Information Extraction to Support Automation of Navigation through 3D Models during Collaborative Design Sessions

Lucian Constantin Ungureanu, Timo Hartmann
Technical University of Berlin, Germany
l.ungureanu@tu-berlin.de

Abstract. This paper explores the possibility of creating a voice-based interface to support collaborative design sessions. The main purpose is to automate changes of viewpoints for 3D models based on ongoing design conversations. We employ a design science research methodology and propose a framework to support this task. Within the structure of the framework, we propose three main components for (1) information filtering, (2) information extraction and (3) domain knowledge encapsulation. A small-scale implementation on two transcripts of real-world design sessions allows us to assess the performance of the components. The classifier for information filtering performs well for identifying the speaker turns not related to viewpoint changes of the 3D model, while additional work needs to be done to reduce the rate of false positives. It also reveals various future research directions.

1. Introduction

Creative design is often seen as a result of collaboration rather than an expression of isolated genius (*Jones, 1983*). Practitioners and researchers in the architecture, engineering and construction (AEC) field embraced this perspective. Consequently, this philosophy shaped the design and engineering thinking in the AEC field, where collaborative and coordinative design sessions represent the manifestation of this perspective. While collaborative design sessions require a strong commitment of all participants towards achieving a design goal (*Chiu, 2002*), coordinative design sessions aim at identifying issues and to define resolution pathways for these (*Mehrbod et al., 2019*).

During such sessions, creativity is enhanced by using various design artefacts. Designers interact with design artefacts and use these to support communication (*Perry and Sanderson, 1998*). Examples of such artefacts are drawings, sketches, and virtual representations such as 3D models (*Perry and Sanderson, 1998*). Most of these artefacts represent visual aids. These visualizations are meant to enhance communication, coordination and collaboration (L*eite et al., 2016; Dossik and Neff, 2011*). Computers started to play an essential role during such sessions, especially considering the high adoption rate of building information modelling (*Golparvar-Fard et al.,2013*). The use of 3D models to support collaboration and communication has been widely adopted. For example, a survey conducted in the U.S. (*Leite et al., 2016*) reported that 3D models represent the most commonly used form of visualization, with a score of 95%. Moreover, technological tools are supposed to “stimulate, cause or assist” creativity (*Achten and Beetz, 2009*). A review of collaborative design studies (*Achten and Beetz, 2009*) reported that 40% of the studies focus on technological aspects, a conclusion also enforced by later studies (*Dossik and Neff, 2011*). At the same time, the static nature of 3D representations does not allow for a natural interaction with the virtual artefacts (*Dossik and Neff, 2011*).

In practical settings, designers expect to be able to points to (gesture), talk about (speech) or sketch on (sketching) (*Perry and Sanderson, 1998*). When it comes to gesture and sketching, emerging technologies such as touch-based hardware are possible alternatives towards a more natural designer-artefact interface. Various studies (*Kim and Maher, 2008; Gu et al., 2011*) started investigating the effect of tangible interfaces on design creativity. A study comparing gesture-based communication with sketching-based communication (*Eris et al., 2014*) highlight that speech is the main communication channel, while gesture and sketching represent secondary means of communication. To our knowledge, speech-augmented designer-artefact interaction does not currently represent a mature research area.

This paper investigates the possibility to automatically extract information from design communications to automatically change viewpoints in 3D models. During design sessions, participants naturally communicate their ideas, intentionally referring to some detail of the design. However, navigating a 3D model to find a viewpoint that shows this detail is a tedious task. Moreover, most of the times, participants in such sessions do not have the required skill set to quickly change a 3D model's viewpoint. A natural language interface between participants and computer might represent a good alternative. However, contemporary voice assistants are intrusive as they require formalised queries.

In this study, we are looking at automating the extraction of viewpoints-related information out of natural human-to-human conversations. To this end, we propose a framework which combines ontological modelling of domain-specific knowledge and the use of probabilistic models to support information filtering and extraction from natural language conversations.

The paper continues with a more detailed overview of state-of-the-art possibilities to interact with digital design artefacts. The paper then introduces the steps we followed to conduct this research as well as the framework we proposed to solve the problem of automatically changing the viewpoints of 3D models. Moreover, we present the implementation of the proposed framework using the transcripts of two real-world design meetings. The implementation is followed by the evaluation of various parts of the framework and discussions of the current status of the framework, where we propose various scenarios for future development. Finally, we present a summary of our findings, limitations and future research in the conclusion section.

