CS260, Spring 09
Nicholas Kong
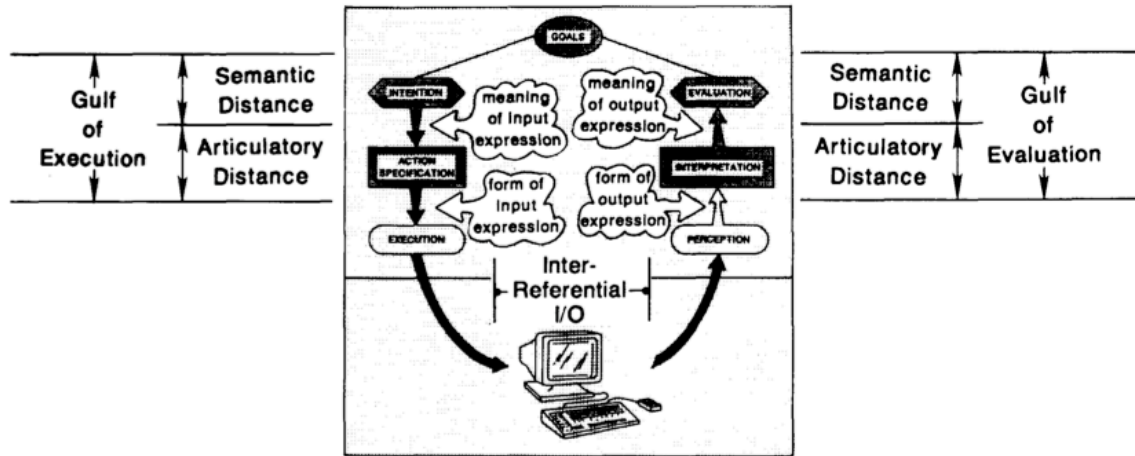
Notes on the gulfs of execution and evaluation

Figure 6 of the paper (shown above) is important and provides a pretty good distillation of the concepts. The paper can be a little confusing because some examples blur the lines between semantic and articulatory distances. I think the key contribution of the paper is to seed thought about what "intuitive" or "obvious" means in an interface. The exact pigeonholing of certain actions into their little bins is not as important.

Definitions

First, let's look at some quotes from the paper:

> The **gulf of execution** is bridged by making the *commands and mechanisms* of the system match the *thoughts and goals* of the user.

In other words, to bridge the gulf of execution, you must translate your ideas or goals into the language of the input.

> The **gulf of evaluation** is bridged by making the output displays present a good conceptual model of the system that is readily perceived, interpreted, and evaluated.

To bridge the gulf of evaluation, you must translate the system's output language into your own internal language.
Both of these gulfs are decomposed into two types of distances:

> **Semantic distance** reflects the relationship between the *user intentions* and the *meaning of expressions* in the interface languages.

There are two interface "languages", one for input and one for output. The input language is comprised of the possible actions and commands; the output language is

comprised of the feedback of the system. To bridge semantic distance, we must therefore provide interface commands which require little translation from the goals of the user. The paper also emphasizes that there is a difference between "automated behaviour", that is, memorization of the system's interface vocabulary, and semantic distance. In the first case, the user has adapted to the system to such an extent that their intents are already formed in terms of the system vocabulary, where in the second the system's vocabulary has been designed to match the user's intent. For example, we are all familiar with the cut-and-paste operation. So instead of thinking "I want to move this paragraph to the beginning of the document.", we may think "I will cut this paragraph and paste it at the beginning.", or at least it becomes immediately apparent what operation we need to use. We've adapted to the system's vocabulary. Note that most word processors do allow users to do a manual click-and-drag of a paragraph (once it's been selected); I would argue the semantic distance is smaller in this case.

> **Articulatory distance** reflects the relationship between the *physical form* of an expression in the interaction language and its *meaning*.

By "physical form" we mean the actual actions required to invoke an "expression" in the interaction language. For example, say we have a "Dismiss all" button in an error dialog. The meaning of this expression in the interaction language is to ignore all further errors. The physical form of this expression is the requirement of the user to move the mouse cursor to the button and click.
On the gulf of evaluation side, the articulatory distance is the translation between the physical form (e.g., string, picture, sound, etc.) into the meaning of that expression *in terms of the output language*. For example, take the example of a mail notifier that changes color when you have new mail. The articulatory distance is not the time taken for the computer rendering that new icon (unlike what I said in some sections, sorry!), but rather the effort required of the user to determine what that change means. An icon change in a mail notifier is a good example of a shortened articulatory distance in the gulf of evaluation.
Some more examples will help here, and I'll try and address the specific questions that arose in section. These definitions are also rather vague, in that the definition of a user's "idea" or "goal" is unclear. The goal could be something as general as "write an e-mail" to something as specific as "shoot zombie".

Examples

ESP game

Gulf of execution

The goal of the player is to match tags of an image with their partner. Let's see if we can split this gulf up into its component distances.

- **Semantic distance.** The player's idea/goal is to guess a tag for an image. The semantic distance is composed of the "translation" of the player's goal into the interface's language. So what's allowed in this interface? We have a textbox in which we can input a tag, so in this case the semantic distance is small. All we must do to guess a tag is to input the tag itself.

- **Articulatory distance.** There are at least two physical forms of the actions we have to underatke in the system, namely the input of the tag into the textbox and the confirmation of that guess. The user must type the tag, but can then either press "Enter" or move the mouse cursor and click "Submit". The two forms have different articulatory distances.

Gulf of evaluation

In the ESP game, the articulatory and semantic distances of the gulf of evaluation are somewhat vague and perhaps conflated. The player wishes to know whether they've successfully matched a tag with their partner. This is achieved in the form of a textbox that notifies the user on what word the user and their partner matched on. So the articulatory distance is very small, in that the string "You matched on jet" is self-revealing of the system's expression "matched words". In addition, the semantic distance is also small, as the goal of the player is to match words, and the feedback string immediately provides this information to the user.
The articulatory distance might be somewhat increased if instead of a pop-up box, the word that the players matched on was *highlighted*. In this case, the player would first have to translate the output of a highlighted word into the system expression "matched word", which might be less clear than the current text feedback.

First person shooter - increasing the articulatory distance in the gulf of execution

In follow-up to the example given in one of the Wednesday sections, one way to increase the articulatory distance in the gulf of execution for, say, Halo, is to provide an RPG-style menu-selection interface for firing a gun. That is, the player would have to select the "Action" menu, then the "Fire" menu, then the "Gun" menu, etc. This increases articulatory distance because the relationship between the physical form of the expression with meaning "fire a gun" is not at all clear.

Minimizing semantic distance in the gulf of execution

The question here is how we can go about minimizing semantic distance in the gulf of execution. This is possible by designing an interface whose vocabulary or language matches that of the goals/ideas of the user. Let's say we want to animate a character walking from point A to point B. In one interface, we may have to manually specify each keyframe, all the way from point A to point B. The semantic distance is quite large: we need to figure out how the actions of walking transfer to the smaller motions of a character, as the vocabulary of the system only allows for changes to individual frames. We could potentially reduce the semantic distance by simply providing a "Walk" button. The downside would be a loss of fine-grained control of the movement of the character, but there would be virtually no translation from the user's idea of "make this character walk" to the language of the interface.