GesCAD: An Intuitive Interface for Conceptual Architectural Design

Sumbul Khan[1]
sumbul_khan@sutd.edu.sg

Hasitha Rajapakse [1,2]
hasitha@ahlab.org

Haimo Zhang[1,2]
haimo@ahlab.org

Suranga Nanayakkara[1,2]
suranga@ahlab.org

Bige Tuncer[1]
bige_tuncer@sutd.edu.sg

Lucienne Blessing[1]
lucienne_blessing@sutd.edu.sg

[1]SUTD-MIT International Design Center, Singapore University of Technology and Design, 8 Somapah Road, Singapore
[2]Augmented Human Lab', Singapore University of Technology and Design, 8 Somapah Road, Singapore

ABSTRACT

Gesture- and speech-based 3D modeling offers designers a powerful and intuitive way to create 3D Computer Aided Design (CAD) models. Instead of arbitrary gestures and speech commands defined by researchers, which may not be intuitive for users, such a natural user interface should be based on gesture and command set elicited from the users. We describe our ongoing research on a speech-and-gesture-based CAD modeling interface, GesCAD, implemented by combining Microsoft Kinect and Rhino, a leading CAD software. GesCAD is based on gestures and speech commands elicited from a specially designed user experiment. We conducted a preliminary user study with 6 participants to evaluate the user experience of our prototype, such as ease of use, physical comfort and satisfaction with the models created. Results show that participants found the overall experience of using GesCAD fun and the speech and gesture commands easy to remember.

CCS CONCEPTS

• **Human-centered computing~Gestural input** • Human-centered computing~HCI theory, concepts and models

KEYWORDS

Gesture recognition; speech recognition; 3D CAD modeling

ACM Reference format:

S. Khan, H. Rajapakse, H. Zhang, S. Nanayakkara, B. Tuncer, L. Blessing. 2017. GesCAD: An Intuitive Interface for Conceptual Architectural Design. In Proceedings of the 29th Australian Conference on Human-Computer Interaction, Brisbane, QLD, Australia, November 2017 (OzCHI 2017), 5 pages.

https://doi.org/10.1145/3152771.3156145

OzCHI '17, November 28-December 1, 2017, Brisbane, QLD, Australia

ACM ISBN 978-1-4503-5379-3/17/11…$15.00
https://doi.org/10.1145/3152771.3156145

1 BACKGROUND AND INTRODUCTION

In the conceptual design stage, especially in the architectural design process, designers create new artifacts whose properties are only partially known, and strive to define the components of the designed object and the relationships between them [5]. As a part of this process, designers perform formal explorations that quickly place the various programmatic, environmental, and urban code related needs and constraints into a geometric and formal framework. These are called massing studies, where designers create various volumetric 'masses' and compositions for building blocks and test how these perform in fulfilling the design requirements. This research aims to support designers in their formal explorations during the massing study phase, when directly creating and manipulating objects in a 3D modeling, or CAD environment.

Speech and gestures are inter-dependent components of human communication. While speech is seen to be suitable for descriptive tasks, gestures are appropriate to handle spatial issues [3]. We define gestures as the hand and arm movements in free space that convey meaningful information. The focus of this research is mid-air, touch free, bimanual gestures.

Traditional CAD systems use graphical user interfaces following the 'Windows, Icon, Mouse and Pointer' (WIMP) paradigm, which is little intuitive and hard to learn, especially when it comes to the conceptual design stage [15]. In order to facilitate 3D modeling at this stage of design, designers would greatly benefit from an efficient and intuitive interaction style that makes spatial manipulation and perception of 3D geometry and transformations easier [19]. Intuitive and flexible interaction techniques will enable designers to model more effectively, and possibly design better, when using CAD during conceptual design. An emerging and promising interaction metaphor is gesture-based interaction [2, 7]. Gesture-based interaction is inherently intuitive, because humans use their body, rather than a device, to interact with machines. We claim that through the ubiquity of gestures when articulating design geometry, the use thereof can help designers communicate with computers and peers about CAD models in a more natural, intuitive, and effective way.