2. Human-Computer Interfaces in Design

Human-Computer interfaces usually require a set of input and output devices. Input devices allow a user to provide a set of commands which are translated in specific machine-interpretable actions. The user accesses the results of the provided commands using output devices. Keyboard and mouse represent the most common input devices, while screens represent the most common output devices. Early research in the field of computer-aided design aimed to assess the human performance while using machines for design tasks (Cuomo and Sharif, 1989) emphasising the need of training to improve the human performance, while more recent research start focusing on the need of designing better interfaces aiming at a more natural and intuitive way of interacting with machines (Turk, 2014).

To this end, the research community highlighted that multi-modality is an inherent characteristic of human communication (Turk, 2014). Aiming to move away from the keyboard-mouse-based input interface, the research community started exploring various interfaces which will allow the humans to provide input through gesture, sketching and speaking. When it comes to design, the purpose is to allow the computers to meaningfully support the design process without hindering creativity.

On one hand, as a result of the rapid technological development, gesture and sketch-enable technologies are slowly integrated into the design process. Studies show that these tangible interfaces increase creativity (Kim and Maher, 2008; Gu et al., 2011). On the other hand, there is a huge gap in the use of speech-based interfaces during collaborative design sessions. Most of the previous research focused on developing voice-based input interfaces with a limited number of voice-based commands which allow users to access various CAD-software functionality such as modelling geometrical primitives (Chu et al., 1997; Kou and Tan, 2008; Xue et al., 2009; Kou et al., 2010) or accessing information linked to 3D elements (Xue et al., 2010, Behera and McKay, 2017). A limitation of the proposed systems is that users always need to remember what they can ask the computer. Additionally, expanding the repository of available commands is very difficult. Finally, these commands are highly structured and disrupt the natural course of design conversations. Advanced approaches which make use of the current state of the art natural language techniques are still missing, but moreover, there is a lack of understanding the specificity of the natural language conversations used in design-related situations. This paper employs a design science research method which aims at laying the foundation for the development of non-intrusive voice-based interfaces, with the main goal of changing the 3D models' viewpoints based on information extracted out of natural language conversations held by the practitioners during collaborative and coordinative sessions.

3. Research Method and Framework Description

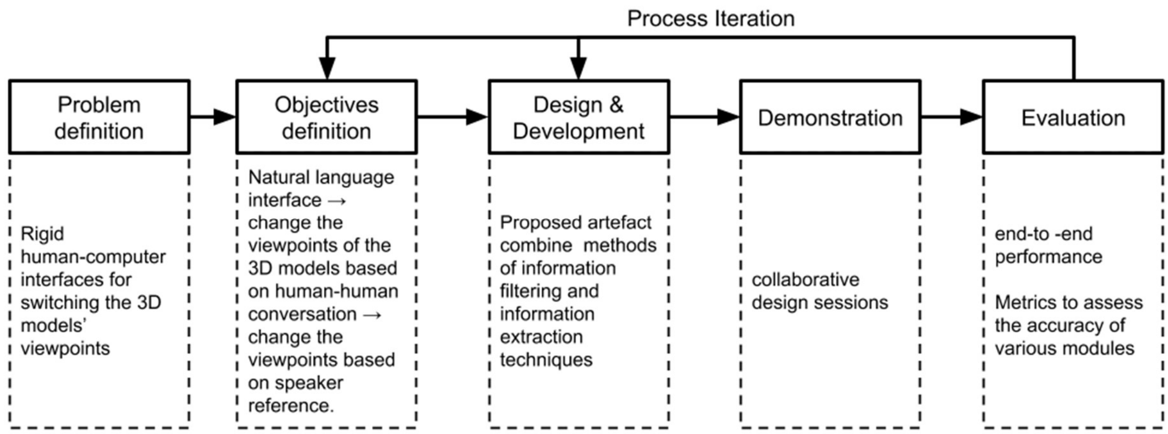

Figure 1-Design Science Research Method with Problem-Centred Initiation

This paper employs a design science research method following the framework proposed by Peffer et al. (2007). Figure 1 presents the problem-centered research method which consists of five main components. Extracting the right information of out human-to-human conversations is a challenging task mainly because there is no distinction between commands-related conversation and background conversation (messy talk). The objective of the current paper is to develop a system that will extract the information related to various parts of the product to be designed and perform operations such as zoom in or isolate the parts mentioned by designers.

