Graphical representation in graphical dialogue

Patrick G. T. Healey
Information, Media, and Communication Research Group,
Department of Computer Science, Queen Mary University of London, E1 4NS
London, UK. email: ph@dcs.qmw.ac.uk

Nik Swoboda
Department of Computer Science; Indiana University, Bloomington, USA
email: nsoboda@cs.indiana.edu

Ichiro Umata and Yasuhiro Katagiri
ATR Media Information Science Research Laboratories, Kyoto, Japan
email: umata@atr.co.jp, katagiri@atr.co.jp

(Received 1 March 2001, and accepted in revised form 4 April 2002)

This paper explores the influence of communicative interaction on the form of graphical representations. A referential communication task is described which involves exclusively graphical dialogue. In this task subjects communicate about pieces of music by drawing. The drawings produced fall into two basic types: Abstract and Figurative. Three hypotheses are developed about the factors influencing the use of these drawing types: efficiency of production, suitability for the task and level of communicative interaction. Experimental evidence is presented which indicates that the drawing types do not differ in the amount of effort required to produce them. The results indicate that (1) Abstract drawings are more effective than Figurative drawings for comparative tasks and (2) a key constraint on their use is level of direct communicative interaction. It is argued that these observations result from differences in the underlying semantic models of music associated with the drawing types and the consequences these differences have for communicative coordination.

1. Introduction

In contrast to speech, gesture, gaze and expression, the use of drawing as a mode of communication has received little attention. Research has primarily focused on the cognitive and computational properties of graphics as systems of representation, passing over their use in communicative interaction (for a review see Scaife & Rogers, 1996). These representational properties are consequently foregrounded when considering the potential benefits of drawing in interaction. For example, the benefits of drawings and sketches in architect’s design interactions seem to depend on their advantages as systems for representing spatial layout and orientation. Experiments with communication tasks that involve a significant spatial component also suggest that the key contribution of drawing to interaction is their use to represent space (e.g.
Whittaker, Geelhoed & Robinson, 1993; Healey, McCabe & Katagiri, 2000). For tasks such as collaborative model building, a drawing can provide a much more concise and efficient representation of a configuration of components than an equivalent verbal description could. Drawings can also be used to represent abstract, non-spatial structures. For example, Venn diagrams and Euler circles capture logical structure. Using a task described in more detail below, Healey, Swoboda, Umata and Katagiri (2001) also report how drawings can be used to communicate musical structure.

The main concern of this paper, however, is not with what drawings contribute to communication as systems of representation per se but with drawing as a mode of interaction. This distinction can be illustrated by analogy to natural language. Models of languages as representational systems typically focus on their syntactic and semantic structure. These models have provided powerful analyses for some instances of language use, such as text or monologue; however, they have proved less effective for analyzing conversational interaction. In conversation, significant aspects of language use are related to sustaining the interaction itself. Some utterances, such as backchannels, primarily contribute to maintaining the coherence of the exchange. There are also devices and procedures that help to manage changes of turn and topic or more sophisticated processes such as detection and resolution of problems in interpretation (Sacks, Schegloff & Jefferson, 1974; Levinson, 1983; Schegloff, 1992). These processes influence the form of language in conversation independent of its effectiveness or efficiency as a system for representing a particular domain. There is also evidence that people use conversational interaction to adapt and modify the conventions they are using to represent a domain (Clark & Wilkes-Gibbs, 1986; Garrod & Anderson, 1987). Contemporary accounts of language use attend both to its properties as a representational system and to its use in communication. Work on graphical interaction has, however, focused primarily on ‘graphical monologue’. The present paper is primarily concerned with graphical dialogue.

There are several sources of evidence which suggest that drawing constitutes a distinctive mode of communicative interaction. A survey reported by van Sommers (1984) indicates that approximately half of informal drawing activities take place with or for an audience. Although van Sommers does not report how often these interactions involve exchanges that could be considered to constitute graphical dialogue, his findings show that drawing occurs in a wide variety of interactional contexts. The category most frequently cited is production of a sketch map of a local area, either as part of an explanation or in order to give directions. The second major category relates to activities with children including games and amusements, teaching or helping with homework and helping children learn to draw. Additional categories of public drawing include sketching of hair, makeup and clothing, sketching house plans, drawing to express feelings, defacing pictures and drawing people.

