Studying Human-Robot Collaboration in an Artistic Creative Process

Extended Abstract

J. Kildal
IK4-TEKNIKER
Iñaki Goenaga 5, Eibar 20600
Spain
johan.kildal@tekniker.es

K. Tahiroğlu
Department of Media, Aalto University
School of ARTS
Espoo, Finland
koray.tahiroglu@aalto.fi

J.C. Vásquez
Department of Media, Aalto University
School of ARTS
Espoo, Finland
juan.vasquezgomez@aalto.fi

I. Maurtua
IK4-TEKNIKER
Iñaki Goenaga 5, Eibar 20600
Spain
inaki.maurtua@tekniker.es

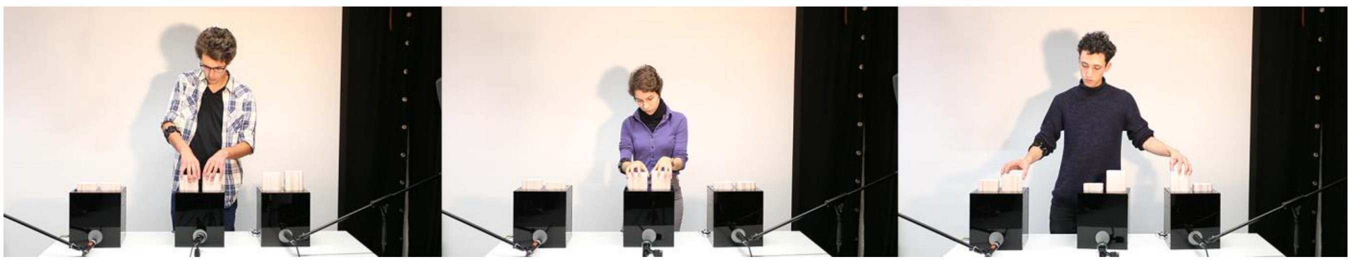

Figure 1: Musicians and network of robotic intelligent sonic agents creating music collaboratively

ABSTRACT

The design and evaluation of a human-robot collaborative system requires following user centered methodologies and processes for its success. A cornerstone of this approach is to collect requirements from three main sources: technological factors, human factors and application-domain factors. In this paper we discuss the relevance that the application domain has in the relative proportion of quantitative and qualitative metrics for an adequate full evaluation. We discuss the particular application domain of collaborative artistic creation, with particular detail on a use case of collaborative music creation through improvisation. We highlight the higher proportion of qualitative metrics that is needed to evaluate the collaboration with a robotic system in such highly-subjective creative scenarios.

CCS CONCEPTS

• **Computing methodologies → Embeded and Cyber-Physical Systems**; **Human-Centered Computing → Human-Computer Interaction (HCI) → HCI design and evaluation methods**; **Applied Computing Robotics → Arts and Humanities → Performing Arts**

KEYWORDS

Human-robot collaboration, user studies, interaction, arts, creative process, autonomous intelligent agents, new interfaces for musical expression

ACM Reference format:

J. Kildal, K. Tahiroğlu, J.C. Vásquez. 2017. Human-Robot Collaboration in an Artistic Creative Process. In *Proceedings of ReHRI'17 – International Workshop on reproducible HRI experiments: scientific endeavors, benchmarking and standardization. 12th ACM Annual Conference on Human-Robot Interaction. Vienna, Austria, March 2017 (HRI'17),* x pages. DOI: xx.xxxx/xxx x

1 INTRODUCTION:

It is clear that collaborative scenarios between a robot and a person have to be designed, researched and developed following user-centered methodologies, with the aim to best realize envisioned scenarios and meet requirements [1, 2]. The primary subject of design in such interactive scenarios is the technological component (a robotic system or an autonomous intelligent agent) as well as the mechanisms through which interactions between robot and human take place. As the product of human design and, they are matter for iterative re-design and re-implementation, in the process towards meeting the goals envisioned for the collaboration. The restrictions and requirements that technology imposes on the design process relate to the state of the art in the technology that is available, its affordability and versatility, as well as the feasibility and complexity of implementing specific interaction techniques.