For example, when designers exchange design ideas about the first floor, the operation expected from the system is to isolate the first floor as shown in Figure 2.

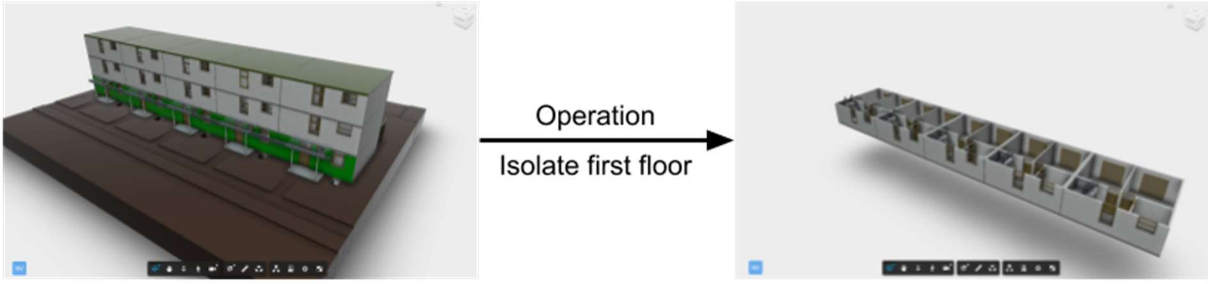

Figure 2-Possible 3D model's viewpoint operation

Figure 3 presents the proposed framework. Within this framework, we distinguish three major components, namely: (1) information filtering, (2) information extraction, and (3) domain knowledge base. Information filtering represents the process of removing data from an incoming stream with the sole purpose of keeping the information relevant for a task (*Belkin and Croft, 1992*). Information retrieval represents the process of extracting the information required to perform a task. Both, information filtering and information retrieval are contexts sensitive. In our framework, this gap is addressed using the third component – domain knowledge.

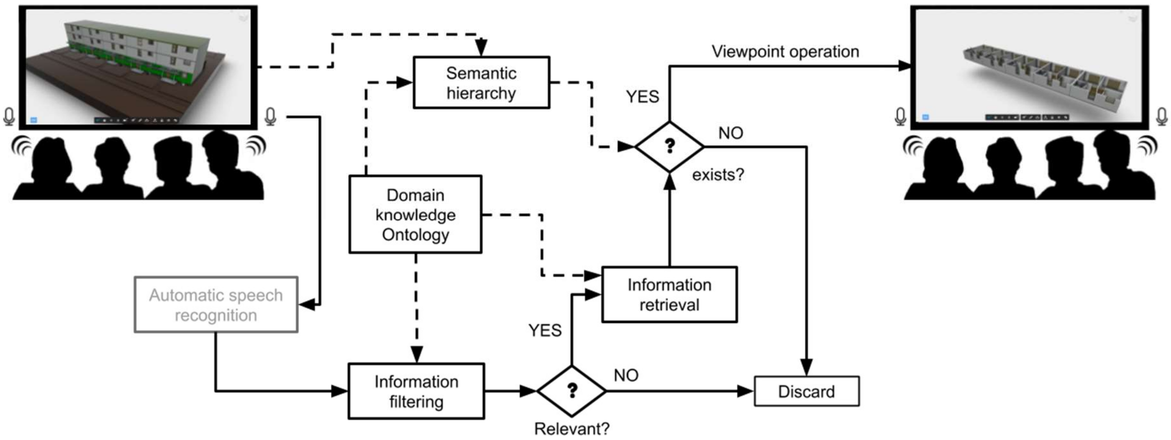

Figure 3-Artefact framework

To develop this system, we use a dataset of recordings of design conversations. We aim at extracting specific information for viewpoint changes for this dataset. As design conversations are unstructured, we structure the cleaning process in two subtasks: (1) automatic detection of those utterances and natural language constructs not related to viewpoints changes and (2) a manual cleaning necessary as preparation for the information extraction. For the automatic detection of task-related constructs, we explore the use of probabilistic classifiers. A classifier can identify constructs not related to any viewpoint operations and discard these. To support information filtering and information extraction, we built an ontology. This ontology aims to capture the knowledge related to the project breakdown. The central assumption we make is that participants use domain-specific named concepts during conversations. We use the error rate to individually evaluate the proposed components for information filtering and information extraction to conduct an end-to-end validation of the entire framework.