Drawing is also tightly integrated into more specialized interactions. For example, Neilson and Lee (1994) provide a detailed analysis of the complex interrelationships between drawing, speech and gesture during architect–client interaction. Similarly, Engle (1998) provides experimental evidence that graphics, gesture and language combine in explanatory dialogues to create composite communicative signals (see also Clark, 1996).
The potential of graphical interaction as a mode of communication is perhaps most clearly illustrated by research on the use of drawing as part of a therapeutic intervention for people with aphasia. This work indicates that drawing can provide a particularly powerful auxiliary mode of communication when language is compromised (e.g. Lyon, 1995a, b; Sacchett, Byng, Marshall & Pound, 1999; Sacchett, this issue).

An additional motivation for focusing on graphical dialogue is that its potential use as a channel of communication has been extended through the development of technologies to support multimodal interactions. Shared ‘virtual’ whiteboards were one of the basic components of early collaboration technologies such as CoLab (Stefik, Foster, Bobrow, Kahn, Lanning, & Suchman, 1987). Greenberg, Hayne & Rada (1995) provide examples of a variety of advanced drawing space applications that have been developed to support drawing in computer-mediated interaction. Evaluations of whiteboard technologies have indicated that drawing can be a valuable mode of communicative interaction (Bly, 1988; Bly & Minneman, 1990; Tang & Minneman, 1991; Whittaker et al., 1993; Healey et al., 2000).

Together, these considerations provide a prima facie case for the significance of graphical dialogue as a mode of communicative interaction. In the studies cited above, drawing occurs as a secondary medium of communication in a multimodal exchange and is rarely the focus of analysis. The present paper focuses on a study of task-oriented interaction in which subjects’ sole means of communication is by drawing (Healey et al., 2001). This study suggests that, like natural language dialogue, the character of the interaction itself is a key constraint on the form of graphical representations; specifically, that subjects ability to interact directly during an exchange alters the type of representation they use.

1.1. STRUCTURE OF THE PAPER

The paper begins by describing a music drawing task and reviewing the main findings from Healey et al. (2001). Two alternative explanations, that appeal to task demands and efficiency of production of the drawings, are described. An experiment is reported that investigates the influence of these two factors. The results suggest that neither of these explanations is sufficient to account for the original findings and that level of direct interaction is a key constraint on the use of the different representation types in this task. The discussion section draws out some implications of the findings and possible further avenues of investigation.

2. The music drawing task

In experimental tasks based on map drawing or kitchen design, subjects can often exploit graphical conventions they have learned performing similar tasks in everyday life; for example, from architectural designs or maps they have been taught to read. The problem of establishing communicative coordination in such tasks is simplified by prior experience with the task and the availability of these conventions. The music drawing task was designed to provide a novel task which involves exclusively graphical
communication about a highly structured domain for which there are few existing representational conventions.

In outline, the task is somewhat like a musical version of the ‘Pictionary’ party game. Subjects are divided into pairs and asked to draw pictures of pieces of music. They are free to draw anything they like, subject to the constraint that no letters or numbers are used. Subjects are seated in separate rooms so that they cannot hear what their partners are listening to. All drawing takes place on a shared virtual whiteboard which provides the only communication link between them. In a basic version of this task the members of a pair have one piece of music each. Each member of the pair produces a drawing of their piece on the whiteboard and they use the drawings to judge whether they have the same piece or not. Drawing and playback of music are self-paced. Once they have decided, each subject presses a button to indicate their choice and they receive feedback about whether they were correct. The task is repeated a number of times with selection of the pieces of music under experimental control. This repetition provides subjects with the opportunity to develop their communicative coordination over time. A series of drawings produced by two pairs over six trials of the task are shown in Figures 1 and 2.

