allow the MAD to quickly find and activate the control? Is the clicking sound emitted by the control when activated readily and correctly understood by the MAD?* Figure 6 provides an example of the Usability Goals Model and with regards to the goal of *efficiency;* how the model helps researchers identify the relevance of the goal and the relevant user characteristics (vision and hearing) involved in the interaction with the interface, and the formulation of general and specific usability questions related to the systems' ability to support the interaction, and thus the usability goal.

6 – Behavioral measures of canine usability

While the design of our prototypes conformed with interaction design principles and canine characteristics, we wanted to empirically evaluate how and to what extent our design met canine usability goals. Empirical evaluation is an important part of the interaction design process; especially so when designing for another species. Given the sensory, physical and cognitive differences between human evaluators and animal users, and the biases that inevitably influence designers' expectations, the assessment of animals' responses against empirical measures is key. In this regard, the accurate analysis and interpretation of data collected during empirical studies with animals is paramount [31, 32]. To minimize the impact of the human evaluator's bias on the interpretation of the animals' behavior, our method applies various data collection metrics and design compliance rating scales to help designers assess the extent to which the interaction being evaluated adheres to relevant usability goals.

In order to define the data collection metrics and design compliance rating scales needed to evaluate the usability of the interaction, we relied on the Design Principles and Usability Goals Models. For example, to evaluate the goal of *efficiency*, we would need to record data related to the time it took MADs to successfully interact with the access controls, broken down into segments related to the actions carried out by the dogs as described in our essential use case (see Stage 2 for detail).

Data Collection

For our study, the data collection was as follows (see Table 1 for details):

a) Efficiency Metrics: aimed at measuring the time taken by the canine participant to carry out the task.

b) Safety Metrics: aimed at measuring and scoring behaviors related to observed errors made by the dogs during the activation of the access controls. To score some of the observed behaviors, we implemented a rating scale commonly used by Dogs for Good to assess the dogs' performance during training. The scale applies a rating range between 1 and 5, where 1 indicates that the MAD made no attempt to approach the interface and 5 indicates that the MAD carried out the task with minimal support from the trainer.

c) Behavioral Annotations: aimed at observing and recording MADs' behaviors which communicated the dog's state during the task trials. These were based on the observations and value judgements of the evaluator, supported by their existing knowledge of canine communicative behavior.

Efficiency Metrics		
Metric	**Unit**	**Behavior Recorded**
Prompt Time	Seconds	Time of issue of verbal and non-verbal commands or prompts by the trainer during each trial.
Nudge time	Seconds	Time record of when the dog's snout or paw came into contact with the control during each trial.
Time to open	Seconds	Total time between the trainer issuing the command and or prompt and the door beginning to open calculated by subtracting Nudge time from Prompt time)

Safety Metrics		
Metric	**Unit**	**Behavior Recorded**
Contact attempts	Count	Number of attempts made during a single trial in which the dog physically came into contact with the control.
No-contact attempts	Count	Number of attempts made during a single trial in which the dog moved towards, but did not come into contact with the control.
Unprompted attempts	Count	Number of attempts to activate control after receiving a reward.
Attempts per control	Count	Total count of attempts (contact & no-contact) divided by number of trials (6) multiplied by the number of canine participants (9).
Access control target area	Label	Specific area of the access control with which the dog came into contact with (top, center, bottom, side, top edge, bottom edge, side edge, or other) .
Activating body part	Label	Body part used by the dog to push the control, or the surrounding wall in case of failed contact (top jaw, bottom jaw, full jaw, side jaw, snout, snout tip, right paw, left paw).
Scoring of attempts per task	Score (1-5)	Assessment per attempt of the amount of handler support needed for dogs to carry out the task -scored using Charity scoring system
Scoring of trial	Score (1-5)	Assessment per trial of the amount of handler support needed for dogs to carry out the task -scored using Charity scoring system
Scoring of control use	Score (1-5)	Average assessment of all trials of the amount of handler support needed for dogs to carry out the task -scored using Charity scoring system
Succesful trial	Yes/No	Record of access control having been successfully activated after a single attempt with an overall trial performance score of 5.