Even though gesture based interaction has developed considerably in the recent years [8, 17, 20], many gesture-based interfaces still mainly use the hands to emulate the mouse with a limited number of gestures [22]. Past studies have largely focused on gesture recognition techniques using small sets of author-defined gestures, overlooking the specific needs of designers in terms of the speech and gesture vocabulary [2, 7, 9]. Often, users would have to learn a device- or application-dependent artificial gestural language that comes with a learning curve [13]. Although user-centred gesture design for natural user interfaces has been investigated in related domains [1, 12, 21], we found no study that employed user-elicited gestures for an interface for 3D CAD modeling for conceptual design.

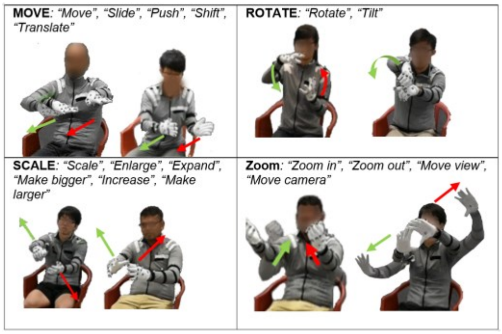

Figure 1: Subset of gestures and speech commands for manipulation commands. Movements of the right and left hands indicated by green and red arrows respectively.

Recent research has conducted small user studies to test gesture- and speech-based interaction for 3D CAD modeling. These studies confirm that users were able to adapt to new environment quickly [6] and were able to complete tasks with mid-high precision within reasonable time [14]. However, limited studies on gestural interfaces for CAD modeling have conducted user studies on factors such as memorability of gestures, ease of using gestures and speech, or user satisfaction – aspects which are fundamental for the success of a new interface for CAD modeling.

We believe that a robust gesture- and speech-based CAD modeling interface must be based on empirically determined gestures elicited from the users, instead of arbitrary gestures designed by creators of the interface. It should be based on an analysis of the natural ways speech and gestures are employed by designers. Furthermore, a robust speech- and gesture-based modeling system must be flexible and extensible. It should adapt to users' needs and gestural techniques [23]. And finally, a successful speech- and gesture-based CAD modeling system must employ viable, low cost technology that can be used with ease in office and educational environments. In this paper, we present GesCAD, a proof-of-concept prototype that runs on a standard notebook computer, which combines a Kinect (v2, see http://www.xbox.com/en-SG/xbox-one/accessories/kinect) and Rhinoceros 5 (https://www.rhino3d.com/). GesCAD is based on a subset of gestures and speech commands elicited from a user experiment previously reported in [10, 11, 18]. We conducted a preliminary usability evaluation of the system such as ease of use, physical comfort and satisfaction with modeling results and report findings in this paper. The novel contributions of this work are as follows:

- Integration of empirically elicited set of specialized gesture and speech commands into a CAD modeling environment
- Development of a proof-of-concept system implementation that combines a mainstream 3D modeling software and commercially available gesture tracker. Our implementation is flexible as it employs a many-to-one mapping of gestures and speech to CAD functionalities.

2 GESCAD SYSTEM

2.1 GESTURE ELICITATION STUDY

We conducted a speech and gesture elicitation study to discover the natural interactions used in CAD modeling. The experiment was conducted with 41 participants (52% female, 48% male) with architecture (49%) and engineering backgrounds (51%), over a period of two weeks at the Singapore University of Technology and Design. Participants sat at a distance of 10' from a screen which showed images and short video clips of CAD modeling referents in three categories: primitives, manipulations and navigations. The participants' task was to articulate the gestures using (1) only gestures, in session A; and (2) gestures and/or spoken words, in session B. The order of the sessions was counter-balanced. Two video cameras were used to capture the participants' gestures from different angles. Video data from the experiment was analyzed by three coders. The coding scheme identified morphological themes in the gestures, and the command words in the speech transcription. For example, for the CAD functionality 'Zoom in', we identified three morphological themes based on how hands were used: (1) Push hands towards screen (2) Pull hands towards self, and (3) Pull hands away from each other [10, 18]. The study revealed that participants preferred bimanual gestures and chose to employ both speech and gestures for communicating CAD functionalities. The study also compiled the most frequent speech terms that designers used for the description of 3D objects and operations; and recorded the multiple gestures and speech commands for the same CAD functionality [11]. For the GesCAD implementation, a subset of the gestures for the CAD manipulation functionalities, namely move, rotate and scale, and the navigation functionality, namely zoom, were selected for implementation (Fig. 1).

