

Put That There: 20 Years of Research on Multimodal Interaction

James L. Crowley
Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LIG
F-38000 Grenoble, France
James.Crowley@inria.fr

ABSTRACT

Humans interact with the world using five major senses: sight, hearing, touch, smell, and taste. Almost all interaction with the environment is naturally multimodal, as audio, tactile or paralinguistic cues provide confirmation for physical actions and spoken language interaction. Multimodal interaction seeks to fully exploit these parallel channels for perception and action to provide robust, natural interaction.

Richard Bolt's "Put That There" (1980) provided an early paradigm that demonstrated the power of multimodality and helped attract researchers from a variety of disciplines to study a new approach for post-WIMP computing that moves beyond desktop graphical user interfaces (GUI).

In this talk, I will look back to the origins of the scientific community of multimodal interaction, and review some of the more salient results that have emerged over the last 20 years, including results in machine perception, system architectures, visualization, and computer to human communications.

Recently, a number of game-changing technologies such as deep learning, cloud computing, and planetary scale data collection have emerged to provide robust solutions to historically hard problems. As a result, scientific understanding of multimodal interaction has taken on new relevance as construction of practical systems has become feasible. I will discuss the impact of these new technologies and the opportunities and challenges that they raise. I will conclude with a discussion of the importance of convergence with cognitive science and cognitive systems to provide foundations for intelligent, human-centered interactive systems that learn and fully understand humans and human-to-human social interaction, in order to provide services that surpass the abilities of the most intelligent human servants.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

ICMI'18, October 16-20, 2018, Boulder, CO, USA.

© 2018 Copyright is held by the owner/author(s).

ACM ISBN 978-1-4503-5692-3/18/10.

DOI: <https://doi.org/10.1145/3242969.3276309>

CCS Concepts/ACM Classifiers

Human computer interaction (HCI), Multimodal Interaction

Author Keywords

Multi-modal Interaction, Perception, Action, Cognition



BIOGRAPHY

James L. Crowley is a Professor at the Univ. Grenoble Alpes, where he teaches courses in Computer Vision, Machine Learning and Artificial Intelligence at Grenoble Polytechnique Institut (Grenoble INP). He directs the Pervasive Interaction research group at INRIA Grenoble Rhône-Alpes Research Center in Montbonnot, France.

Over the last 35 years, professor Crowley has made a number of fundamental contributions to computer vision, robotics and multi-modal interaction. These include early innovations in scale invariant computer vision, localization and mapping for mobile robots, appearance-based techniques for computer vision, and visual perception for human-computer interaction.

Current research concerns context aware observation of human activity, Ambient Intelligence, and new forms of Human-Computer Interaction based on machine perception.

REFERENCES

1. Bolt, R.A., Put-that-there, *Voice and gesture at the graphics interface*. Vol. 14, No. 3, pp. 262-270, ACM, 1980.
2. Turk, M., Multimodal interaction: A review. *Pattern Recognition Letters*, 36, pp.189-195, 2014.
3. Oviatt, S., Schuller, B., Cohen, P., Sonntag, D. and Potamianos, G., *The Handbook of Multimodal-Multisensor Interfaces: Foundations, User Modeling, and Common Modality Combinations*. Morgan & Claypool, 2017.