Behavioral Annotations		
Metric	**Unit**	**Behavior Recorded**
Distraction	Label	Record of exhibited behaviours which denote distraction during each trial indulging sniffing, turning away from the access control or trainer, focusing on other dogs or researcher.
Concentration	Count	Count of the number of trials in which the dogs exhibited behaviours that denote distraction divided by the total number of trials.
Focus of attention	Label	Record of the dog's general area of focus during the trial (floor, forward, access control, trainer's face, trainer's body, training room, and other)
Overall state	Label	General assessment of the dogs' state throughout the trial (focused, focused on the trainer, unfocused, confident, relaxed, excited, hesitant, confused).

Table 1. Detail of data collection metrics

7 - Improving the canine UX

The application of the MEAU supports the systematic evaluation of usability for an animals' (i.e. MADs') interaction with a technological interface (i.e. access control) which results in the evaluator being able to identify where the interface underperformed relatively straightforwardly. This, in turn, makes it easier to identify the emerging canine user requirements that should inform the next iteration of the interface design in order to provide improved usability for the user. For example, in regard to the *efficiency* metric of control target area (See Table 1 for detail), the data shows a significant percentage of trials in which the canine participants erroneously came into contact with a surface other than the control's push plate while interacting with the prototype access controls. This could potentially indicate that, while allowing the MADs to easily perceive the control as a whole, the current mono-chromatic design of the prototypes was not allowing MADs to quickly and correctly discriminate between the parts of the control which were operational (the push plate) and the parts that were structural (sides). Thus, for the next design iteration of the canine-friendly access controls, an emergent user requirement would be: *the access control should be designed in a way that helps canine users better discriminate between its operational and non-operational parts*. Knowing this, designers can then explore new design solutions to meet the requirement*;* for example, a round shape - as opposed to a square one - could help users target the push plate by offering a larger area relative to the perimeter; or contrasting colors could be used to better distinguish the operational and non-operational parts of the control.

THE STUDY

Research Context and Participants

As a part of a project that aims to improve mobility assistance dogs' working environments through the design of canine-centered technological interfaces, we wanted to investigate how the usability of canine-friendly prototype controls similar to those developed by Mancini *et al.* [11] (Figure 7) compared with the usability of a standard issue access control; whether the prototype controls posed any usability challenges for the dogs; and, if so, how their design would need to be improved in order to improve usability and the resulting MAD UX. The comparative usability study was carried out at Dogs for Good's training facilities. Here, dogs undergo training during which they learn all the behaviors they are expected to perform in order to become Mobility Assistance Dogs. One such behavior is assisting their human partners with opening (motorized) doors. To this end, trainers teach MADs to respond to a "nudge" command which indicates that they must: 1) locate the access control; 2) activate the control with their snout; 3) grasp that they have activated the control successfully; 4) wait for the door to open; and 5) once the door has fully opened, follow their human partner through it.

Nine MADs, nearing the end of or having finalized their training, and their trainers took part in the study. Six of them were crosses between Golden and Labrador retrievers, two were Labrador retrievers and one was a Labradoodle. Six males and three females took part in the study. The dogs' age ranged between 1 year and 2 months, and 2 years and 3 months. To control for environmental variables and to exclude any environmental sensitivities that might have affected the participants, the study took place inside the charity's main training room. One of the room's access points was outfitted with a motorized door opened using a standard issue access control. All dogs at the charity used this control to practice the opening of the door using the "nudge" command.

Figure 7. Access control prototypes

We tested a total of three access controls. The first was a standard issue access control which consisted of a smooth black plastic backing with a flush spring-loaded metal push plate. It had an approximate size of 132 x 127 x 114mm and a recommended installation height between 750mm and 1200mm from the floor [33]. The other two controls, one blue and one yellow, consisted of slightly textured plastic backings with slightly protruding spring-loaded push plates of the same color and material. The controls' height was adjustable through the use of fasteners that could easily adhere to the training room's walls. They had a size similar to that of standard issue access controls, and the same wireless pairing feature that allowed them to open the motorized door. The most notable differences between the existing standard access control and the prototypes were: 1) the material (metal vs plastic); 2) the color (metallic grey vs bright blue and yellow; 3) the finish of the push plate's surface (smooth vs lightly textured); 4) the amount of pressure needed to operate the control (firm vs soft); 5) the amount of protrusion of the push plate (flush vs an approximate protrusion of 5mm); 6) the clicking noise produced by the control when pushed (just audible vs loud); and 7) the installation of the controls (fixed vs adjustable).