2.1. MUSIC DRAWING TYPES

The drawings that subjects produce in this task can be reliably classified into three types: Abstract, Figurative or Composite (Healey et al., 2001). Abstract drawings, like

\[\text{\^{P}ictionary is a registered trademark of Pictionary Inc., Nevada, USA.}\]
those produced by the pair shown in Figure 1, can be characterized as depicting formal aspects of the music. Typically they involve a contour or graph-like representation constructed from combinations of continuous lines, blobs, short strokes or column-like elements. These drawings are used to represent a variety of aspects of musical structure, e.g. pitch, stress, harmonic structure, chord structure, rhythm, tempo, texture or intensity. The Figurative drawings, by contrast, depict recognizable scenes or situations. They may include, amongst other things, faces, landscapes, scenes, objects or figures. A sequence of Figurative drawings by a pair is shown in Figure 2. These drawings appear to represent the music in a more ad hoc manner, drawing on a mixture of associations between pieces and scenes or situations. For example, a cocktail bar or a Manhattan sky-line might be used to represent a canonical Jazz piece and a landscape or a sleeping figure might be used for a restful piece. Composite drawings are the smallest category and involve some combination of the two styles for example, a smiling face as an annotation to a simple contour.

The results reported by Healey et al. (2001) provide two lines of evidence that the use of these drawing types is primarily determined by the pattern of communicative interaction. The first piece of evidence is that individuals show a reliable tendency to match the type of drawing produced by their partner. This pattern of entrainment is illustrated by the sequences in Figures 1 and 2. On each trial the members of the pair each produce drawings of the same type. The tendency to match the type of drawing produced by a partner appears to be independent of the characteristics of the piece being drawn, which is systematically varied. It also seems to be independent of the expertise, graphical or musical, of the individual producing the drawings. This pattern...
of matching or entrainment parallels results for natural language dialogue where conversational partners show a strong tendency to match both content and syntax of their utterances (Garrod & Anderson, 1987; Brannigan, Pickering & Cleland, 2000).

The second piece of evidence for the influence of communicative interaction derives from a contrast between two versions of the music drawing task: Interactive and Non-interactive. In the Interactive task, described above, both individuals can draw on the whiteboard simultaneously and both are able to modify or erase each other’s drawings. In the Non-interactive version only one member of the pair draws on each trial. In this case, the members of a pair alternate between “Drawer” and “Chooser” roles. On each trial the Drawer has a single target piece that they draw on the whiteboard. The Chooser watches the drawing develop on the whiteboard and uses it to select between two pieces, the target and a distracter. As for the Interactive task, feedback about the accuracy of their choice is provided at the end of each trial; however, no direct graphical interaction is possible because the whiteboard is configured to block the Chooser from drawing. Subjects do, however, swap between Drawer and Chooser roles on each trial.

Healey et al. (2001) found that the distributions of drawing types in the Interactive and Non-interactive tasks were reliably different. The results, summarized in Table 1 show that in the Interactive task there is a strong preference for Abstract drawings whereas in the Non-interactive task there is a strong preference for Figurative drawings (the “None” column indicates the proportion of occasions on which no drawing was produced).

2.2. EFFECTS OF INTERACTION

Healey et al. (2001) propose that the difference in use of the Abstract and Figurative drawing types is explained by differences in the semantic models of the music they invoke and the implications this has for communicative coordination.