In sharp contrast, the non-technological partner in the interaction (i.e., the human actor) cannot be redesigned or otherwise altered easily in any substantial way. At most, humans engaged in collaborating with robotic systems can be required to acquire new skills or accept new paradigms, but they cannot be subjected to a design process. Instead, human factors that are relevant for the interaction with a robot need to be studied and understood, and these factors will be necessary additional input for the requirements that the interactive collaboration with the robotic system has to fulfill.

There is still a third main source of requirements for the design of collaborative human-robot system: the application domain in which the collaboration is framed, which encompasses additional requirements from stakeholders and from the environment. The application domain can modulate significantly the set of requirements that were introduced by the analyses of both technological and human factors. Very importantly, application domain factors also determine which metrics will be most relevant from the perspective of the experimental research of the interactive system.

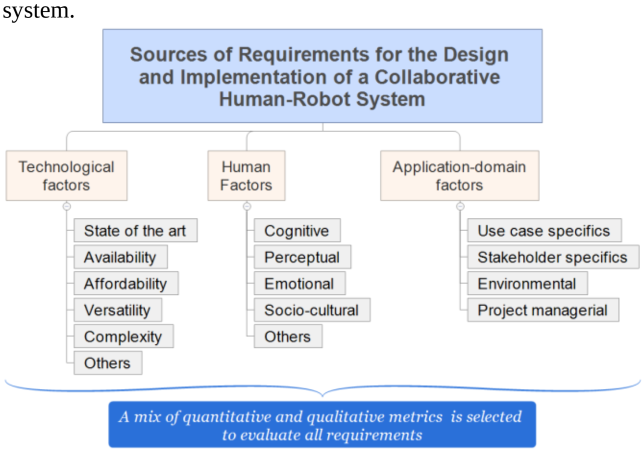

Figure 2: Sources of requirements for the design and implementation of a typical human-robot collaborative system.

The next sections in this paper will focus on discuss this last aspect: the selection of experimental research metrics based on the application domain. We will illustrate our points with a case study from research that we have conducted in a very specific application domain: collaborative musical creation between an expert musician and a robotic autonomous and intelligent instrument.

2 APPLICATION-DEPENDENT METRICS

When designing and developing a human-robot collaborative solution through user-centered methods, a set of various relevant metrics is generally needed in order to be able to evaluate all the important aspects about the system (well beyond the regular metrics used for the evaluation of robot developments alone [3]), with respect to the requirements that have been defined for it. The application domain has a deciding influence on the selection of such set of metrics.

As a general rule, it is a good practice to collect a mix of quantitative and qualitative data through experimental user studies at every stage of the development. In that way, it is easier to form a complete picture of the degree to which a solution adjusts to requirements from different origin (technology factors, human factors and application area). For specific aspects of the system, and in our experience, it is also a good strategy that the various selected metrics show a degree of redundancy in the data that is collected with them. For instance, if *time-to-complete-task* is considered to be important in a particular scenario, this can be measured both objectively (a quantitative measure) and subjectively (a qualitative measure). Quantitative data would usually be the absolute time actually elapsed till completion of the task, which would help evaluate the compliance with the efficiency requirements established for a particular application domain. Complementing this, a qualitative measure of the time to complete task could also be considered. For instance, the subjective perception that the person in the interaction loop has of the time elapsed until completion of the task. This second measure can help evaluate the user experience (UX) that the collaboration is capable of offering to the human collaborator. Considering both measures jointly, researchers could provide a nuanced and multi-perspective answer to the question, *is time to complete task fast enough in this collaboration*?

2.1 A Productive Application Domain

As already stated, forming a meaningful mix of quantitative and qualitative metrics depends heavily on the application domain. In most collaborative scenarios, robots are intended to perform sub-tasks with a degree of autonomy, while also maintaining coordination with the human they are collaborating with for the successful joint completion of an overall task. In such cases (for instance, in the collaborative assembly of flat-pack furniture), both the robot and the person can share the same detailed description of the work to be carried out. As a result, quantitative metrics tend to have a predominance in studies that evaluate the robot´s performance and the collaboration as a whole. Metrics such as correctness, productivity, efficiency, effectiveness, cost-efficiency, waist (of time and material), resulting quality and similar others are commonplace. Beyond the objective and pragmatic measures just listed, qualitative measures are also necessary to account for factors that derive from the direct collaboration between person and robot, and the UX that can result from it. Important metrics that account for an overall UX can include perceived safety, mental workload, quality of rapport with their technological collaborator, and overall satisfaction with the resulting work done or service provided, to name some.