Experimental Set-up and Procedure

The study took place over the course of one day, lasting a total of 8 hours; 6 of those hours were dedicated to access control task trials, 1.5 hours was used for trainer interviews and 0.5 hours for lunch. In order to prevent task fatigue on the part of dogs and trainers, the task trials were divided into three sessions, one for each access control tested, with an approximate duration of 2 hours each. Each session consisted of a total of 54 "nudge" task trials, divided into nine separate sequences of individual dog and trainer teams. Each trainer handled three separate dogs, trainer 1 (T1) handled canine participants 1, 2 and 3 (D1, D2, D3); trainer 2 (T2) handled canine participants 4, 5 and 6 (D4, D5, D6); and trainer 3 (T3) handled canine participants 7, 8 and 9 (D7, D8, D9). A sequence involved each trainer and canine participant team carrying out the task trial of opening the door a total of six times. Each task trial started with the trainer positioning the dog next to the wheelchair at an approximate distance of 3 meters from the control. Then, once the trainer had the dog's attention, they would slowly move towards the access control. At an approximate distance of 500-750mm from the control, the trainer would issue the "nudge" command. If the dog was not successful after the first issue of the command, the trainer would offer other verbal (e.g. "good lad") and non-verbal (pointing, looking at the control) cues. If the dog had made multiple unsuccessful attempts to carry out the command, the trainer would reposition the wheelchair to better direct the dog towards the control. Then, once the dog had successfully activated the control using their snout, the trainer would mark the behavior with a verbal cue (e.g. "yes") and offer the dog a reward in the form of kibble. As the door opened the trainers would slowly move away from the door and reposition themselves and the dogs for the next trial waiting for the door to come to a complete close. A total of 9 sequences per access control were carried out in the following order: T1 and D1, T2 and D3, T3 and D6; followed by T1 and D2, T2 and D4, and T3 and D8; ending with T1 and D3, T2 and D6, T3 and D9. In total 162 task trials were recorded, with each dog performing the task of opening the door 18 times, corresponding to 6 trials per control tested.

Mirroring a standard ergonomic testing set-up [34], a 100mm grid overlay on the area adjacent to the control was glued to the wall and floor using thin 4mm black tape, to better understand the position of the dog's body in space during the trials (See Figure 8). All controls were installed at a height of 750mm from the floor to the control's bottom edge.

Figure 8. Study set-up featuring T3 and D6 using the standard issue access control

Ethical Considerations

The study was carried out under the ethical approval of The Open University's Animal Welfare Ethical Review Body. In addition, the study was conducted in accordance with current ACI ethical frameworks [35], which require researchers to protect animal participants' welfare and autonomy at all times when working with them. Mediated consent [35], (art. 2, par. 1) for the dogs' participation was previously obtained from Dogs for Good and the dogs' trainers.

Data Collection

All trials were video, and audio recorded using a GoPro Hero 4 camera mounted atop an overhead boom pole (commonly used for scene lighting), a GoPro Hero 1 camera mounted on a small tripod, and an iPhone 7 mounted on a small tripod. The 3-camera set-up allowed the capture of an overhead, posterior and side view of each trial. This was done to provide a multi-view recording of the dog's complete interaction with the control (approach, operation and reward). The degree of detail of the recording was needed to accurately capture the resulting data collection metrics previously identified by the MEAU (Table 1). All videos were analyzed using ObjectusStudio software, which enables frame-by-frame analysis and millisecond timing.

Data Analysis

To analyse collected data, we used the metrics described in Table 1.

FINDINGS

Table 2 below summarizes the data collected during the study as a result of the interaction of the dogs with each access control tested. The data shown is aggregated for all participants per access control and pertains to the applicable metrics, illustrating the application of the MEAU.

