# Graphically Speaking:
## Do Graphics Affect Perspectives in Event Conceptualization?[1]

**Ichiro Umata** (umata@atr.co.jp)
ATR Media Information Science Laboratories;
Seika Soraku Kyoto, 619-0288 Japan
**Yasuhiro Katagiri** (katagiri@atr.co.jp)
ATR Media Information Science Laboratories;
Seika Soraku Kyoto, 619-0288 Japan
**Atsushi Shimojima** (ashimoji@jaist.ac.jp)
Japan Advanced Institute of Science and Technology;
1–1 Asahi Tatsunokuchi Nomi Ishikawa 923-1292 Japan

**Abstract**

To see is to believe. A picture is often worth a thousand words in everyday communication settings. A graphical representation actually "talks" in such communication, integrated with some other representation systems like spoken language. Because of its power, however, a graphical representation can affect the way people grasp a target situation it describes. This paper presents an empirical investigation of language usage in graphical communication. Drawing on actual dialogue data, we show that the configuration of graphics affects linguistic expressions of motion when people collaboratively work on a task. This effect of graphics on language usage demonstrates that the configuration of graphics has an influence on perspectives in event conceptualizations.

## Introduction

Daily communication is by nature multi-modal, and graphics often serve as strong visual aids for information exchange. People communicate with each other effectively by integrating information from linguistic and graphical sources (Neilson and Lee (1994); Umata, Shimojima, and Katagiri (2000)). People grasp described situations via graphics, taking advantage of their "handiness." Because of this role they play, however, the way people grasp target situations can be affected by graphics.

Recent investigations have demonstrated that the perspectives of spatial descriptions and linguistic expressions show some correspondence. Levinson (1996) observed that some languages make almost exclusive use of absolute coordinates while European languages tend to use egocentric or relative coordinates, reflecting people's strategies for spatial memory and inference. Taylor and Tversky showed that there are three perspectives of spatial descriptions (i.e. *a gaze tour/a route/a survey*) that roughly correspond to the frames of reference distinguished by Levinson in a linguistic method.

It has also been observed that the existence of graphics provides two graphics-based perspectives in addition to those used in only describing a world, and that there is a trade-off between cognitive costs and alignment-failure-robustness (Umata, Katagiri, and Shimojima (2002)). One of those perspectives is called the Observer-to-Graphics Perspective, in which people conceptualize the target events from the viewpoint of the observer relative to the graphics. Suppose one says: "From Baker Street, we'll travel on the Bakerloo Line and then go right at Oxford Circus to Holborn" while holding the map of the London Underground shown in Figure 1. The movement is described from the viewpoint of the observer relative to the map. On the other hand, if one says: "From Baker Street, we'll travel on the Bakerloo Line and then go left at Oxford Circus to Holborn," the speaker "goes into" the map world as an imaginary agent, taking the other perspective called the Protagonist Perspective.

When people see and talk about world situations through graphics, it is quite likely that the features of the graphics affect the way people conceptualize the target situations. In this research, we studied how the availability and configuration of graphics affect language usage in communication and problem-solving. We focused on the influence of graphical representations on the conceptualization of motion events.

Suppose that John and Mary are at the Goodge Street tube station, discussing where to have dinner together. Mary might suggest a place by saying (1) below, but she would not do it by saying (2):

(1) Let's go down to Waterloo Station on the Northern Line and eat at Livebait.

(2) Let's come down to Waterloo Station on the Northern Line and eat at Livebait.

The current position where the two people are located becomes the reference point of the movement in this case, and the movement can only be conceptualized as a movement away from the reference point, and hence the use of "go." Suppose, on the other hand, that John and Mary are discussing their evening plans over the map of the London Underground shown in Figure 1. Mary could use, in this case, either (1) or (2). The availability of the map and the configuration of icons on the map affect the conceptualization of the movement here: the nearness of the Waterloo Station icon from them makes it possible for her to conceptualize the movement, in addition to the previous distal movement conceptualization, as a movement *in the map-world* toward the reference point, their current position. Graphical representation can have an influence on language usage.

The use of "come" in (2) is possible because the map and the graphical objects contained in it are readily avail-

[1]This research was supported in part by the Telecommunications Advancement Organization of Japan.