The first part of this explanation is based on the suggestion that the two drawing types differ in their levels of compositionality and generalization. Abstract drawings support quasi-compositional inferences about the internal structure of the pieces of music. For example, the relative length or height of two different parts of a contour can signal the relative duration or pitch of two parts of the music. Figurative drawings, by contrast, do not support structural inferences of this kind. The layout of objects in a scene or landscape is not interpreted as indicating something about the musical structure of a piece. Rather, it is the drawing taken as a whole that signals the identity

<table>
<thead>
<tr>
<th>Task Version</th>
<th>Abstract (%)</th>
<th>Figurative (%)</th>
<th>Composite (%)</th>
<th>None (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interactive</td>
<td>59</td>
<td>21</td>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td>Non-interactive</td>
<td>27</td>
<td>64</td>
<td>8</td>
<td>1</td>
</tr>
</tbody>
</table>
of the music. Abstract drawings also generalize across instances. If the interpretation of an Abstract drawing has been fixed, for example as a pitch contour or a representation of tempo or intensity, then it has a relatively straightforward generalization to other pieces: there is a pitch contour or intensity profile for every piece of music. Figurative drawings, by contrast, do not readily generalize across pieces. Although some limited generalizations are possible, a distinguishing landscape or bar scene cannot be generated for every piece of music.

The second part of the explanation depends on the implications, for communicative coordination, of these differences between Abstract and Figurative drawings. Effective communication using Abstract drawings, it is argued, depends on both members of a pair coordinating on a relatively specific interpretation of the drawings with respect to the music. Specifically, they are under pressure to coordinate on the particular parameter or regularity they are using to classify the pieces (e.g. pitch, intensity, tempo, texture, melody, rhythm) and on how this maps to the particular form of Abstract drawing they use (e.g. components, scale, axes). Figurative drawings, by contrast, require weaker coordination with respect to the underlying semantic model of the music. Members of a pair do not need to coordinate on their basic classification of the pieces of music. They rely instead on ad hoc associations between particular pieces and a variety of possible scenes, activities, situations or moods. Different associations are generated for each piece and their pictorial representations rely on general conventions that make minimal reference to the musical structure of each piece. The claim is that the Figurative drawings invoke a relatively underspecified, or more liberal, semantics for the music.

The final step in this explanation is to provide an account of the difference in the distribution of these two drawing types between the Interactive and Non-interactive tasks. Healey et al. (2001) propose that the difference in use of drawing types in the Interactive and Non-interactive versions of the task arises because of differences in the level of interaction required to support their coordinated use. Specifically, because direct interaction provides subjects with opportunities for directly adapting and modifying each others’ drawings, it supports the development of the semantic coordination necessary for the use of the Abstract drawings. Where these mechanisms of direct interaction are not available, subjects rely on Figurative drawings.

2.3. DRAWING EFFECTIVENESS AND COLLABORATIVE EFFORT

Two alternative explanations of the use of the drawing types can be developed which do not appeal to problems in sustaining semantic coordination.

The first alternative addresses a possible confound between level of interaction and task demands in the two versions of the music task. The Interactive and Non-interactive versions of the music task vary not just in the level of interaction between participants but also in the degree of comparison they involve. The Non-interactive version of the music task is similar to a simple definite reference task. Only one piece of music is drawn, by one subject, on each trial. In the Interactive version two pieces are drawn, one by each subject, and directly compared on each trial. By hypothesis, Abstract drawings should be more effective for making direct comparisons between pieces of music. As noted above, they are designed to invoke a common frame of reference which
facilitates comparisons both within and between pieces. Since the Interactive task involves more direct comparison of pieces than the Non-interactive task, it could promote the use of Abstract drawings independently of level of interaction.

The second alternative explains the differences in the use of Abstract and Figurative drawings in terms of differences in efficiency of production. Intuitively, drawings of bar scenes and landscapes are more complicated to produce than the graph-like Abstract representations. Clark and co-workers have shown that in conversational interaction there is pressure to minimize the joint effort required to communicate (Clark & Wilkes-Gibbs, 1990). One effect of this pressure is a reduction in the complexity of descriptions produced by conversational partners. For example, recurrent referring expressions require progressively fewer words and turns to complete, in extended conversational exchanges. The semantics of these referring expressions remains constant as they are reduced. However they become more difficult to interpret for anyone who is not actively involved in the exchange and hence, in one sense, more abstract (e.g. Clark & Wilkes-Gibbs, 1990; Wilkes-Gibbs & Clark, 1992).