Metric	Standard Issue Access Control (SI)	Blue Prototype Access Control (BP)	Yellow Prototype Access Control (YP)
Time to Open	2.58 seconds	1.69 seconds	2.09 seconds
Contact attempts	1.63 all trials	1.26 all trials	1.17 all trials
Attempts per control	2.17 all trials	1.60 all trials	1.62 all trials
Overall score	4.31/5.00	4.33/5.00	4.2/5.00
Successful Trial	52.22%	76.46%	65.56%
Concentration	12/54	8/54	7/53
Focus of attention	Trainer	Trainer	Trainer
Overall state	Focused (4 MADs) Relaxed (3 MADs) Enthusiastic (1 MAD) Distracted (1 MAD)	Focused (8 MADs) Relaxed (1 MAD)	Focused (5 MADs) Relaxed (4 MADs)

Table 2. Overall participant results per access control tested

The blue and yellow controls had been prototyped to comply with relevant interaction design principles (*perceptibility*, *feedback*, and *affordance*) taking into account dogs' physical, sensory and cognitive characteristics, as per the process described above, in order to achieve good canine usability. Therefore, we expected them to score better on usability compared to the standard issue control. Albeit not to a major extent, this hypothesis was confirmed by the empirical data, in relation to the usability goals that we tested for (*efficiency* and *safety*).

With regards to *efficiency* (i.e. how quickly the dogs could complete the task using the controls), the difference between the controls seems to be confirmed. The dogs took an average time of 2.58 seconds to activate the standard issue access control (SI), 1.69 seconds to activate the blue prototype access control (BP), and 2.09 seconds to activate the yellow prototype access control (YP). In other words, it took the dogs in average 26.74% less time to open the door using the prototype controls, with the BP performing best.

With regards to *safety* (i.e. how the controls facilitated error-free interaction) the number of attempts the dogs made before successfully activating each control is of relevance. For the SI, participants made an average of 2.17 attempts before they succeeded in activating it, for the BP, they made an average of 1.60 attempts, and for the YP, they made 1.62. Of those attempts the ones that resulted in contact were respectively, 1.63 (75.11%) for the SI, 1.26 (78.75%) for the BP, and 1.17 (72.22%) for the YP.

The difference in performance between the BP and the YP could be due to a lower adherence of the YP to the interaction design principle of *perceivability*. Although the YP is made in a bright yellow color that should be highly visible to dogs, the contrast between the prototype and the color of the white wall on which it was installed was significantly lower compared to the contrast of the BP against the same wall. This lower compliance of the YP with the principle of *perceivability* might have affected the control's usability with regards to *safety* and *efficiency*. This finding was also reflected in the work of Mancini *et al*. [11].

Furthermore, we measured an overall trial score based on the sum of the results obtained from the application of the performance rating scale commonly used by Dogs for Good divided by the number of attempts per trial. The BP scored highest of all (4.33); however, although the SI scored higher than the YP (4.31 vs 4.2) both prototypes outperformed the SI (BP: 76.46%, YP: 65.56%, SI: 52.22%) in terms of number of successful trials (i.e. less attempts per trial). This means that even though, compared to YP, SI resulted in more successful attempts when operating it during trials, it required a higher number of attempts per trial. This could indicate that, even though the dogs might have needed a higher degree of trainer support while completing the task, the prototypes provided a higher degree of *efficiency;* or it could also be attributed to the fact that MADs were more familiar with using the SI than the prototypes. The results of the *safety metrics* also show that MADs were able to activate the prototypes efficiently and without error more often than when using the SI, with the BP outperforming the SI by 24.4% and the YP outperforming the SI by 13.34%.

In terms of behavioral metrics, while activating the SI, MADs exhibited distraction related behaviors during 12 out of the 54 trials recorded, when activating the BP during 8 out of the 54 trials, and during 7 out of the 53 trials for the YP. In other words, dogs were distracted 22.22% of the time when activating the SI, as opposed to 14.81% of the time when using the BP and 13.2% of the time when using the YP. These results could potentially indicate that the prototypes slightly outperformed the SI in terms of keeping the dog's attention in spite of possible distractors in the environment.

In regard to their overall state, when operating the SI the majority of MADs were either focused (D1, D2, D7, D8) or relaxed (D4, D5, D9), with one dog being enthusiastic (D3) and another one (D6) being distracted throughout. When interacting with the BP only D5 was relaxed, while all other dogs were focused. When interacting with the YP five dogs were focused (D1, D3, D7, D8, D9) and four where relaxed (D2, D4, D5, D6).

