Experimenting with a New Proposal for Digital Design Instrumentation

A Wizard of Oz method to study its impact on activity

Gaelle Baudoux[1,2], Pierre Leclercq[1]
[1]LUCID-ULiege from University of Liege [2]ndrscr from University of Liege
[1,2]{gbaudoux|pierre.leclercq}@uliege.be

This paper is a continuation of a previous research in which we highlighted the limits of the current transition between ideation and digital production, in particular to generate the Building Information Modeling models, and in wich we proposed an alternative transition through the semantic and digital formalisation of the building based on an automatic interpretation of the architectural sketches. We must now study how to test this proposed transition technology. This paper presents the test set up to determine (1) how to extract meaning from the often ambiguous, incomplete and personal graphical traces for generating the building models and (2) how to return these generated models so that they constitute an added value for the design activity. This arrangement consists of a Wizard of Oz type experiment immersing expert designers in this technology for a design capsule. The protocol includes the elaboration of the instrumented work environment and the collection of data via cameras and interviews. These experiments allow to obtain data documenting (1) the activities required to interpret architectural sketches and produce models, and (2) the design activities and human-machine interactions of the architects.

Keywords: *Design, Human-computer interaction, Ideation-CAD transition, Sketch interpretation, Experimentation protocol, Wizard of Oz.*

INTRODUCTION

Architectural design processes have changed in recent years to meet the pressing demand for economic and ecological performances. Indeed, they have moved from batch-by-batch sequential processes to multi-actor integrated processes. New technologies, such as BIM, have been introduced to meet these new challenges. Unfortunately, while BIM does help in the construction phases, it is not well suited to the early creative phases of design

We have shown in a previous article that the transition from ideation to BIM modelling, i.e. from the activity of idea generation and development to the exhaustive and precise specification of building characteristics, suffers from several limitations in its current implementation (Baudoux *et al.*, 2022). « The convergent loops of ideation and rough design are followed by the first split, aimed at producing the rough BIM model. In this first split, there is firstly an imposed change in the nature of activities from design to production. Secondly, a change of design media is imposed. And thirdly, there is a loss of information, which is only partly offset by the addition of complementary information and exchange media in parallel with BIM models. Once a rough BIM model has been produced, a second split arises as it is hardly usable and has to be manually remodeled to serve the rest of the BIM process. During this BIM stage, several digital models are created and then modified, which therefore coexist. They are also supplemented by several complementary documents of various kinds, which evolve in parallel. » (Baudoux *et al.*, 2022, p. 3 and 7).

In this previous article, we also proposed an alternative to this current transition by using automated semantic and digital modelling from sketches to feed the BIM model. This new method would « automatically collect, throughout the process, useful information from the graphical traces of the designers. This information will be automatically modeled in a semantic model, rich in meaning. Then, based on a specific ontology, a software will interpret these project's attributes to automatically generate a 3D model of the building under design that can be more directly implemented in the BIM digital model. » (Baudoux *et al.*, 2022, p. 8).

EXISTING DEVELOPMENTS OF SIMILAR PROPOSALS

When it comes to developing design support tools based on architectural sketch interpretation, Safin (*et al.,* 2008) differentiates three possible approaches:

- A digital transposition of traditional tools, replacing paper with a tablet, pencils with a digital pen, etc.
- Tools that analyse and interpret sketches to produce a digital model of the building.
- Tools that allow for 3D modelling using gestural pen commands or perspectives.

In this case, we have proposed a tool of the second type.

Originality in comparison to similar software

The 90s and 2000s saw the growth of numerous sketch interpretation softwares. Our current proposal goes beyond ones such as *ASSIST* (Davis, 2002), *SketchIt* (Stahovich, 1996), *UDSI* (Notowigdo, 2004), by offering 3D model generation. It differs from *Sketching Reality* (Chen et al., 2008), which interprets sketches by hand to generate realistic 3D renderings, from *Isom* (Jamagne, 1991), which produces a semantic model from 3D geometry, and from *Topologic* (Aish et al, 2018), which generates topological 3D models from geometric 3D models), by proposing a direct generation of digital 3D models and semantic models from freehand or CAD sketches.