2.2 Development of Interface

The prototype was developed in the form of a plugin for the Rhinoceros 5 software, in the C# language. It consists of 3 main parts: a CAD Module, a Speech Module, and a Gesture Module. The CAD module supports various operations in 3D CAD modeling, such as creating geometric primitives, performing selected geometric transformations and changing viewpoint. The CAD module was developed using RhinoCommon API in Rhino.NET SDK (Fig. 2). The Speech Module captures speech commands, converts them to text, and maps them to the respective functionalities. Speech commands are captured by the Microsoft Kinect device through its microphone array; the captured audio data is then processed using Kinect for Windows SDK for noise cancellation; the audio data is transcribed into English text using Microsoft Speech Platform SDK and supported language packages, which accommodates various accents in English. Finally, the text is mapped to a functionality provided by the CAD module. In order to allow for flexibility in natural language speech control, multiple speech commands of similar meanings are mapped to single functionalities, in what we term as a 'many-to-one mapping'. For example, 'move', 'translate', and 'shift' are all mapped to the functionality of geometric translation in the CAD module. When using multiple objects the 'switch' command is used to switch between objects.

The Gesture Module maps continuous gestures to control parameters in a CAD functionality. For example, the direction of the user's hand motion is mapped to the direction of translation of the 3D model. Kinect for Windows SDK recognizes 25 joints in a detected body. Gestures are recognized as movements of 10 selected joints on both arms between consecutive frames in 3D space. Selected gestures from a previous study was implemented with the help of these detected joints. Users can use multiple gestures to control the parameters of a single CAD functionality, to allow flexibility, another example of a many-to-one mapping (Fig. 1). The prototype can be installed in Rhinoceros as a regular plugin using Rhino Plugin Manager. Once the plugin is installed, it is activated by typing the word 'GesCAD' in the Rhinoceros command line. The program starts listening to the speech commands as soon as the plugin is activated. When a speech command is matched with a functionality, it activates the gesture recognition process and performs the relative CAD software action in real time until the 'STOP' speech command is given.

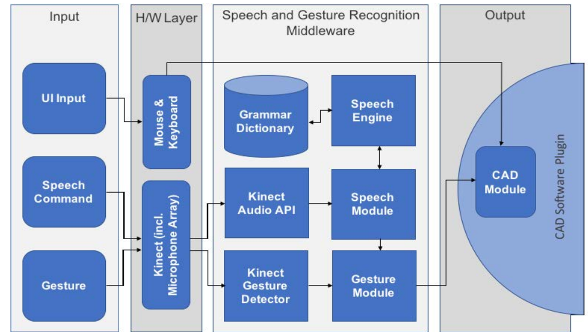

Figure 2: GesCAD system diagram

3 EVALUATION OF THE SYSTEM

A preliminary user study was conducted to analyze the usability of the GesCAD system. Six participants (3 females, 3 males) from architectural background took part in the study. Average age of the participants was 32 (SD=3.89). Participant occupation included PhD students and Research Assistants. All participants were proficient in English and had considerable experience with at least one 3D CAD modeling software using mouse and keyboard.

The user study was conducted over a period of two days, in an experiment setup at the Singapore University of Technology and Design. The computer display was projected on a screen. The participant sat on a chair at a distance of six feet from the screen with the Kinect sensor placed on a table at a distance of three feet between the screen and the participant. Participants were first given a demonstration on how to execute commands in Rhino using gestures and speech. Participants were then given a custom-made manual showing the commands with their associated gestures and speech terms. Then, they could freely practise using the interface for a maximum of fifteen minutes. During this time, they could refer to the manual, or ask the experimenter any questions they had.

After the training, participants were given four tasks that involved the creation of a 3D box and manipulating it by (1) rotating, (2) scaling and (3) moving, and (4) navigating view by zooming in/out. They were also given a basic modeling task that involved the manipulation of two boxes. All tasks were given in the form of before and after images printed on handouts. The order of the tasks was randomized. After the completion of every task, participants were asked to rate the task on a 5-point Likert scale on two aspects: (1) the ease of performing the task, (2) satisfaction with the modeling results: how well the results matched what participants intended. After the end of all tasks, participants also rated their overall experience of using the system on the following aspects: (1) fun in performing the tasks, (2) how easy it was to remember commands, (3) physical comfort, and (4) potential for creative exploration. Participants also expressed their opinion about the system in an open-ended question. An experimenter took notes during each session of the user study, which took a maximum of one hour to complete per subject.