The metrics of focus and overall state were used as a means to qualitatively ground the findings where the results between participants showed more significant differences compared to the quantitative *efficiency* and *safety* related metrics. For example, when comparing the individual metrics for D3 and D8 (Table 3) in regard to their interaction with the SI, D8 appears to have outperformed D3 on all counts. However, D3's overall state was enthusiastic during most of the trials. Excited or enthusiastic dogs can sometimes find it hard to focus on the task at hand, which can lead to distraction [36]. Indeed, during the trial D3 was distracted 83.33% of the time, as opposed to D8 who was not distracted at all. Furthermore, in their enthusiasm, dogs might be quick to act-out the command issued, finding themselves repeating the expected action quickly yet inaccurately [36], as is evidenced by the fact that D3 had no successful trials and a high average of attempts per control (5.16).

Metric / Participant	D3	D8
Time to Open	8.40 seconds	1.85 seconds
Contact attempts	2.5 all trials	2 all trials
Attempts per control	5.16 all trials	2 all trials
Overall performance	3.72/5.00	4.6/5.00
Successful Trial	0%	33.33%
Concentration	5/6	0/6
Focus of attention	Trainer	Trainer
Overall state	Enthusiastic	Focused

Table 3. Results for D3 and D8's operation of the SI

DISCUSSION

As the findings above exemplify, through the application of the MEAU, we conducted a detailed evaluation of the interaction between a group of MADs and different access controls. Although complex and time consuming to apply, we suggest that the MEAU can guide designers though the systematic and comprehensive analysis of an interaction, and that such a method can support the accurate assessment of an interface's usability. This involves examining the structure of a task, identifying requirements for the implementation of the relevant interaction design principles and identifying questions for the assessment of relevant usability goals, based on the characteristics of both the system in question and the intended user. Such an analysis can help identify appropriate metrics against which to evaluate the extent to which the interaction meets appropriate usability goals.

Dealing with the interspecies gap

One way of helping ACI designers deal with the challenges they face due to interspecies differences and communication barriers is to develop methods that help them systematize their research approach and quantify relevant aspects of an interaction through detailed analysis and through the progressive identification of relevant design requirements and questions to be empirically tested. Of course, accurately measuring usability does not mean understanding the experience of other animals, which as human evaluators we may never fully comprehend. However, it does mean that we can at least control for some of the many important factors involved, so that we know what it is that we are measuring, and why. In this regard, a range of interaction design methods can be adapted and applied.

Support throughout the Design Process

The MEAU delineates a rigorous process to help designers understand the degree to which an interface's design meets canine user requirements and affords good usability. The method can be applied at different stages of the design process. Firstly, it can be applied when the interface being designed is at an early stage of development, when its features are still undefined. At this phase, stages 1-4 can be applied and the Design Principles Model can be used to identify requirements for compliance with interaction design principles that are relevant to the interaction in question, before proceeding any further down the design path. If, in its early phases of development, the interface does not adhere to the relevant interaction design principles, it will likely not only fail to meet usability goals and score poorly when empirically tested but will possibly also be unusable by the intended user.

Where an interface is already developed, and its features are for the most part defined, the method can help designers to heuristically assess the extent to which it complies with interaction design principles and, through stages 5-6, it can enable them to identify and empirically test relevant usability goals. Additionally, as previously described in stage 7, it can help designers to consider the relevance of the interaction design principles and goals they should focus on for future design iterations in light of empirical data.

Application to other interactions

Although this paper illustrated the application of the MEAU as it relates to the specific interaction between MADs and three access controls, its structure can be easily applied to support the heuristic and empirical investigation of any other interactions between animals and technology in order to achieve good usability. Following the stages described herein, ACI researchers can apply the method to inform the interaction between other species and products, interpreting the design principles and usability goals based on the specific physical, cognitive and sensory characteristics of the animals as well as the qualities of the products the animals are expected to interact with (e.g. cattle milking systems, captive animal enrichment, veterinary products and services).