We continue the work of the *NEMo* software (Demaret & Leclercq, 2011), developed after *EsQUIsE* (Leclercq & Juchmes, 2002; Juchmes, Leclercq & Azar, 2004), which can recognise architects' freehand drawings on the basis of the various conventional symbols used spontaneously in architecture, without the obligation to use a specific machine-understood representation, even if they present incompatibilities, in order to generate a 3D model of the building.

Technical feasibility

From a feasibility perspective, we believe that current technologies are sufficiently mature. Already in the 2000s (Stahovich, 1996; Davis, 2002; Notowigigdo, 2004) it was possible to interpret the pixels of lines and combine direction and speed information to generate any abstract shape. The software could recognise text, shapes and predictable symbols, all in 2D, from scanned paper traces or from tablets. More recently, Ding and Liu (2016) show in their literature review that inputs have evolved, accepting 2D drawing, 3D drawing by crossing 2D planes, drawing in immersion in a 3D model or even perspectives. However, it is still necessary to have expertise and to have knowledge of the domain's design rules. In terms of output, 2D representations, 3D models or semantic diagrams can be generated.

ISSUE

Our issue is to determine how to collect and formalise the information from the first design phases in order to generate the expected 3D models and thus feed the design process. Concretely, our study consists in **elaborating a test protocol** for an instrumentation system allowing us to understand how to extract the **meaning of the architects' graphical traces** and in what form to return the generated models in order to constitute an **added value to the design** process, beyond the production of deliverables. This protocol and the data collected will thus allow us to derive results on (i) the mechanisms of sketch interpretation necessary to make this technology efficient, and (ii) its impact on the designer's activity and on the human-computer dialogue.

Our paper focuses on the elaboration of the whole test protocol and will give a preliminary preview of results to demonstrate its feasibility..

GENERAL METHODOLOGY

To address this issue, we are setting up a laboratory experiment protocol using a Wizard of Oz technique. This technique consists in simulating the functionalities of an innovative technology by replacing them with equivalent human work, hidden and in real time. In this way, the observed subject believes that he/she is using the so-called technology except that it does not need to be developed yet. This makes it possible to assess in advance its impact on users and their interaction with the machine (Dahlbäck et al., 1993; Browne, 2019; Rietz et al., 2021) and thus help to figure out the development needs.

This Wizard of Oz method is widely employed in research on Human-Robot interactions (Riek, 2012; Clabaugh & Matarić, 2019) and on automated driving (Frison et al., 2020). This method is also used in the field of clinical psychology with conversational agents (Gaffney, Mansell & Tai, 2019) and in the development of virtual or augmented reality (Cordeiro et al., 2019; Freitas et al., 2020). It is rarely implemented in design-related research, and when it is, it is in NLP-based text editor design (Yang et al., 2019), video game design (Yesilbek & Sezgin, 2021), or ornamental object design based on speech descriptions (Cuadra et al, 2021) but not, to our knowledge, in the context of architectural project design.

Our goal, through the implementation of this experiment in the form of a Wizard of Oz, is to immerse the subjects, expert designers in architecture, in this technology, thus replaced by a team of human modelling agents, and to ask them to carry out a preliminary design session of a fictive architectural project for which we provide them with the expected programme.

EXPERIMENTAL CONTEXT

To conduct our experiments, we have access to the Lucid Usability Lab (fig. 1). This Digital Collaborative Studio is an "augmented" space of 120m2, offering a variety of technological devices for graphic collaboration (graphic tablets, connected desks, interactive desks and wall), as well as a suite of software for digital sketching (modelers), graphic communication (SketSha) and multi-actor design process analysis (Common Tools).

Figure 1
Lucid Usability Lab

The observed subjects are divided into two role groups: the human modelling agents, hidden behind the "intelligent software", and the expert architects, one-day users of this "software".

The human modelling agents are all architect-engineers students, in their third or fourth year out of five. They work in teams to model the designed project in real time. For logistical reasons of participants' availability, two teams of agents were composed, each composed of three drafters and one coordinator.

The design subjects, on the other hand, have varied profiles, shown in table 1. We intentionally diversified the profiles in order to eliminate biases of background, digital sensitivity or gender. The study population consists of 9 subjects, men and women, with different qualifications and an average of 10±8 years of experience. This number enables us to produce analysable results and to diversify the profiles.