2.2 An Artistic Creative Application Domain

To illustrate the dependency of evaluation metrics on the application domain, it is useful to compare the productivity-oriented application domain example just outlined with another one in which the intended result of the collaboration cannot be known in detail by the robot (or even by the human actor in most cases, for that matter) until its completion: a scenario of collaborative artistic creation. The subjective nature of the artist's vision and creative agenda is a main defining feature of a collaborative artistic creation process. In such a context, the robotic collaborator may be aware of a general frame of collaboration previously agreed with the artist (e.g., the creative material to be used and, to some degree, the structure of the resulting work), but a detailed description of the steps to be taken and of the exact final result intended cannot exist beforehand, by the very nature of the artistic creative process.

This main difference between productive and artistic creative collaborative scenarios will be reflected in the mix of metrics that should be selected for evaluation. It is still true that there should be a mix of both quantitative and qualitative data sources for a complete analysis. However, there is not an objective "right or wrong" criterion that can be described easily in a mathematical way, and correctness of the final result will depend largely on purely artistic criteria. Having said that, the robot will still have clearly-defined tasks that have to be evaluated through quantification (against an *objective correctness* reference). Therefore, while both quantitative and qualitative data are still necessary for evaluation, it is likely that the mix of metrics shifts towards qualitative in this kind of application domain.

3 CASE STUDY: HUMAN-ROBOT COLLABORATIVE MUSIC CREATION

To illustrate our discussion in the previous section, we offer a brief outline of our NOISA project, as a case study of a collaborative robotic system designed and developed for the application domain of collaborative improvised music performance [4–6]. NOISA (Network of Intelligent Sonic Agents) consists of a set of three networked robotic autonomous agents, each of which is a musical instrument (see Figure 3 for an external and internal view of one of the agents).

3.1 The Rules of the Collaboration

The objective sought by the collaboration between the human performer and its instrument is to amplify the human musician´s creativity and to maximize her capacity to complete a musically-satisfactory improvised composition.

As a musical instrument, each NOISA agent provides an interface in the form of two handles that can be slid up and down by hand, producing synthesized sound that the musician can model and control with a great level of precision (Figure 1 shows stills of three musicians performing on NOISA.)

From the perspective of the musician, the three instruments, each with its signature sound space, provide a rich and expressive and non-trivial instrument that allows the performer to develop mastery over time through practice and experimentation.

As robotic autonomous agents, each agent can actively move its own motorized handles, displacing them physically in space, and producing its own sonic output (self-performing capacity). However, and crucially, the network of autonomous agents does not take the initiative in the performance. Instead, it stands respectfully and discretely in the background, acting as a mere passive musical instrument in the hands of its performer. From that position, it observes the musician perform and it observes itself being played. With the data collected during the performance, it learns from the musician's musical discourse, identifying the main motives used by the artist and the artist´s discourse as it unfolds.

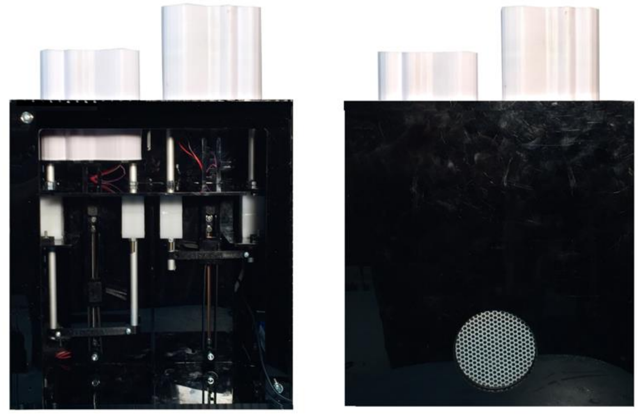

Figure 3: One of the robotic intelligent and autonomous agents in the NOISA networked instrument.