For a small-scale practical implementation of the framework, we used the transcripts of two architectural meetings recorded by the Design Thinking Research Group (Llyod et al. 2005). These meetings are part of an effort to design a crematorium and were recorded in their natural setup. Both meetings combined consist of 2318 speaker turns and almost four hours of recorded time. Three participants were involved during the first meeting (A1) and five participants for the second meeting (A2). In both meetings, a researcher was present as an observer. During the design sessions, participants focus on designing various parts of the crematorium while using paper-based drawing to visualize various parts of the design. At this point, it is worth mentioning that no computer-based artefacts were used during these sessions. In this paper, we address the issue of information filtering and information extraction supported by an ontology created for the crematorium domain.

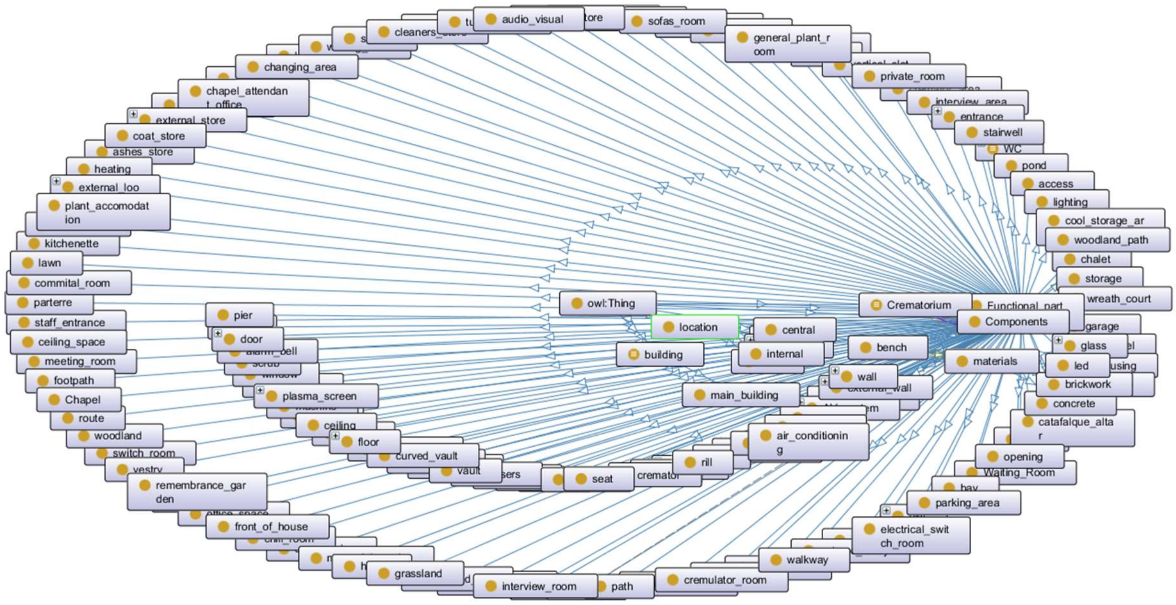

Figure 4-Crematorium Ontology

Figure 4 presents the crematorium-related domain concepts included in the ontology. The ontology has concepts *functional parts (*which include concepts such as *waiting room* or *chapel* which describe the functional aspects of a crematorium project), components (which included concepts such as *wall* or *roof*), locations (which include concepts such as *back* or *top*), and materials (which include concepts such as *concrete*). These concepts were used to automatically label speaker turns which contain these concepts. For the information filtering, we prepared the transcripts for a binary classification problem, where a new feature was created to indicate if a specific speaker turn is or is not a possible candidate for a viewpoint change, mainly based on content related to the ontological concepts. After labelling 541 speaker turns were labelled as related, while 1777 were labelled as not related. In preparation for the classification, we performed additional cleaning tasks such as remove numbers, stop words, and punctuation, steaming and characters case folding from upper to lower case. After cleaning, a document term matrix was created. In a document-term matrix, the number of rows is equal to the number of documents while the number of columns is equal to the number of unique terms in the dataset. For our implementation, the number of rows is equal to the number of turns (2318) and the number of columns corresponds to 2224 unique terms. Each term is weighted based on its frequency on each turn. For the classification, 75% of the dataset was used for training and the other 25% were used for testing. Within the training subset, 1320 turns correspond to negative class and 418 turns correspond to the positive class. Within the testing subset, 457 turns correspond to negative class and 123 turns correspond to the positive class. We implemented a simple Naïve Bayesian classifier for which we disabled the Laplace smoothing (Laplace = 0).