Table 1
Characteristics of the observed population of designers

Subject	1	2	3	4	5	6	7	8	9
Date -dd.mm	18.03	21.03	30.03	01.04	15.04	20.04	25.04	27.04	09.05
Gender	Man	Woman	Man	Woman	Woman	Man	Man	Man	Man
Age -years	52	24	25	34	30	30	48	30	28
Background	Eng. Archi	Eng. Archi	Eng. Archi	Archi	Eng. Archi	Eng. Archi	Eng. Archi	Archi	Eng. Archi
Activity	Agency	Agency / Research	Agency	Research	Agency	Agency	Agency	Research	Agency
Experience	Senior	Junior	Junior	Interm.	Interm.	Interm.	Senior	Junior	Junior

DATA COLLECTION

We will start by explaining the general principle of this particular experiment before detailing its implementation in the experimental space and the data collected.

General principle of the experiment

As illustrated below, in figure 2, each experiment consists in asking a subject architect to design a family house in an urban environment. We provide him/her with the architectural programme listing the different spaces required by the client as well as a plan of the site along with photos of it. The subject designer has one and a half hour to make a preliminary design of his project. He is equipped with a graphic table and digital pen, which is the input connected to the "software". The software, embodied by the three modelling agents, sends back to him, based on the dynamically evolving sketches received, (1) a basic 3D digital model of the project, (2) the plans, sections and elevations, if applicable, set out in CAD format, and (3) an inspirational board or realistic rendering of the project. These various elements are sent back to the subject designer every 5 minutes, a regular delay set to allow the agents to sufficiently update the progress of the modelling while getting closer to a real time response. In addition to these regular spontaneous sends, the subject can at any time send a more specific command to the "software" to request a particular point of view in the 3D model, a particular section or inspirational images based on 3 specified keywords. The conversation between design subject and the " computer " continues in this way until the end of the design session.

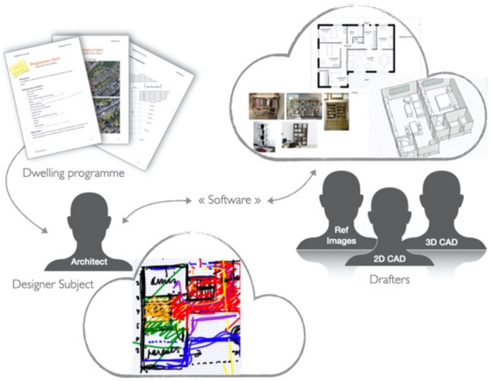

Figure 2
General structure of the experiment and examples of sketches and generated visuals for subject 1

Here is an exemple of this interaction between designer and "software": designer 9 draws the back façade of his building; the modellers recognise the neighbouring houses and the number of floors as well as the shapes of the openings and their distribution; they update the 3D model with this façade composition as well as the plan and propose a new inspirational board; the designer welcomes this update of the model then lingers over the images, one of which catches his attention; he then use it as inspiration to create a large opening across the entire façade's width.

At the end of the design session, the subject is asked to present the project to a fictive client based on the sketches drawn on the graphic table and along with the visuals sent throughout the session. The fictive client is played by a colleague from the laboratory.

Concretely, the data collected by this protocol aims to qualify (1) the reaction behaviours of designers when receiving the automatically generated visuals, (2) the interpretation capacity of the modellers and (3) the added value of this software assistance for the project.

Experimental space

The experiment takes place in two adjacent rooms. In the first room (fig. 3 - "ST1") take place the **subject designer and the researcher**. The design subject sits at a virtual desk consisting of a computer with three screens and an A2-sized graphics tablet embedded in the table. The graphics tablet, via the shared digital annotation software SketSha (Elsen & Leclercq, 2008), is the digital sketching interface serving as the input for the drawings. On the three screens appear, from left to right, (1) the timer of the experiment and the current board of inspirational images or realistic renderings of what the project would look like, (2) the current CAD plans and sections, and (3) the current 3D model as well as the 'software' control terminal.