An example of how this process could operate in the music task is provided in Figure 3. In this figure three pairs of drawings are shown that illustrate the representations being produced by one pair during early, middle and late trials of the

![Figure 3. Transformation of a keyboard into a graph.](image-url)
interactive version of the task. In the early trials Figurative pictures of keyboards predominate. Schematic hands are drawn to indicate which notes are being played and lines represent lateral motion to suggest tempo and progression. In the middle trials the details of the keyboard and hands begin to drop out. By the final trials the pairs are predominantly using graph-like representations. These still reflect the “x-axis” of the keyboard with pitch running left to right but now the hands and piano keys have been replaced by arrows, apparently deriving from the original lines to indicate hand motion, to represent tempo and progression. The semantics of these drawings appears to have remained more or less constant while their pictorial complexity has been significantly reduced.

An explanation constructed on these lines thus treats the Abstract drawings as more efficient, abbreviated forms of the Figurative drawings (cf. Bartlett, 1932; Tversky, 1995). In the present case, the difference in the distribution of Abstract and Figurative drawings can be accounted for by reference to the mechanism through which the abbreviations or reductions are achieved. In Clark’s model (Clark & Wilkes-Gibbs, 1990) this mechanism is grounding. A referring expression is grounded if and only if the addressee signals that they accept it. Successive cycles of presentation and acceptance provide pairs with a means of gradually reducing and refining their referring expressions. In the Non-interactive version of the music task, subjects are restricted to between-trial cues to grounding: feedback about success, speed of choice of a drawing and matching of drawing type. In the Interactive task, by contrast, subjects have a variety of within-trial opportunities for grounding. For example, they can indicate directly using question marks, ticking, circling or even redrawing whether they accept their partner’s drawing. This allows them to complete several cycles of grounding within a single trial. Consequently, the predominance of Abstract drawings in the interactive task might be explained as a reduction in drawing effort facilitated by the enhanced opportunities for grounding.

These two alternative explanations for the distribution of drawing types in the music task were investigated experimentally.

3. Experiment: “Matching” and “Discrimination” tasks

Two versions of the Interactive music drawing task were developed. In the first, ‘Matching’, version (the basic task described above) subjects have one piece of music each. Half the time they have the same piece and half the time they have different pieces. On each trial their task is to determine, using drawing alone, if the pieces they have are the same or different. In the second version, the ‘Discrimination’ task, participants each have two pieces of music. One piece is common to both and the other two are distractor pieces. In this case they must try to determine which of their two pieces is the one held in common.

Both tasks provide opportunities for grounding through direct interaction; however, they differ in the degree of comparison between pieces involved. In the Matching task each subject draws one piece of music and makes one comparison: their drawing with their partner’s drawing. In the Discrimination task each subject draws two pieces of music and must make four comparisons: each of their two drawings with each of their
partner’s. If more comparative tasks favour the use of Abstract drawings, then the Discrimination task should promote greater use of Abstract drawings than the Matching task.

3.1. MATERIALS
In order to circumvent a strategy, identified in pilot studies, of drawing pictures of the instruments being used in a particular performance only solo piano pieces were used. Several criteria were used in the selection of the music. To help ensure a mixture of styles, no more than two pieces by any given composer or performer were used. In order to reduce the possibility of simple visual associations with names (e.g. “Swan lake”) or patterns of exposure (e.g. themes used in advertisements) easily recognizable pieces were avoided. A total of 112, 30 s, piano solo pieces, were used.

All drawing was carried out on an LCD tablet (combined graphics tablet and screen) using shared virtual whiteboard software written specially for the task. The LCD tablets support a relatively naturalistic stylus based form of graphical input that allows subjects to draw directly onto the screen and erase using the reverse end of the stylus.

