The Bridge Project

Jake Beal, Nick Caldwell, Jimmy Lin, Justin Schmidt, Marc Spraragen & Patrick Winston

Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

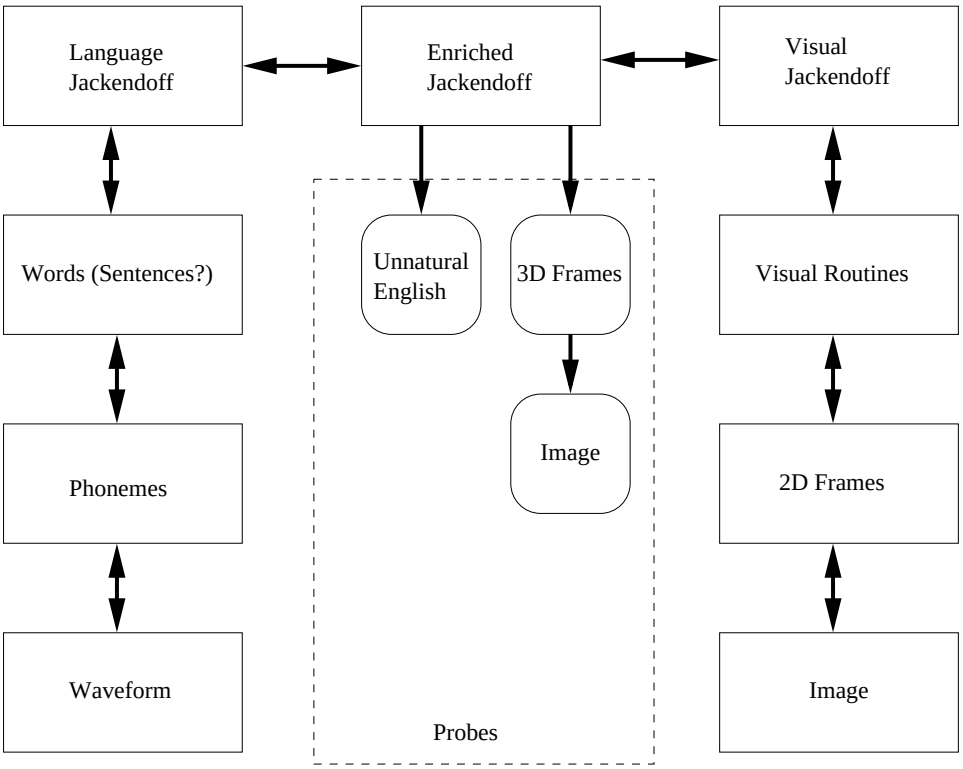

Figure 1: The Bridge architecture: boxes are buffers holding representations of the external world at various levels of abstraction, arrows are constraints connecting adjacent buffers. Probes may be attached to the architecture to provide insight into its operations.

The Problem: We need better perception systems, both to support applications and to support research on intelligence. We believe that each subsystem of a better perception system will rely on contextual information from other perception subsystems and on knowledge of the evolving situation in the external world. Accordingly, we are building a system, the Bridge System in which simple machine vision and natural language subsystems work together, sharing information, asking each other questions, and ultimately, solving problems that machine vision and natural language systems, working alone, cannot.

Motivation: We humans are amazingly good at interpreting sensory data. Images we see or words we hear are often noisy, ambiguous, and context dependent. In fact, there is much evidence that our human perception of the world is partly hallucinated, based on our expectations. For example, when we say “gas stove” most people produce one ’s’ but hear two; similarly, when we look at a colored lump on a chair, we understand it as a shirt even though it has practically no distinguishing features.

We theorize that we humans handle such situations because our perception subsystems depend on context, both from other senses and from the evolving situation. If this is true, then primary sensory processing need not be very good—its job is not to determine what is seen or heard, but a range of what *might* be, to be disambiguated by contextual information.

The Bridge System is to test the theory: if the theory is correct, then the Bridge System should produce much better performance on vision and language tasks when the two halves are connected than when they are separate.

Previous Work: Ullman's work on vision systems and bidirectional search is an important source of inspiration for this project. In this work, Ullman observed the bidirectional structure of the visual cortex and suggested an algorithm by which incoming visual data searching upward and lexicon images searching downward could simplify object recognition tasks.[3]

More directly, one of the foundations of the Bridge project is unpublished work by Marc Spraragen. Spraragen's system takes a set of simple sentences expressing that various people walk and push, birds walk and fly, and balls roll; then, Spraragen's system generates a causal interpretation connecting the events, producing a slide show rendering its interpretation into the visual world.

Approach: The architecture in which we are implementing the Bridge System is a system of constraints connecting data buffers. Each buffer contains a complete representation of the world at some level of abstraction. At the lowest level, this is direct sensory data: audio waveforms on the language side, a pixel map on the vision side. Interpretations ascend through complexity until the highest-level semantic interpretation, which actually bridges between the two senses.

The actual processing work in the system is done by bidirectional constraints operating between adjacent buffers. For example, between the phoneme-level and sentence-level buffers on the language side of the bridge structure, a constraint interprets clusters of phonemes as words, and likewise turns words in the sentence buffer into sequences of phonemes.

The semantic structure used at the highest level is a version of Jackendoff's LCS frame structure,[2] enhanced with the addition of concepts from Borchardt's transition space.[1]. This high-level structure can be interpreted into sensory information for language and vision with equal ease.

Impact: If successful, the Bridge project will provide an important new approach to machine perception. Most important is a paradigm shift from "What does this data mean?" to "What does this data let me rule out?", explicitly involving multiple parts of the system in what otherwise would be a local sensor interpretation task.

Future Work: Further investigation will involve connecting more senses to the system, as well as memory and knowledge database capabilities.

References:

[1] Gary C. Borchardt. Causal reconstruction. Technical Report 1403, MIT Artificial Intelligence Laboratory, 1993.

[2] Ray Jackendoff. *Semantics and Cognition*. MIT Press, 1983.

[3] Shimon Ullman. *High Level Vision: Object Recognition and Visual Cognition*. MIT Press, 1996.