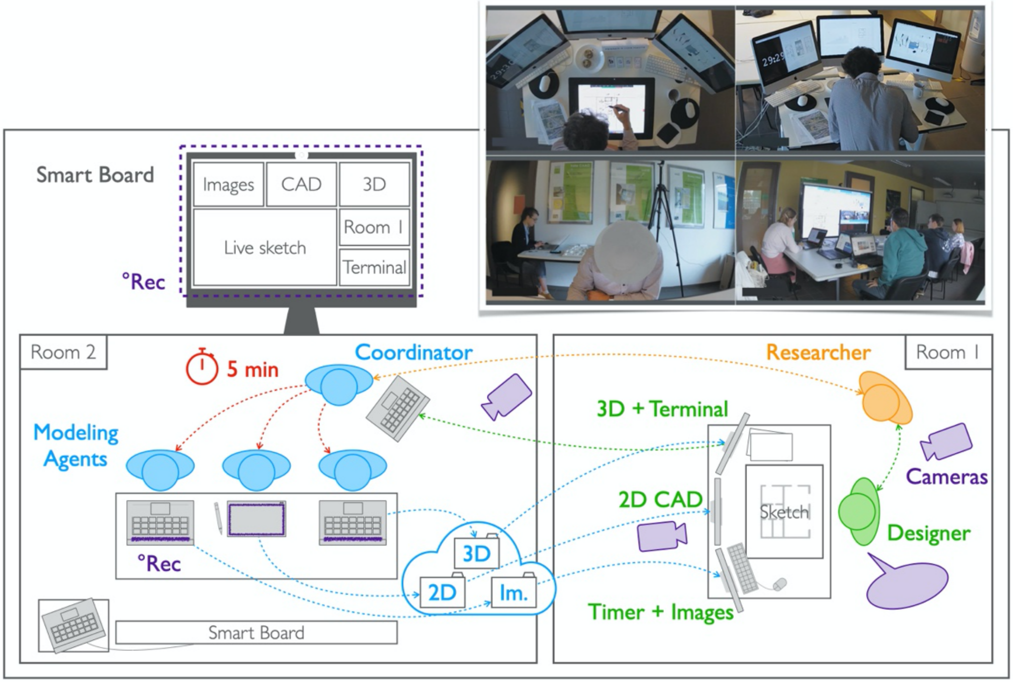

Figure 3
Spatial and technical set-up
of the experiment

The subject designer also has paper documents on his desk showing the architectural programme and site information. The researcher is present to facilitate the use of the software, to regulate the different phases of the experiment and to collect the first observation data, e.g. the phasing of the design activity and the questioning moments to which they will return later. Several cameras are also installed in the room to record the experiment: a camera on the ceiling, framing the uses of the virtual desk, a camera behind the subject's back, framing the representations sent back by the screens and a camera in front of the subject, framing his/her facial expressions and where he/she looks. The subject is also asked to think aloud, according to a think aloud protocol, thus expressing his design rationale and his feelings about the computer.

In the second room (fig. 3 - "ST2"), we have the team of **three modelling agents and their coordinator**. They face a control screen that allows them to monitor the ongoing design in the adjacent room. This control screen displays, from left to right and from top to bottom, (1) the image board, the 2D CAD documents and the 3D model currently being displayed to the subject, (2) the real-time evolution of the sketch, (3) the overall view of the design room, relayed through the cameras, and (4) the software control terminal. The modelling agents work continuously on their respective modelling task, i.e. generating an inspirational board of images suitable for the current focus of the project or a realistic rendering of what the project would look like, producing clean plans and sections, and producing the 3D model based on the sketches received. The coordinator exchanges with the subject designer via the software's control terminal, transmits the subject designer's requests to the modelling agents, triggers the sending of the representations to the subject designer's screens after the elapsed five minutes and, finally, provides a fourth opinion to help interpret the sketches received. He is also in private informal communication with the researcher so that he can report on any technical problems.

To carry out their task and place them in the same conditions as the future technology they are simulating, as well as to ensure that they behave consistently, the modellers have been given specific instructions to follow:

- forbidding them to design an architectural proposal, their role being limited to translating the received representations;
- providing them with a 2D and 3D library of standard furnishings;
- specifying the by-default measurements to be assumed, unless otherwise stipulated by the designer, for wall thicknesses, ceiling heights, window sill heights, roof slopes, etc.
- providing them with the layout plan and the 3D model of the site;
- informing them of the content of the design brief.

It should be noted that the modelling agents in charge of producing the 2D clear documents and the 3D model do not hear the designer's think aloud, so that they are placed in the same input data conditions as the software, i.e. only the sketch and the architectural program provided.