The experimental whiteboard consists of a shared drawing area, a strip palette of eight colours and a task bar containing buttons for playing back the pieces of music and making responses as illustrated in Figure 4. The whiteboard was configured so that both participants could draw and erase anywhere on the whiteboard. All playback of pieces was self-paced and controlled through “Stop” and “Play” buttons in the task bar. Response buttons were also positioned in the task bar: “Same” or “Different” for the Matching task and “Piece A” or “Piece B” for the Discrimination task. In order to prevent subjects from using response button positions as a basis for coordination, the

![Figure 4. Experimental whiteboard (Discrimination task).](image-url)
assignment of the “Piece A” and “Piece B” buttons and the “Same” and “Different” buttons was randomized across trials.

3.2. SUBJECTS

Twenty-four participants, all Japanese, were recruited from a variety of disciplines at local colleges and universities. Sixteen were male and eight female, average age 19, and they were paid an honorarium for taking part.

3.3. PROCEDURE

Subjects were divided into 12 pairs, each member of a pair seated in a separate room, and given written instructions explaining the two versions of the task. They were asked to collaborate to carry out the task as quickly and as accurately as possible. As soon as one subject made a decision by pressing the appropriate button on the task bar, further drawing was blocked and the other member of the pair was prompted for a decision. If a 2 min time limit expired before either subject had decided, a dialogue window appeared to prompt a final choice. Subjects received feedback after each trial about whether their choice had been correct. Because it is a collaborative task the feedback that they were “correct” was given only if both subjects chose correctly. If one or both chose incorrectly they both received feedback that their choice was “incorrect”. After instruction, each pair was given a short demonstration of simultaneous drawing on the whiteboard in order to familiarize them with its basic operation.

3.4. DESIGN

The experiment employed a within-subjects, factorial design with all pairs carrying out both tasks. The independent variables were Task (Matching vs. Discrimination) and Experience, coded as four blocks of trials corresponding to increasing levels of exposure to the task. An additional manipulation of medium (Mouse vs. Stylus) is discussed in Healey et al. (2001). Order of treatments was counterbalanced across pairs using a semi-Latin square design. Selection of music was constrained so that no subject heard the same piece of music twice. This design resulted in a total of 68 trials per pair with order of conditions and materials counterbalanced across pairs.

4. Results

To assess the effects of task structure on performance, the proportion of correct responses out of 68 was calculated for each pair. An analysis of variance with Task (Matching vs. Discrimination) and Experience (levels 1–4 corresponding to four blocks of 17 trials, one for each quarter of the experiment) as within-subject factors showed simple main effects of Task \(F(1,47) = 12.07, p = 0.00\)\(^1\) and Experience \(F(3,47) = 4.84, p = 0.01\). Pairs were reliably more accurate in the Matching task.

\(^1\) A criterion level of \(p < 0.05\) was adopted for the interpretation of all statistical tests.
The percentage accuracy for each level of task experience is given in Table 2. The pattern suggests a steady increase in accuracy with experience. Linear trend analysis confirmed that participants became more accurate across trials ($t(47) = 3.48$, $p$(one tailed) = 0.00).

Average response times for each pair in each condition were calculated to provide a second measure of performance. Analysis of variance with Task (Matching vs. Discrimination) and Experience (levels 1–4) showed a simple main effect of Task ($F_{(1,47)} = 23.89$, $p = 0.00$) and Experience ($F_{(3,47)} = 4.07$, $p = 0.02$). Subjects were slower to complete the Discrimination task (101 238 ms) than the Matching task (81 067 ms). Table 3 indicates that subjects become faster with experience and this is confirmed by a linear trend analysis ($t(47) = 2.38$, $p$(one tailed) = 0.01).

### Table 2

Percent correct identification with experience (matching and discrimination conditions combined)

<table>
<thead>
<tr>
<th>Block of trials</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>45%</td>
<td>54%</td>
<td>59%</td>
<td>64%</td>
</tr>
</tbody>
</table>

### Table 3

Changes in response times with experience (ms)

<table>
<thead>
<tr>
<th>Block of trials</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>99 425</td>
<td>97 537</td>
<td>85 142</td>
<td>84 707</td>
</tr>
</tbody>
</table>

62% than in the Discrimination task 49%. The percentage accuracy for each level of task experience is given in Table 2. The pattern suggests a steady increase in accuracy with experience. Linear trend analysis confirmed that participants became more accurate across trials ($t(47) = 3.48$, $p$(one tailed) = 0.00).