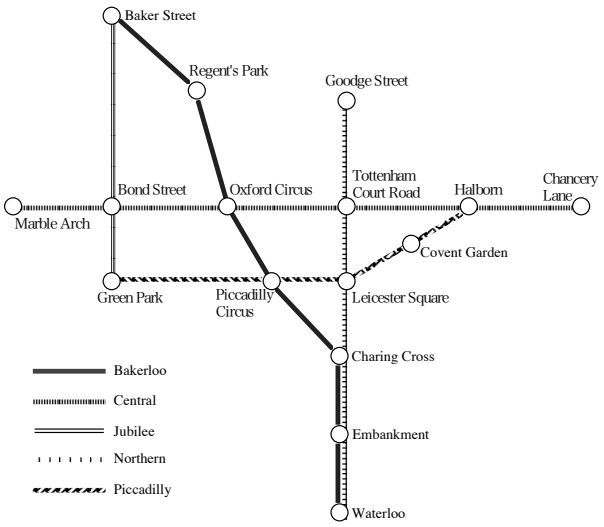


Figure 1: Route Map of London's Underground System

able to the speaker as a resource to formulate messages to be communicated and problems to be reasoned about. The locations and arrangements of objects can be expressed in terms of the relationships between graphical objects and the speaker, as well as those between objects themselves and the speaker. This availability, or the ease of accessibility, of graphical representations should work to amplify our communicative and reasoning capabilities by providing us with a novel set of possibilities for constructing perspectival event conceptualizations.

We investigated the effect of graphical representations on perspectival event conceptualizations through an empirical analysis of the use of motion verbs in actual two-party task-oriented dialogues that make use of diagrams. We first outline, in the next section, the classification of the types of perspectival conceptualizations available in communications that involve graphical representations. We then examine, in the following section, the Japanese dialogue data from our corpus involving a "Missionaries and Cannibals" type puzzle. We found that the configuration of graphics affects the way people grasp the target situations and that the perspectival conceptualizations of the graphical objects are more prominent than those of the real-world objects they represent.

## Perspectives in Graphical Communication

In graphical communication, graphical representations work as "windows" through which we can see the target situations the graphics describe. However, they also serve as information processing "sites" because we can take advantage of their handiness. Graphical representations are so deeply ingrained in how we grasp and describe target situations that their existence raises the possibility of setting novel perspectives in conceptualizing events. Two such perspectives were observed in Umata, Katagiri and Shimojima (2002), in addition to two perspectives concerned solely with the target world. There are four possible categories of perspectives on motion events in graphical communication.

(a) Observer-to-World Perspective

> A movement is taken as a movement in the real world and conceptualized from the viewpoint of the observer within the real world. This perspective is concerned solely with the real world.

(b) Agent Perspective

> A movement is taken as a movement in the real world and conceptualized from the viewpoint of the agent of motion. This perspective is concerned solely with the real world.

(c) Observer-to-Graphic Perspective

> A movement is taken as a movement in the map space and conceptualized from the viewpoint of the observer relative to the map. This perspective concerns both the real world and the graphic space and creates a bridge between the two.

(d) Protagonist Perspective

> A movement is conceptualized from the viewpoint of an imaginary agent in a narrative world. In graphical communication situations, a graphic provides the narrative domain for this perspective. The agent can be identified with either the speaker or the listener. This perspective belongs solely to the narrative world.

The first two categories are perspectives concerned solely with the target world, which can also be observed in communication without the use of graphics. When you say "John is coming to my place from Goodge Street" without a map, you are taking the Observer-to-World Perspective. If you are actually driving to somebody's place, you might say "I'm now going to the right-hand side of Piccadilly Circus," taking the Agent Perspective. If you are explaining the way to somebody via a cellular phone, you might say, "Go south and turn left at Leicester Square," taking the Agent Perspective of the person to whom you are talking.

In graphical communication, the latter two perspectives are available in addition to (a) and (b). Examples from the HCRC Map Task Corpus analyzed in Umata, Katagiri and Shimojima (2002) are shown below.

### HCRC Map Task Corpus

The examples shown in the following sections are from the HCRC Map Task Corpus. This map task is a cooperative one involving two participants. The two speakers sit opposite one another, and one speaker gives instructions for a route to the other. Each has a map that the other cannot see, and a route is marked on the Instruction Giver's map while no route is marked on that of the Instruction Follower. The speakers are told that their goal is to reproduce the Giver's route on the Follower's map. Their maps are not exactly identical and the speakers are

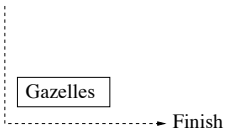


Figure 2: Movement described in (3)

told this explicitly at the beginning of their first session. It is, however, up to them to discover how the two maps differ. The maps describe fictitious areas.