The work of the modelling agents is also recorded by a camera placed in Room 2 and through a recording of their working computer screens.

This is the final protocol after adjustment following the zero experiment. The modelling agents have also been trained on this zero subject and are therefore fully operational from the very first experiment.

Time course

The detailed course of the experiment is the following, for a total time of 2 hours 55 minutes for the subject designer and 3 hours 55 minutes for the modelling agents:

1. Setting up the experimental space and the modelling team (respectively 30' and 15')
2. Explanation of the experiment to the subject designer without revealing the Wizard of Oz principle (10')
3. Think aloud training (10')
4. Presentation of the dwelling programme (10')
5. Design and interaction with the software (90') - breaks can be taken at the subject's convenience
6. Closure of the design session (5')
7. Presentation to client (10')
8. Self-confrontation of the subject designer about the specific questioning moments (20')
9. Directed debriefing interview with the design subject (20')
10. Individual self-confrontation interviews with the modelling agents (3 x 15')

Collected Data

First of all, the data documenting the **experiment's course** are collected by recording the control screen and through the cameras in Room 1. We thus obtain all the data characterising the stages of the design, the different visuals sent, their time code and any software commands given by the subject designer. We also gather data on the use of the various visuals for project communication purposes, in this case towards the fictive client.

The data documenting the **subject's activities and reasoning** are collected through the cameras in Room 1, especially through the camera framing the designer's face, through the think aloud protocol during the design process, and through the self-confrontation interview carried out at the end of the experiment using the camera recordings and the questioning moments identified by the researcher during the design capsule. We thereby capture the designer's focus of attention at each moment, as well as declared errors in design or in software interpretation, evaluations of the relevance of the received visuals, validations and invalidations of project elements, decision-making and reasons for choosing certain proposals, as well as the analogies made.

Then, data documenting the **designer's opinion on the potentialities and limitations** of the proposed instrumentation are collected through the questionnaire at the end of the experiment. In this way, we evaluate the perceived usefulness of the different representations, the content points of the visuals that the user found lacking, and the level of disturbance felt due to the spontaneous appearance of the representations.

Finally, data documenting the **interpreting and producing activities of the modelling agents** are collected through the camera in room 2 and screen recordings as well as through individual self-confrontation interviews. We thus compile all the visual and verbal data characterising each moment of their activity as well as a synthesis of the reasons expressed justifying their different strategies of interpretation and modelling.

DISCUSSION

We will now discuss the analyses allowed by this protocol and its feasibility.

Protocol feasibility

For an immersion of this scope, the protocol remains fairly light. It only mobilises common, intuitive or familiar tools, thus not requiring any adaptation time. The design station fits on a movable virtual desk, the modelling station consists of a common table and a large screen, to which we add 4 cameras. The whole system can be set up and configured in less than half an hour. The whole experiment only takes the participants respectively 3 or 4 hours to complete.

The data collection methods, including the think aloud, all succeeded in collecting the large amount of data expected. The design subjects were also satisfied with the architectural results of their design session and noted the great help potential of this technology. The modelling agents, on the other hand, were able to achieve good production results.

Possible analyses set

This simulation protocol is rich in gathered data. They provide information for 4 analysis tracks:

- The interpretability of architectural sketches: which elements are (in)understandable, (in)used, transformed by the agents, and are the produced models (in)useful or (in)correct?
- Design by analogy with the received visuals: what are the roles of exploration, (in)validation, rediscovery; do these analogies vary over time; what is their impact? What profile of interaction behaviour can we highlight?
- The evolution of the project: how does each feature of the building evolve from its emergence to its final validation, and how is it represented?
- Mutual learning: how do designers and modellers (and subsequently the machine) construct a mutually understandable graphic communication and is this a form of progressive learning?

Preliminary results

As a first result, we are able to analyse the activity of the modelling agents. We have already been able to identify the different parameters allowing the recognition of architectural sketches. Thus, the known context, including the expected architectural programme and the ground plans, as well as the common architectural codes, provide a good basis for the understanding of the received sketches. In addition, the information deduced from the shape associations and their relative proportions and arrangement as well as the common colour codes observed, such as blue or black for anchored elements, warm colours for annotations or furniture, green, blue and orange respectively for vegetation, water and wood, etc., provide a good framework for the understanding of the sketches. We also observe a phenomenon of progressive learning of the personal representation codes used by each subject.