3.1 Results

All participants completed all tasks successfully. Median ratings were higher for the aspects of 'fun' (Mdn=4.5) and 'easy to remember' (Mdn=5), and slightly lower for the aspects of 'physical comfort' (Mdn=4) and 'potential for creative exploration' (Mdn=4) (Fig. 3a). During training, it was observed that participants made little effort to go through the manual or enquire about gestures from experimenters. Instead, they freely improvised the gestures they had seen in the demo. Participants' perceptions of the ease of performing the task was high for all tasks, except for Rotate and Modeling tasks, which were slightly lower (Fig. 3b). Median ratings for satisfaction with modeling results were also high for Move and Scale; and slightly lower for Rotate, Zoom and the Modeling tasks. Participants took the least time to complete the zoom task (Mean=20s, SD=6.06), and the most time to complete the rotate task (Mean=50s, SD=42.16). The modeling task, which involved the creation of two boxes, and manipulating them by scaling, rotating and moving, took approximately two minutes to complete on average (Fig. 3c). In the open-ended question, participants indicated that they found the commands ‘easy to use’ (participant 3), ‘intuitive’ (participant 2) and ‘great to explore conceptual ideas’ (participant 5). However, two participants (2 and 6) voiced concern over issues of precise input. Another two participants expressed concern that the rotate functionality was ‘hard to control’ (participant 1) and ‘not very responsive’ (participant 3).

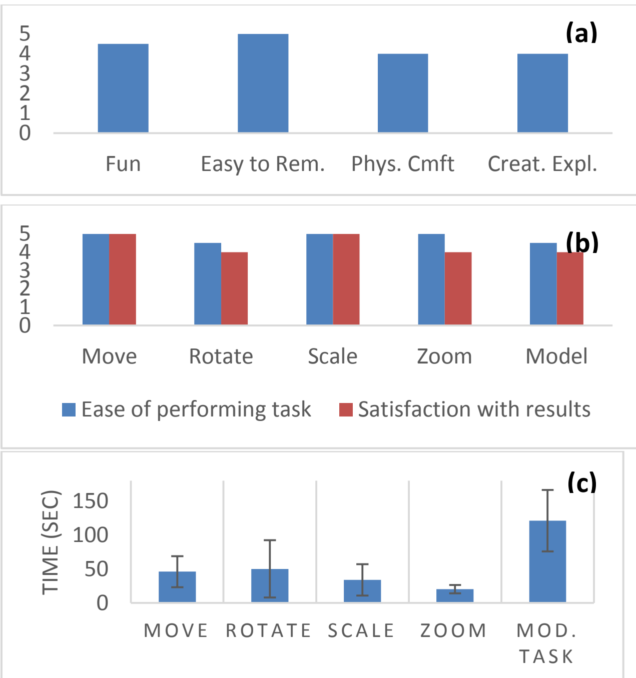

Figure 3: (a) Median ratings for overall experience (b) Median ratings for tasks (c)Average time taken to complete tasks, in seconds.

4 DISCUSSION AND CONCLUSION

In this paper, we presented a proof-of-concept system implementation of a speech- and gesture-based interface, GesCAD, for 3D CAD modeling in Rhino for conceptual design. We conducted a preliminary user study to test our implementation, whose gesture and speech command set is elicited from users. Results show that the participants found the overall experience of using GesCAD ‘fun’ and the speech and gesture commands ‘easy to remember’. Ratings of ease and satisfaction with modeling results were positive for all aspects.

Two key points in our approach differentiate the GesCAD system from previous multi-modal interfaces for CAD modeling [6, 14]. Firstly, the gestures and speech terms used in the GesCAD system were determined by a gesture elicitation experiment for CAD modeling conducted with actual designers, and based on in-depth research on the use of speech and gestures in CAD modeling [10, 11]. This approach ensured that the participants in the user study found the gestures easy to perform and intuitive. Secondly, we employed a many-to-one mapping from gestures and speech to functionalities, similar to the use of gestures in natural human interaction [4]. This ensures flexibility in using gestures and speech input. Such flexibility is crucial to user experience, since participants will not have smooth user experience if they are constrained by a limited vocabulary or gestures that need to be learned. Using the GesCAD interface, participants freely improvised gestures, without the spending time to learn the gestures through the manual. We attribute participants’ positive response to GesCAD to these two key points in our approach and consider these to be the chief contributions of this study.