### Observer-to-Graphic Perspective

First, consider the following utterances:

(3) (The Giver is showing the Follower the movement shown in Figure 2.)

Giver: Okay? Now you need to drop straight down towards the gazelles.
Follower: Right, coming in at the top of them.
Giver: That's right, and then go round the bottom of the gazelles.
Follower: On the left-hand side?
Giver: And head off to the right-hand side ... So you go under the gazelles

The expression "drop straight down to" would not have been suitable without a map. It describes the motion from the Observer-to-Graphic perspective, making use of the spatial relation on the map. Other spatial expressions ("at the top of," "bottom of," "right-hand side," and "under") also describe the spatial relation of the target situation via the graphical relations. For example, "go round the bottom of the gazelles" does not mean actually going under the gazelles in the target situation. Also, notice that the deictic spatial expression "to the right of" is based on the Observer-to-Graphic Perspective. It would have been "to the left-hand side" if the giver was taking the Protagonist Perspective.

### Protagonist Perspective

Now we will examine the examples of the Protagonist Perspective shown in (4):

(4) (The Giver is showing the Follower the movement shown in Figure 3.)

Giver: Then down.
Follower: What do you mean down? Towards the bottom of the paper?
Giver: Uh-huh
Follower: Uh-huh. Do I ... do I go by the collapsed shelter?
Giver: Uh-huh
Follower: Uh-huh
Giver: And then ... so that ... until you've got ...
Follower: The collapsed shelter's on my right?
Giver: Uh-huh

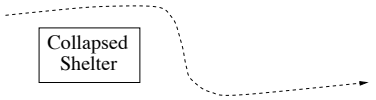


Figure 3: Movement described in (4)

Follower: Right
Giver: And then go round to your left
Follower: My left?
Giver: As you're the wee guy

The spatial expressions "on my right," "to your left" and "my left" are clearly based on the Protagonist Perspective. If the speakers were talking based on the Observer-to-Graphic Perspective, they would have said "on my left," "to your right" and "my right," respectively. The giver confirmed it with the utterance "as you're the wee guy," introducing an imaginary agent explicitly. Thus, we can find many such expressions spoken from the two graphic-based perspectives in communications that make use of graphics.

### Perspectives and Alignment of Coordinates

We have observed that the existence of graphics provides two novel perspectives, namely the Observer-to-Graphics Perspective and the Protagonist Perspective. The former requires less cognitive resources because one can grasp motion events as seen on graphics. This perspective works if the coordinates are firmly aligned between speakers. However, misalignment of coordinates leads to serious miscommunication. On the other hand, the Protagonist Perspective is robust against misalignment, because conversation participants "go into the graphics," and the spatial relations between the actual speakers and their graphics are not crucial in this perspective. Such a "trade-off" between description cost and misalignment robustness was actually observed in the HCRC Map Task Corpus (Umata, Katagiri and Shimojima (2002)).

## Dialogues Involving a "Missionaries and Cannibals" Type Puzzle

We have observed that the availability of graphics provides a novel set of perspectives based on graphics. Because people "see" the target situations via graphics from those perspectives, it is quite likely that the features of the graphics affect the way people conceptualize the target situations.

The dialogue data analyzed in this section were taken from two collaborative problem solving experiments that involved back-and-forth movement. Another important feature of this task was that it involved two real-world places that subjects were familiar with. We also examined how much effect the Observer-to-World Perspective had on the usage of motion verbs.

## Motion Verbs in Corpus

Verbs like *come* and *go* reflect a speaker's reference point, as shown in (1) and (2). *Go* indicates motion to a location that is distinct from the reference point. *Come* indicates motion toward the reference point[2].

The Japanese language also has a pair of motion verbs similar to English *come* and *go*: *kuru* and *iku*. *iku* (go) and *tsurete-iku* (take) indicate motion to a location that is distinct from the reference point. *kuru* (come) and *tsurete-kuru* (bring) indicate motion toward the reference point[3]. *Iku*-type verbs are used more widely than *kuru*-type verbs in the sense that the former expresses movements that are neutral with respect to the reference point locations. There are several verbs that can be classified into these two classes. We examined the usage of these two classes of verbs in the following two experiments.