For the information extraction, we implemented a natural language parser and we extracted the noun phrases (NP) from each turn. These phrases were once again filtered, using pattern matching to keep only those which are part of the crematorium ontology. The performance of the information filtering module depends on the language used during design and on the concepts captured in the ontology. When it comes to language use, one of the major concerns relates to syntactic ambiguity. Syntactic ambiguity means that a sentence can have more than one parse trees, which means that some parts can have multiple syntactic roles within the uttered sentence. As mentioned previously, those concepts captured in the ontology must reflect the structure of the design artefact (e.g. 3D models). For an in-depth analysis, we can consider the following sentence extracted from the transcripts.

"*Basically we're arriving in the new car park in this area and from the car park we'll enter the building through a waiting area*".

The syntactic parser identifies the following items as noun phrases: "*we*", "*the new car park*" "*this area*", "*the car park*", "*we*", "*the building*", "*a waiting area*". Using the concepts included in the ontology, the following terms are extracted after filtering: "*building*", "*waiting area*", "*car park*". We notice that the quantifier "*new*" used by the speaker to make a separation between the existing car park and the new car park is missing.

4. Evaluation and Discussions

The results of the classifier indicate a sensitivity of 53.91% and a specificity of 94.42%, with a balanced accuracy of 74.16%. The classifier performs well in terms of identifying those turns not related to viewpoint changes. The true positive rate is low, and future efforts should focus on improving this and reducing the false positive rate as these correspond to turns related to viewpoint changes. Moreover, one limitation of the current exploration is given by structuring the conversations on speaker turns. This approach does not capture the real behaviour of the speaker. Segmentation of the speaker turns in small units called utterances must be employed in this direction. Moreover, experimentation with other classification methods should be the subject of future studies, especially with those which perform well on short text classification.

The use of natural language processing techniques, such as parts of speech tagging and syntactic parsing, is usually computationally intensive. In this direction, the use of the classifier to discard the turns not related to viewpoints helped to reduce the workload for a such as task. For an utterance such as "*basically we're arriving in the new car park in this area and from the car park we'll enter the building through a waiting area*" the extracted parts will be [*"car park", "car park", "building", "waiting area"*]. "*car park*" and "waiting area" represent specific parts which can be easily identified, and the system can perform operations such as isolating or zooming in. When it comes to the concept "building" it is not clear to what building the speaker refers to. A limitation of this study is that it is not considering the relationships defined in the ontology. A relationship such as "*crematorium*" has some "*waiting room*", "*waiting room*" has some "*wall*", "*wall*" has some "*window*" will allow us to identify when the speaker refers to "*that specific window which is part of the wall which is part of the waiting room*", and not to another window.

Ontologies also play an important role in improving the performance of the voice assistants such as Alexa or Siri. Most of the voice assistants make use of WikiData repository which is mainly built for general purpose tasks. Design is a very specialized activity for which repositories such as WikiData do not manage to capture all required knowledge.

One major limitation of both components proposed for information filtering and information extraction is that it covers only those situations where the speaker directly refers to the thing, without considering that the speaker might refer to a thing using various expressions such as "*this"* or "*it"*. Future studies will aim to investigate the use of coreferential expressions in design conversations. Moreover, future efforts will focus on creating an instantiation of the proposed artefact for the purpose of end-to-end testing. Moreover, any operation leading to the change of viewpoints in 3D models need to clearly identify the input data required to be performed. Making the operation indicated in Figure 2 requires additional domain knowledge such as knowing which elements are part of level 1. This points out toward the fact that the ontology needs also to capture knowledge related to the semantic hierarchy of the 3D models.

5. Conclusions

The current paper proposes an artefact which aims to address the gap of voice-based interfaces in collaborative design sessions. This research follows the design science methodology with the aim of creating a framework to solve the above-stated problem. Precisely, the framework aims to analyse the participants' conversations with the purpose of filtering and extracting information related to the viewpoints of the 3D models. For the filtering task, we proposed the use of a classifier, while information extraction relies on natural language processing techniques such as syntactic parsing.