Implementation of the gestures for zoom and scale commands was more successful than that of move and rotate, both of which involve object manipulation in 3D space. This may be due to the mismatch between users’ perception of objects’ orientation in physical space and that in Cartesian space. For instance, slight misalignment in hands may affect the identification of the direction of rotation. We believe such issues led to longer completion time for Move, Rotate and the Modeling tasks, and affected users’ ratings of physical comfort and potential for creative exploration, which were slightly lower than other aspects. Such issues of precision in user input may be addressed by improving the speech input for more complex commands [16], such as specifying distance, axis of rotation, and angles.

This study, which tested only four basic CAD functionalities with a small number of participants, was a preliminary investigation and hence findings can only be considered indicative. We acknowledge that the use of Likert scale has limited validity in case of small sample sizes. Future investigation includes expanding the CAD functionalities in the GesCAD system and testing them with more participants. A detailed comparison of the GesCAD system with traditional mouse and keyboard input to analyze aspects such as completion time, accuracy, and cognitive load such as the NASA Task Load Index (NASA-TLX) would yield more comprehensive findings regarding the usability of the GesCAD system. The system can be extended by adding a simple GUI and system training module, so that users can expand the CAD functionalities with new gestures and speech commands [22, 23]. In the future, the system will also incorporate a machine-learning algorithm that supports individual preferences of modeling using gestures and speech. As opposed to previous studies that focus on accuracy of gesture recognition, we employed a user centered approach for the design of the speech and gesture interface. Flexibility in user input and speech modality makes this system less dependent on gesture recognition accuracy or complexity. Hence this study demonstrates a practical application of combining speech and low-resolution gesture input for conceptual design using CAD modeling. It provides designers with a technique that is touch-free and intuitive. An approach such as this promises to bring multi-modal interfaces to the forefront for conceptual design.

ACKNOWLEDGMENTS

This work was supported by the SUTD-MIT International Design Centre (IDC) grant number IDG21500109, under the Sustainable Built Environment Grand Challenge, and Visualization and Prototyping Design Research Thrust.

REFERENCES

[1] R. Aigner, D. Wigdor, H. Benko, M. Haller, D. Lindbauer, A. Ion, S. Zhao and J. T. K. V. Koh. 2012. Understanding Mid-Air Hand Gestures: A Study of Human Preferences in Usage of Gesture Types for Hci. *Microsoft Res. Tech. Report*, 2012, 10.

[2] S. Arroyave-Tobón, G. Osorio-Gómez and J. F. Cardona-McCormick. 2015. Air-Modelling: A Tool for Gesture-Based Solid Modelling in Context During Early Design Stages in Ar Environments. *Computers in Industry* 66, 73-81. http://dx.doi.org/10.1016/j.compind.2014.10.007

[3] U. Athavankar. 1999. Gestures , Mental Imagery and Spatial Reasoning. *Preprints of the International Conference on Visual Reasoning (VR 99)*, 103-128.

[4] J. Cassell. 1998. A Framework for Gesture Generation and Interpretation. In *Computer Vision in Human-Machine Interaction*,191–216. http://dx.doi.org/10.1017/CBO9780511569937.013

[5] N. Cross. 1982. Designerly Ways of Knowing. In *Designerly Ways of Knowing*, Springer, London, 1-13. http://dx.doi.org/10.1007/1-84628-301-9_1

[6] D. Dave, A. Chowriappa and T. Kesavadas. 2013. Gesture Interface for 3d Cad Modeling Using Kinect. *Computer-Aided Design and Applications* 10, 4, 663-669. http://dx.doi.org/10.3722/cadaps.2013.663-669

[7] J. Huang and R. Rai. Year. Hand Gesture Based Intuitive Cad Interface. In *Proceedings of ASME 2014 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. http://dx.doi.org/10.1115/DETC2014-34070