## Data

The data analyzed here was gathered from experiments involving problem solving. In this task, two subjects collaboratively worked on a "Missionaries and Cannibals" type puzzles using a diagram given to them. The structure of the puzzle was basically the same as the original one, except that it involved two actual places that the subjects were familiar with. The time limit was seven minutes.

The subjects recruited from local universities were seated in separate, soundproof rooms and worked together as a pair using a shared virtual whiteboard and a full duplex audio connection. The diagram was shown on their whiteboards, and the subjects could draw and erase freely except that they could not erase the original diagram. All inputs to the screen were by stylus, and any writing or erasing by one participant would appear simultaneously on the partner's screen. A pointing action with the stylus was shown by a cursor on the screen, and the subjects could see what their partner was pointing to. The subjects were video-taped during the task.

## The Motorcycle Gang Task

The puzzle was almost the same as the original one, except that we used two actual places and replaced the missionaries and cannibals with two teams of motorcycle gangs. The subjects were told to work out how all members of both gangs could be transported safely. This task involved only two kinds of motion: forward and backward motion between two places. The time limit was seven minutes, including the time they used to read the problem sheet.

[2]Actually, *come* and *go* have more complicated semantics, as shown in Fillmore (1997). The scheme presented here is a rather simplified version, but serves well enough for the present purpose.

[3]There is one clear difference between English and Japanese, though. When a speaker is trying to go to the hearer, s/he will say, "I'll come to you," while *iku* (go) is used rather than *kuru* (come) in Japanese. However, this difference is not relevant here.

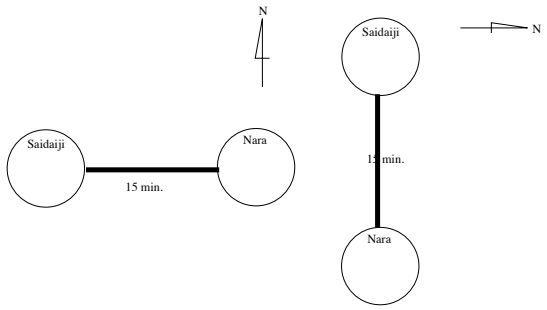


Figure 4: Horizontal and Vertical Diagrams

## Experiment 1

The first experiment was conducted to examine the effect of the configurational features of graphics when a perspective based on graphics is taken. The motion in this task was much more simplified, though there was back-and-forth motion that was not in the HCRC Map Task. The problem involved motion between two actual places on a motorbike so that the subject could also directly access the real world. This task has a general direction of motion: all six boys have to move from *Saidaiji* to *Nara*. Each of these locations is about the same distance from the experiment site. The bike was supposed to be able to carry only two people at one time, and someone had to ride back on it. Two kinds of graphical representations were provided as shown in Figure 4. One had a horizontal configuration, in which the two icons of the places are at about the same distance from the subject[4]. The other one had a vertical configuration, which had a variation in the distance from the subject to each place. Each condition had eight pairs of subjects.

The speakers were not supposed to be the ones moving between the two places in this task setting, so the Agent Perspective was not possible. Thus, the possible perspectives were expected to be the Observer-to-World, the Observer-to-Graphic and the Protagonist Perspective. Because the motion is taken as that in the map from the real-world observer in the Observer-to-World Perspective, it is likely that the spatial relation between the speaker and the graphical objects plays an important role. The assumption was that, under that perspective, *kuru(come)*-type verbs would be used for the motion to Nara more frequently in the vertical condition than in the horizontal one, because they would be affected by the nearness of the Nara icon in the diagram to the speaker.

**Results of Experiment 1** The distribution of motion verbs was as follows:

*Iku*-type verbs and *kuru*-type verbs exhibited significantly different distributions depending on the direction of movement in both conditions (horizontal: $\chi^2{}_{(1)} = 93.64, p < 0.01$; vertical ($\chi^2{}_{(1)} = 53.98, p < 0.01$). In both the horizontal and vertical conditions, *iku*-type

[4]Note that it is common for maps in Japan to have a direction other than north at the top.