As a second preliminary result, we can encode the activity of the subject designer as illustrated below in figure 4. As an example, this allows us to reconstruct the reaction of subject 6 to the visuals received in the 51st minute: after receiving the visuals, he consults them and positively comments the 3D model received as well as its adequacy with the sketch of the current project; he then declares "I see what doesn't work and that I must change" and returns to his sketch to implement changes; he then says "I see what doesn't work and what I need to change" and returns to his sketch to implement changes on the critical points.

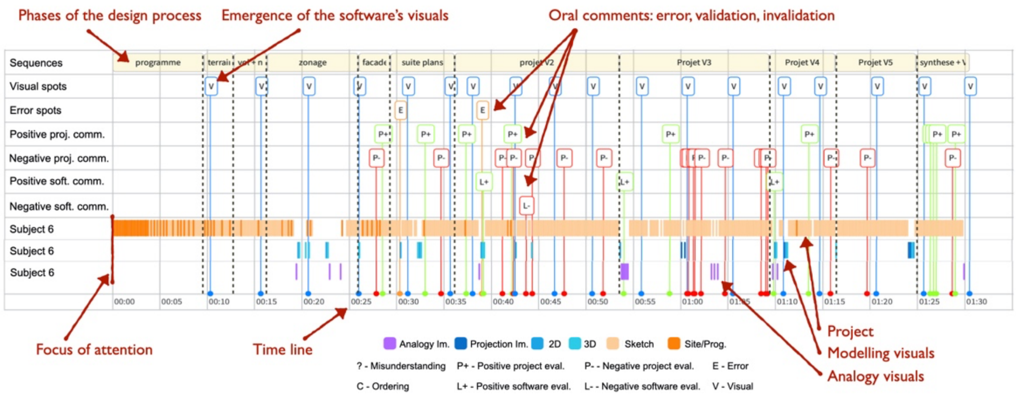

Figure 4
Visualisation of the design activity and human-computer dialogue of the subject architect

In doing so, we note different patterns in the cognitive activities of the designer as well as phenomena of reflexive conversations with the machine that progressively jointly shape the project. We also observe the impacts of the received modelling visuals, which modify the design activities, and which trigger project evaluations and/or modifications of the building characteristics.

CONCLUSION

The contribution of this paper is the development of a Wizard of Oz type protocol that allows the evaluation of a new technology proposed to smooth the ideation-BIM transition based on sketch interpretation.

This protocol consists of immersing 9 expert designers for an hour and a half in a work environment equipped with this technology. We asked them to design a house project on a graphic table. This graphic table is used to send in real time the graphic traces produced by the designer to a team of human modelling agents located in the next room. This team models the designed project in real time and sends back, via visualization screens, analogy images, cleaned up plans and a rough 3D model of the project being designed.

The protocol includes post-experimental self-confrontation interviews with the designers and with the modelling agents. The proposed protocol also includes the strategic implementation of observations based on multiple cameras and control screen to document (i) the activities of the modellers, in order to gather data on their ability to successfully interpret the received sketches, and (ii) the design and human-computer interaction activities of the designers.

The main limitation of this test protocol is related to the intrinsic structuration of a Wizard of Oz, which by hypothesis does not perfectly simulate the reaction timing of a technology, as humans are still slower at production. We observed some impact in the activities of the designers: they sometimes changed tasks while waiting for the ordered visual to be received. In consequence, the protocol is not perfectly without influence on the studied activity but did not biased the results.

In terms of perspectives, the ongoing analysis of the preliminary results will allow us to qualify the activities of the designer equipped with this digital proposal. We also will describe the modellers' ability to interpret a design subject's intent, in addition to analysing the subjects' ability to engage with the tool. We plan to carry out a series of modelling experiments with a batch of moddellers, using similar dynamically evolving sketches of architectural projects in order to generalise the rules of understanding and interpretation specific to the field of architecture observed here.

We would like to thank all the expert architects and student-jobbers who participated in this experiment as well as the F.R.S.-F.N.R.S. funding this research.