A small-scale implementation was performed using the transcript of two real-world design sessions. The design science research method proposes an iterative approach to develop the framework. Each iteration reveals various aspects which need to be addressed by future research. In the case of our implementation, the classifier proposed for information filtering performs well for the classification of the true negatives (turns not related to 3D models' viewpoints). The performance of the classifier on true positives indicates that future research should focus on reducing the false positives, aim to capture the real speakers' behaviour by segmentation of the transcript in small utterances, and exploring the performance of various classification methods especially focusing on short text classification. Implementation of the information extraction module indicates that future research should focus on making better use of the knowledge encapsulated in the ontologies.

References

Achten, H. and Beetz, J., 2009, September. What happened to collaborative design?. In *Proceedings of the eCAADe Conference* (pp. 357-366).

Behera, A.K. and McKay, A., 2017, September. Designs that talk and listen: Integrating functional information using voice-enabled CAD systems. In Proceedings of 2017 25th European Signal Processing Conference. Leeds.

Chiu, M.L., 2002. An organizational view of design communication in design collaboration. *Design studies*, *23*(2), pp.187-210.

Chu, C.C.P., Dani, T.H. and Gadh, R., 1997. Multi-sensory user interface for a virtual-reality-based computer-aided design system. Computer-Aided Design, 29(10), pp.709-725.

Cuomo, D.L. and Sharif, J., 1989. A study of human performance in computer-aided architectural design. International Journal of Human-Computer Interaction, 1(1), pp.69-107.

Dossick, C.S. and Neff, G., 2011. Messy talk and clean technology: communication, problem-solving and collaboration using Building Information Modelling. *The Engineering Project Organization Journal*, *1*(2), pp.83-93.

Eris, O., Martelaro, N. and Badke-Schaub, P., 2014. A comparative analysis of multimodal communication during design sketching in co-located and distributed environments. *Design Studies*, *35*(6), pp.559-592.

Golparvar-Fard, M., Tang, P., Cho, Y.K. and Siddiqui, M.K., 2013. Grand challenges in data and information visualization for the architecture, engineering, construction, and facility management industries. Computing in Civil Engineering (2013), pp.849-856.

Gu, N., Kim, M.J. and Maher, M.L., 2011. Technological advancements in synchronous collaboration: The effect of 3D virtual worlds and tangible user interfaces on architectural design. *Automation in Construction*, *20*(3), pp.270-278.

Jones, J.C., 1983. Continuous design and redesign. *Design Studies*, *4*(1), pp.53-60.

Kim, M.J. and Maher, M.L., 2008. The impact of tangible user interfaces on spatial cognition during collaborative design. *Design Studies*, *29*(3), pp.222-253.

Kou, X.Y. and Tan, S.T., 2008. Design by talking with computers. Computer-Aided Design and Applications, 5(1-4), pp.266-277.

Kou, X.Y., Xue, S.K. and Tan, S.T., 2010. Knowledge-guided inference for voice-enabled CAD. Computer-Aided Design, 42(6), pp.545-557.

Leite, F., Cho, Y., Behzadan, A.H., Lee, S., Choe, S., Fang, Y., Akhavian, R. and Hwang, S., 2016. Visualization, information modeling, and simulation: Grand challenges in the construction industry. *Journal of Computing in Civil Engineering*, *30*(6), p.04016035.

Lloyd, P., McDonnell, J., Reid, F., Luck, R., 2005. DTRS7-Dataset from 7th design thinking research symposium, http: //design.open.ac.uk/dtrs7

Mehrbod, S., Staub-French, S., Mahyar, N. and Tory, M., 2019. Characterizing interactions with BIM tools and artifacts in building design coordination meetings. *Automation in Construction*, *98*, pp.195-213.

Peffers, K., Tuunanen, T., Rothenberger, M.A. and Chatterjee, S., 2007. A design science research methodology for information systems research. *Journal of management information systems*, *24*(3), pp.45-77.

Perry, M. and Sanderson, D., 1998. Coordinating joint design work: the role of communication and artefacts. *Design studies*, *19*(3), pp.273-288.

Turk, M., 2014. Multimodal interaction: A review. Pattern Recognition Letters, 36, pp.189-195.

Xue, S., Kou, X.Y. and Tan, S.T., 2009. Natural voice-enabled CAD: modeling via natural discourse. Computer-Aided Design and Applications, 6(1), pp.125-136.

S. Xue, X. Y. Kou & S. T. Tan (2010) Command Search for CAD System, Computer-Aided Design and Applications, 7:6, 899-910