Table 1: Distribution of *iku*-type and *kuru*-type verbs.

| | | *iku*-type | *kuru*-type |
|---|---|---|---|
| Horizontal | Saidaiji → Nara | 66 | 2 |
| | Nara → Saidaiji | 4 | 45 |
| Vertical | Saidaiji → Nara | 47 | 18 |
| | Nara → Saidaiji | 1 | 45 |

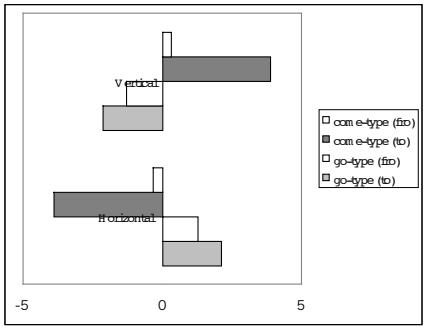


Figure 5: Difference in distribution of motion verbs

verbs were predominantly employed for the movement from Saidaiji to Nara, whereas *kuru*-type verbs were predominantly employed from Nara to Saidaiji. This distributional imbalance indicates that our four perspective types are not sufficient to explain the whole story of graphical communication in problem solving settings, since, for the horizontal condition, movement in each direction could be considered neither toward nor away from the reference point under all four perspectives. Hence one would predict that all movement in the horizontal condition should be expressed with *iku*-type verbs.

We therefore stipulated an additional perspective, the Problem Perspective, to account for this distributional imbalance. As the task for "Missionaries and Cannibals" type problems is to transport all parties collectively from one location to another, the nature of the problem itself induces a general direction of movement from the source location to the goal location. This is equivalent to positing the reference point at the source location, which is Saidaiji in our experimental setting. Under the Problem Perspective each movement is taken as a movement either toward or away from the reference point as posited by the structure of the problem itself, and this perspective is expected to appear in transportation type problems in both concrete and abstract domains.

Comparing the horizontal and the vertical conditions, we notice that *kuru*-type verbs are more frequently used for the movements from Saidaiji to Nara. Figure 5 shows the difference in the distribution of motion verbs between the two conditions. The motion verbs exhibit significantly different distributions between the horizontal and the vertical conditions ($\chi^2_{(3)} = 17.65, p < 0.01$). More concretely, the frequency of the *kuru*-type in Saidaiji to Nara motion is significantly larger in the vertical condition (adjusted residual: horizontal = -3.87, vertical = 3.87).

The Problem Perspective sets a general reference point at Saidaiji, the source of the whole transportation process. It appears, however, that the graphics configuration may be able to modify the reference point settings. The *kuru*-type showed higher frequency for Saidaiji to Nara in the vertical condition than in the horizontal condition. This shows that the spatial relation between the speaker and the graphical objects makes the transition to the Observer-to-Graphic Perspective, and this affects the reference point setting accordingly. The handiness of the graphical representation can cause a switch in perspectives and thus a shift in the reference point.

In contrast to the increase in *kuru*-type verbs in the vertical condition, we notice no increase in *iku*-type verbs for the movements in the opposite direction. The low frequency of the *iku*-type for Nara to Saidaiji suggests that the perspective switch from the Problem Perspective to the Observer-to-Graphic Perspective is preferred when the resulting perspective takes the movement as a toward movement rather than as an away-from movement. This asymmetry also suggests that the Protagonist Perspective was not playing a significant role, as switching to it would have increased the occurrence of *iku*-type verbs here.

Thus, it was shown that the effect of the Problem Perspective defined by the task was the most prominent factor in reference point setting, but that the configuration of a graphical representation often affects this reference point setting.

## Experiment 2

The previous experiment showed that the Problem Perspective was the most influential factor, while the configuration of graphics also affects the usage of motion verbs. Now we will look into the effect of the real-world configuration in conversations involving graphics. The setup of Experiment 2 differs from that of our previous experiment as follows:

- Both of the two conditions had a vertical configuration.
- One of the locations was the current position of the subject (ATR), and the other was a place some distance away.
- One condition was consistent with the physical world, while the other condition was inconsistent with the physical world.

The difference was that one diagram had a configuration consistent with the real-world relationship, while the other did not; that is, the nearer icon in the graphics represented a farther place in the real world. The general starting point was placed at the top of both diagrams. These diagrams are shown in Figure 6.