Individualized measures of usability

In our description of the proposed method, the emphasis has been on species-specific characteristics rather than on individual particularities. However, Ruge *et al.* [37] highlighted the importance of accounting for variation in the users' individual traits and their implications for the meaning of the behavior they exhibit during an interaction. So, individual differences should be part of the equation when applying the method. Indeed, our analysis of the dogs' behavior while interacting with the access controls revealed individual differences, including significant variations in the findings related to the goals of *efficiency* and *safety;* if the dogs' overall state and focus had not been recorded as part of the behavioral annotations, the findings could have led us to reach different conclusions about the controls' usability. Thus, while overall compliance with interaction design principles and usability goals may be assessed at the species level, individual user characteristics can significantly influence an interface's usability in specific cases. Therefore, individual characteristics (e.g. overall state of the participant when interacting with the interface, particular personality traits, any sensory, physical or cognitive limitations) should be investigated and taken into account when applying the method.

CONCLUSION

In designing user centered interactions between animals and technology, the field of ACI would benefit from the systemic, detailed and holistic evaluation of usability and its impact on the animal UX. By carrying out a comparative usability study of a MADs' interaction with access controls we have attempted to demonstrate the importance of applying interaction design principles and usability goals in conformity with canine user characteristics to inform the design of interfaces that meet the requirements of canine users. Through the application of the MEAU, we have illustrated and followed a rigorous process that can support the detailed measuring of usability in the face of animal users' individual characteristics and any existing human evaluator biases, which may influence the measure of an animals' behavior and the resulting evaluation of usability. The MEAU reframes available usability design and evaluation methods to support the systematic formulation of canine-centric user requirements for future design iterations.

ACKNOWLEDGMENTS

We thank Dogs for Good and their staff who have welcomed us into their facilities and participated in this research, providing invaluable support and insight. We are especially grateful to the Charity's Client Relations Manager Duncan Edwards. We would also like to thank Dr. Rachael Luck and Research Fellow Elizabeth Cox for their continued support during this research. This research is funded by the Pet Plan Charitable Trust.

REFERENCES

1. Mancini, C., *Animal Computer Interaction: A Manifesto.* Interactions, 2011. July, August: p. 69-73.

2. Mancini, C., S. Lawson, and O. Juhlin, *Animal-Computer Interaction: The emergence of a discipline.* International Journal of Human-Computer Studies, 2017. 98: p. 129-134.

3.=Majikes, J., et al., *Balancing noise sensitivity, response latency, and posture accuracy for a computer-assisted canine posture training system.* International Journal of Human-Computer Studies, 2017. 98: p. 179-195.

4. Baskin, S. and A. Zamansky, *The Player is Chewing the Tablet!*, in *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*. 2015. p. 463-468.

5. Westerlaken, M. and S. Gualeni, *Grounded Zoomorphism*, in *Proceedings of the 2014 Workshops on Advances in Computer Entertainment Conference - ACE '14 Workshops*. 2014. p. 1-6.

6. Pons, P.J., Javier; Catala, Alejandro, *Towards Future Interactive Intelligent Systems for Animals: Study and Recognition of Embodied Interactions*, in *IUI 2017 Intelligent Training and Educational Interfaces*. 2017, ACM: Limassol, Cyprus.

7. Zeagler, C., et al. *Going to the dogs: towards an interactive touchscreen interface for working dogs*. in *ACM UIST 2014 on User Interface Software and Technology*. 2014. Honolulu, HI: ACM.

8. Robinson, C., et al., *Designing an emergency communication system for human and assistance dog partnerships*, in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15*. 2015. p. 337-347.

9. Jackson, M.M., et al., *FIDO—Facilitating interactions for dogs with occupations: wearable communication interfaces for working dogs.* Personal and Ubiquitous Computing, 2014. 19(1): p. 155-173.

10. Resner, B., *Rover@Home: Computer Mediated Remote Interaction for Dogs*, in *Media Arts and Sciences* 2001, Massachusetts Institute of Technology: Cambridge, MA, USA.

11. Mancini, C., et al., *Towards Multispecies Interaction Environments: Extending Accessibility to Canine Users*, in *Proceedings of the Third International Conference on Animal-Computer Interaction - ACI '16*. 2016. p. 1-10.