REFERENCES

Aish, R., Jabi, W., Lannon, S., Wardhana, N. and Chatzivasileiadi, A. (2018). Topologic: tools to explore architectural topology.

Baudoux, G., Calixte, X. and Lecelrcq, P. (2022). From architectural ideation to BIM: Method of Semantic Formalisation of the Building, 40th eCAADe conference, Belgium.

Browne, J. T. (2019). Wizard of oz prototyping for machine learning experiences. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-6).

Chen, X., Kang, S. B., Xu, Y. Q., Dorsey, J. and Shum, H. Y. (2008). Sketching reality: Realistic interpretation of architectural designs. *ACM Transactions on Graphics (TOG)*, *27*(2), 1-15.

Clabaugh, C. and Matarić, M. (2019). Escaping oz: Autonomy in socially assistive robotics. *Annual Review of Control, Robotics, and Autonomous Systems*, *2*, 33-61.

Cordeiro, E., Giannini, F. and Monti, M. (2019). A survey of immersive systems for shape manipulation. *Comput. Aided. Des. Appl*, *16*, 1146-1157.

Cuadra, A., Goedicke, D. and Zamfirescu-Pereira, J. D. (2021). Democratizing Design and Fabrication Using Speech: Exploring co-design with a voice assistant. In Proceedings of the 3rd Conference on Conversational User Interfaces (pp. 1-8).

Dahlbäck, N., Jönsson, A. and Ahrenberg, L. (1993). Wizard of Oz studies–why and how. *Knowledge-based systems*, *6*(4), 258-266.

Davis, R. (2002). Sketch understanding in design: Overview of work at the MIT AI lab. In *Sketch Understanding, Papers from the 2002 AAAI Spring Symposium* (pp. 24-31).

Demaret, J.-N. and Leclercq, P. (2011). An adaptative multi-agent system for architectural sketch interpretations.

Ding, C., & Liu, L. (2016). A survey of sketch based modeling systems. *Frontiers of Computer Science*, *10*, 985-999.

Elsen, E. and Leclercq, P. (2008). Sketch power to support collaborative design. CDVE, 5th Internationnal Conference on cooperative Design, Visualisation and Engineering. Springer.

Jamagne, P. (1991). Combien Project. Progress report DTP3. ISOM, a software to design dimensioning and functional organization of inner spaces. Rapport LEMA-ULg.

Jumches, R., Leclercq, P. and Azar, S. (2004). A Multi-agent system for the interpretation of architecture sketshes. In SBM (pp. 53-61).

Freitas, G., Pinho, M. S., Silveira, M. S. and Maurer, F. (2020). A systematic review of rapid prototyping tools for augmented reality. In *2020 22nd Symposium on Virtual and Augmented Reality (SVR)* (pp. 199-209). IEEE.

Frison, A. K., Forster, Y., Wintersberger, P., Geisel, V. and Riener, A. (2020). Where we come from and where we are going: A systematic review of human factors research in driving automation. *Applied Sciences*, *10*(24), 8914.

Gaffney, H., Mansell, W. and Tai, S. (2019). Conversational agents in the treatment of mental health problems: mixed-method systematic review. *JMIR mental health*, *6*(10), e14166.

Leclercq, P. and Juchmes, R. (2002). The absent interface in design engineering. *AI EDAM*, *16*(3), 219-227.

Notowidigdo, M. J. (2004). *User-directed sketch interpretation* (Doctoral dissertation, Massachusetts Institute of Technology).

Riek, L. D. (2012). Wizard of oz studies in hri: a systematic review and new reporting guidelines. *Journal of Human-Robot Interaction*, *1*(1), 119-136.

Rietz, F., Sutherland, A., Bensch, S., Wermter, S. and Hellström, T. (2021). WoZ4U: An Open-Source Wizard-of-Oz Interface for Easy, Efficient and Robust HRI Experiments. *Frontiers in Robotics and AI*, *8*.

Stahovich, T. F. (1996). SketchIT: A sketch interpretation tool for conceptual mechanical design. USA: PhD thesis

Yang, Q., Cranshaw, J., Amershi, S., Iqbal, S. T. and Teevan, J. (2019). Sketching nlp: A case study of exploring the right things to design with language intelligence. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).

Yesilbek, K. T. and Sezgin, M. (2021). On training sketch recognizers for new domains. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2142-21492