A further key functionality is enabled by sensor data obtained from multiple sensors on the instrument itself (position sensors, touch sensors on the handles), on the musician (EMG sensors on the forearm) and in the environment (external RGB and depth cameras analyzing the musician's movements and face expressions). The system is able to combine all the sensor data with the learning from the musician's performative discourse, and estimate the *level of engagement* that the musician has with the process of improvisation at each moment during the performance of a piece. With this information, when NOISA detects a drop in the level of engagement of the musician with the performance activity, the robotic instrument kicks in as a collaborator. It carefully executes physical (and consequently also musical) actions that are based on the motives and discourse learned from the artist through observation. These interventions are intended as cues that the collaborating system offers the musician to regain engagement with the performance and help maintain her musical discourse. When NOISA observes that the level of engagement has recovered, the system retreats again to continue observing and learning from the behavior of passive instrument.

3.2 Evaluation Metrics

Like with the design and development of any human-robot collaborative system, the evaluation of NOISA required a set of metrics that included both quantitative and qualitative ones. The themes for requirements derived from technological factors, as outlined in the introduction (see also Figure 2) were still relevant for the construction of the hardware system, the musical instrument and the programming of the intelligent autonomous agents. Similarly, regular human factors requirements applied, in particular for the design of the interaction. When compared with a productivity-centric collaborative system, the main differences in the metrics employed arose from the application domain. Some of the classic production and productivity-related metrics still applied, when considering the musical production as *the product*. However, some of the quantitative metrics have less relevance of become fussier to apply. For instance, productivity and efficiency of production were less relevant, as a performance session could be seen like a single unit production batch, where trial and error was not a plausible approach. Instead, quality of the outcome was still important, but the way of assessing it had a marked qualitative weight, as quality ratings would depend neatly subjective artistic criteria. Evaluating system performance also showed interesting challenges, in particular when evaluating the engagement estimation engine. Such engine was developed to estimate an evolving curve of engagement as accurately and faithfully as possible (a quantification of engagement curves that could serve for decision making for the system). However, the assessment of such AI engine relied on the self-assessed levels of engagement provided by musicians over many performances. These master references were, in turn, highly subjective.

4 CONCLUSIONS

Like with any interactive technology, a cornerstone of the user centered processes for the design and development of collaborative human-robot systems is the comprehensive capture of relevant requirements. They are the ultimate checklist for the evaluation of a system, which should satisfy requirements to the largest possible extent. This is observed by selecting an adequate mix of metrics for the data collection in experimental studies.

The mix of metrics usually compiles both quantitative and qualitative data, necessary to respond to the varied nature of requirements that is always found in one such hybrid human and robotic environment.

We distinguish three broad areas as the main sources for the design and implementation requirements: technological factors, human factors and application domain factors. Of these three sources, in this paper we have stated that the application-domain factors have the largest influence on the relative proportion of quantitative and qualitative evaluation metrics in an optimal experimental design. We have illustrated this outlining a typical productive human-robot collaboration scenario and another more unusual one in which robot and person collaborate in developing artistic creative activity. Through a use case based on our NOISA research project, we have defended that, in an artistic creative domain, the proportion of relevant qualitative metrics needed to evaluate as system is necessarily higher.

REFERENCES

1. Holtzblatt, K., Wendell, J.B., Wood, S.: Rapid contextual design: a how-to guide to key techniques for user-centered design. Morgan Kaufmann (2005).
2. Adams, J.A.: Human-robot interaction design: Understanding user needs and requirements. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. pp. 447–451. SAGE Publications (2005).
3. Aaron Steinfeld, Terrence Fong, David Kaber, Michael Lewis, Jean Scholtz, Alan Schultz, Michael Goodrich: Common metrics for human-robot interaction. Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction. , Salt Lake City, Utah, USA (2006).
4. Tahiroğlu, K., Svedström, T., Wikström, V.: NOISA: A Novel Intelligent System Facilitating Smart Interaction. In: CHI EA '15. pp. 279–282. ACM (2015).
5. Tahiroglu, K., Svedström, T., Wikström, V.: Musical Engagement that is Predicated on Intentional Activity of the Performer with NOISA Instruments. In: Berdahl, E. and Allison, J. (eds.) NIME '15. pp. 132–135 (2015).
6. Tahiroglu, K., Vasquez, J.C., Kildal, J.: Non-intrusive Counter-actions: Maintaining Progressively Engaging Interactions for Music Performance. Presented at the NIME '16 (2016).