12. Norman, D., *Affordances, Conventions and Design.* ACM Interactions Magazine 1999(May-Jun): p. 30-42.

13. Norman, D.A., *The design of everyday things.* 1998: London : MIT Press

14. Preece, J., H. Sharp, and Y. Rogers, *Interaction design: beyond Human-Computer Interaction* 4th ed. 2015: John Wiley & Sons.

15. Mancini, C., Petre, Marian., van der Lindien, Janet, *TM356 Interaction Design and the User Experience*. Introduction to interaction design. Vol. Block I. 2016, Milton Keyens, UK: The Open University.

16. Kirksey, S.E. and S. Helmreich, *The Emergence of Multispecies Ethnography.* Cultural Anthropology, 2010. 25(4): p. 545-576.

17. Muto, T. *Service Dog Definition*. 2014; Available from: http://www.canineprofessionals.com/service-dog-definition.

18. *Dogs for Good*. Available from: https://www.dogsforgood.org/.

19. Mancini, C., et al., *Re-Centering Multispecies Practices*, in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. 2015. p. 2673-2682.

20. Westerlaken, M. and S. Gualeni. *Digitally Complemented Zoomorphism: a Theoretical Foundation for Human-Animal Interaction Design*. in *DPPI: Praxis and Poetics*. 2013. Newcastle upon Tyne, UK.

21. Byrne, C., et al., *A method to evaluate haptic interfaces for working dogs.* International Journal of Human-Computer Studies, 2017. 98: p. 196-207.

22. Brown, T., *Change by Design: How Design Thinking Transforms Organizations and Inspires Innovation*. 2009: HarperBusiness. 272.

23. Constantine, L.L., *Essential Modeling: Use Cases for User Interfaces.* ACM Interactions, 1995. 2 (2): p. 34-46.

24. Constantine, L.L.L., L. A. D. , *Software for Use: A Practical Guide to the Models and Methods of Usage-Centered Design*. 1999, Boston, USA: Addison-Wesley.

25. Hekkert, P.S., Hendrik N.J., *Introducing product experience: Product Experience*. 2008, Oxford: Elsevier Ltd.

26. Ford, M., *Vision in dogs and Cats.* CVMA/SBCV 2014.

27. Miller, P.E. and C.J. Murphy, *Vision in Dogs.* JAVMA, 1995. 207(12): p. 1623-1634.

28. Wever, E.G., *Sound Reception*, in *Encyclopædia Britannica*. 2017, Encyclopædia Britannica, inc.

29. Heffner, H.E., *Hearing in Large and Small Dogs: Absolute Thresholds and Size of the Tympanic Membrane.* Behavioral Neuroscience, 1983. 97(2): p. 310-318.

30. Thimbleby, H., *User Interface Design*. 1990, Harlow, Essex, UK: Addison-Wesley.

31. Ruge, L., C. Mancini, and R. Luck. *Requirements Engineering Elicitation for Mobility Assistance Dogs:*

Meeting Canine User Needs Through Technology Enabled Interpretation. in *24th International Conference on Requirements Engineering: Foundation for Software Quality*. 2018. Utretch, The Netherlands.

32. Zamansky, A., D.v.d. Linden, and S. Baskin, *Pushing Boundaries of RE: Requirement Elicitation for Non-human Users*, in *2017 IEEE 25th International Requirements Engineering Conference (RE)*. 2017. p. 406-411.

33. *Access Control Systems and Disability* 2015.

34. Van Cott, H.P.K., Joel J.; Pezoldt, Valmon J.; Porter, Louis G.; Fried, Charles; Fechter, John V.; Persensky, J. J.; Teichner, Warren H, *A standard ergonomics reference data system: the concept and its assessment*, in *NBSIR 77-1403*, U.D.o. Comerce, Editor. 1978, National Bureau of Standards: Las Cruces, New Mexico, USA.

35. Mancini, C., *Towards an Animal Centered Ethics for Animal-Computer Interaction.* International Journal Human-Computer Studies, 2017. 98: p. 221-233.

36. Carmichael, L., *Dogs for Good*, L. Ruge, Editor. 2017.

37. Ruge, L., et al., *User centered design approaches to measuring canine behavior*, in *Proceedings of the Fifth International Conference on Animal-Computer Interaction - ACI '18*. 2018. p. 1-12.