If the real-world configuration has some effect on setting the reference point, the motion verbs would show

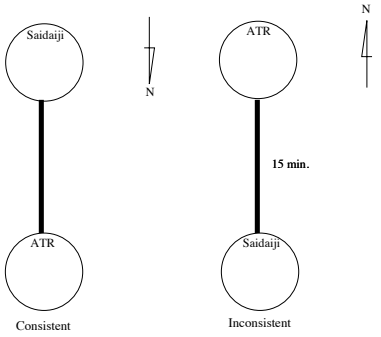


Figure 6: Consistent and inconsistent diagrams

different distributions between the consistent condition and the inconsistent condition. The frequency of *kuru*-type verbs in start-to-goal motion is expected to be lower in the inconsistent condition than in the consistent one. If the real-world configuration did not have much effect, then the distribution would be almost the same between these two conditions. Each condition had eight pairs of subjects.

**Results of Experiment 2** The distribution of motion verbs was as shown in Table 2.

Table 2: Distribution of *iku*-type and *kuru*-type verbs in Exoeriment 2.

| | | *iku*-type | *kuru*-type |
|---|---|---|---|
| Consistent | Start → Goal | 70 | 17 |
| | Goal → Start | 1 | 48 |
| Inconsistent | Start → Goal | 47 | 25 |
| | Goal → Start | 4 | 42 |

No significant difference was observed between the consistent and inconsistent conditions ($\chi^2_{(3)} = 7.00 < 7.81, p = 0.05$). Furthermore, a comparison of each condition with the vertical condition in Experiment 1 also did not show any significant difference (consistent vs. vertical: $\chi^2_{(3)} = 2.14 < 7.81, p = 0.05$; inconsistent vs. vertical: $\chi^2_{(3)} = 2.83 < 7.81, p = 0.05$). The distribution in each condition was similar to the distribution in the vertical condition of Experiment 1. The *kuru*-type was observed in the Start-to-Goal motion in both conditions. The frequency of the *iku*-type for the Nara to Saidaiji motion was again quite low.

The results show that spatial consistency did not contribute to shifting the reference point. The Observer-to-World Perspective did not have a strong influence in conversation with a diagram. The Protagonist Perspective was also weak in this setting.

It turned out that the spatial property of graphics had a stronger effect on event conceptualization than that of its target world. This suggests that the Observer-to-Graphic Perspective is stronger than the Observer-to-World Perspective in graphical communication settings.

## Conclusion

We have analyzed the effect of graphics on language usage in communication. Based on an empirical analysis of the uses of movement verbs in actual conversational data, we have shown that the configuration of a graphical representation affects the reference point setting when people conceptualize motion events from graphics-based perspectives. We found that: (1) the task settings provided yet another kind of perspective, the Problem Perspective, setting a general reference point to the general origin; (2) the Problem Perspective is stronger than graphics-based perspectives; (3) the Observer-to-Map Perspective is the next strongest; and (4) the real-world perspective does not contribute, in comparison with the graphics-based perspective, to the reference point shift.

These results suggest that we are mainly grasping an event of the target world via its representation, rather than from the event itself, in graphical communication situations. The point of using graphical representations is the convenience and the ease of access they give us, which helps us to grasp an event through the mediation of graphics. This mediation makes it possible to talk about distal objects by manipulating their proximal counterparts, thereby facilitating both communication and reasoning processes. This provides us with a novel set of perspectives based on graphics, and conceptualization of target events may be affected by the features of the graphics when people rely on those perspectives.

## References

Fillmore, C. J. (1997). *Lectures on Deixis*. Stanford, CA.: CSLI Publications.

Levinson, S. (1996). Frames of Reference and Molyneux's Question: Cross-linguistic Evidence. In P. Bloom, M. A. Peterson, L. Nadel, and M. Garrett. (Eds.) *Space and Language*, 109–169. Cambridge: MIT Press.

Neilson, I., and J. Lee (1994). Conversations with Graphics: Implications for the Design of Natural Language/Graphics Interfaces. *International Journal of Human-Computer Studies 40*, 509–541.

Taylor, H. A., and B. Tversky (1996). Perspective in Spatial descriptions. *Journal of Memory and Language 35*, 371–391.

Umata, I., A. Shimojima, and Y. Katagiri (2000). Talking through Graphics: An Empirical Study of the Sequential Integration of Modalities. *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*, 529–534.

Umata, I., Y. Katagiri, and A. Shimojima (2002). Movement Conceptualizations in Graphical Communication. In Hegaty, M., Meyer, B. and N. Hari Narayanan (Eds.) *Diagrammatic Representation and Inference LNAI 2317*, 3–17. Berlin: Springer.