[8] W. Hürst and C. Van Wezel. 2013. Gesture-Based Interaction Via Finger Tracking for Mobile Augmented Reality. *Multimedia Tools and Applications* 62, 1, 233–258. http://dx.doi.org/10.1007/s11042-011-0983-y

[9] J. Kang, K. Zhong, S. Qin, H. Wang and D. Wright. 2013. Instant 3d Design Concept Generation and Visualization by Real-Time Hand Gesture Recognition. *Computers in Industry* 64, 7, 785-797 http://dx.doi.org/10.1016/j.compind.2013.04.012

[10] S. Khan and B. Tuncer. 2017. Intuitive and Effective Gestures for Conceptual Architectural Design. In *Proceedings of ACADIA 2017 conference: Disciplines, Disruption*. Massachusetts Institute of Technology, Cambridge, MA, USA.

[11] S. Khan and B. Tuncer. (Forthcoming). Gesture and Speech Based 3d Cad Modeling for Conceptual Architectural Design

[12] C. Kühnel, T. Westermann, F. Hemmert, S. Kratz, A. Müller and S. Möller. 2011. I'm Home: Defining and Evaluating a Gesture Set for Smart-Home Control. *International Journal of Human-Computer Studies* 69, 11, 693-704. http://dx.doi.org/10.1016/j.ijhcs.2011.04.005

[13] A. Malizia and A. Bellucci. 2012. The Artificiality of Natural User Interfaces. Commun. ACM 55, 3 (March 2012), 36-38. http://dx.doi.org/10.1145/2093548.2093563

[14] V. G. Nanjundaswamy, A. Kulkarni, Z. Chen, P. Jaiswal, Sree Shankar S., A. Verma and R. Rai. 2013. Intuitive 3d Computer-Aided Design (Cad) System with Multimodal Interfaces. In *Proceedings of ASME 2013 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. http://dx.doi.org/10.1115/DETC2013-12277

[15] F. Naya, M. Contero, N. Aleixos, J. A. Jorge and P. Company. 2008. Sketch-Based Interfaces for Parametric Modelling. In *Human Computer Interaction*, I. Pavlidis Ed. InTech. http://dx.doi.org/10.5772/6310

[16] S. Oviatt. 1999. Ten Myths of Multimodal Interaction. *Commun. ACM* 42, 11, 74-81. http://dx.doi.org/10.1145/319382.319398

[17] H. T. Regenbrecht, M. Wagner and G. Baratoff. 2002. Magicmeeting: A Collaborative Tangible Augmented Reality System. *Virtual Reality* 6, 3, 151–166. http://dx.doi.org/10.1007/s100550200016

[18] B. Tuncer and S. Khan. 2017. User Defined Conceptual Modeling Gestures. 2nd Cultural DNA Workshop notes. Korea Advanced Institute of Science and Technology, Seoul, South Korea

[19] R. Wang, S. Paris and J. Popović. 2011. 6d Hands: Markerless Hand-Tracking for Computer Aided Design. In *Proceedings of 24th annual ACM symposium on User interface software and technology*. ACM, New York, NY, USA, 549-557. http://dx.doi.org/10.1145/2047196.2047269

[20] A. Withana, R. Peiris, N. Samarasekara and S. Nanayakkara. 2015. Zsense: Enabling Shallow Depth Gesture Recognition for Greater Input Expressivity on Smart Wearables. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 3661–3670. http://dx.doi.org/10.1145/2702123.2702371

[21] J. O. Wobbrock, M. R. Morris and A. D. Wilson. 2009. User-Defined Gestures for Surface Computing. In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1083-1092. http://dx.doi.org/10.1145/1518701.1518866

[22] Y. Yin and R. Davis. 2014. Real-Time Continuous Gesture Recognition for Natural Human-Computer Interaction. In *Proceedings of 2014 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. IEEE, Melbourne, VIC, Australia. http://dx.doi.org/10.1109/VLHCC.2014.6883032

[23] B. Zamborlin, F. Bevilacqua, M. Gillies and M. D'inverno. 2014. Fluid Gesture Interaction Design: Applications of Continuous Recognition for the Design of Modern Gestural Interfaces. *ACM Trans. Interact. Intell. Syst.* 3, 4, 30. http://dx.doi.org/10.1145/2543921