Average response times for each pair in each condition were calculated to provide a second measure of performance. Analysis of variance with Task (Matching vs. Discrimination) and Experience (levels 1–4) showed a simple main effect of Task ($F_{(1,47)} = 23.89$, $p = 0.00$) and Experience ($F_{(3,47)} = 4.07$, $p = 0.02$). Subjects were slower to complete the Discrimination task (101 238 ms) than the Matching task (81 067 ms). Table 3 indicates that subjects become faster with experience and this is confirmed by a linear trend analysis ($t(47) = 2.38$, $p$(one tailed) = 0.01).

#### 4.1. DRAWING TYPES

Figures 1 and 2 show two sequences of drawings produced by pairs over six trials in the Matching task. As the pictures illustrate, the typical strategy for the Matching task was to divide the screen into two halves and produce drawings of each piece. Once their initial drawings had been produced, subjects employed a variety of devices including arrows, crosses, circles and ticks to identify possible correspondences between the pieces and to signal whether they judged the pieces to be the same. Figures 5 and 6 show examples of four Figurative and four Abstract drawings, respectively, produced in one

---

5 In both cases, performance is substantially above the chance level of 25%. It should be noted that although the “real” chance level is 25%, the effective level of chance changes depending on subject’s strategies. To be successful, subjects must coordinate two things, their interpretation of what the drawings represent and their joint decision based on that interpretation (i.e. to be correct they must both choose “Same” or “Different”, or “Piece A” or “Piece B”, respectively). Once subjects solve the simpler problem of communicating their choice, for example by using ticks and crosses or arrows to indicate combinations of pieces, the effective level of chance becomes 50%.
trial of the Discrimination task. The typical strategy for the Discrimination task involved dividing the screen into four quadrants and producing a drawing of one piece in each quadrant. As for the Matching task, subjects employed a variety of graphical devices to indicate possible correspondences and differences between the pieces and to coordinate on their final decision.

The whiteboard log files from each trial were separated into two files each corresponding to the drawing activity of one subject. These files were used to generate separate images of each subject’s drawings and then classified into one of the following

**Figure 5.** Figurative drawings from a trial of the Discrimination task.

**Figure 6.** Abstract drawings from trial of the Discrimination task.
categories: Abstract, Figurative or Composite. The classification was made following the same criteria as used in Healey et al. (2001) and the original pieces of music were not played back during classification. Previous work has established that inter-judge agreement on this coding classification is high (Kappa = 0.9, N = 287, k = 2, Healey et al., 2001). An additional category of "None" was introduced to deal with a small number of cases 4% in which one or both of the subjects had not drawn a picture of a piece on a given trial. In some cases this was due to problems with the whiteboard, in other cases subjects used auxiliary devices such as arrows, ticks and crosses to indicate a choice but did not draw the piece itself.

The proportion of trials on which pairs produced Abstract drawings in each condition was calculated. An analysis of variance with Task (Matching vs. Discrimination) and Experience (levels 1–4) as within-subject factors showed a main effect of task ($F_{(1,47)} = 6.53, p = 0.02$) and no interactions. Although the unfocused comparison of all four levels of experience showed no effect (omnibus $F_{(3,47)} = 2.84, p = 0.06$), as Figure 7 illustrates there was a small but steady increase in the use of Abstract drawings as task experience increases. This is confirmed by a linear trend analysis ($t_{(47)} = 2.71, p$(one tailed) = 0.01).

The effect of the Matching and Discrimination tasks on the use of drawing types is illustrated in Figure 8. In the Matching task pairs used Abstract drawings 54% of the time, in the Discrimination task they used them 63% of the time. For comparison, Figure 8 also includes the proportions of drawing types for the Non-interactive task from Healey et al., (2001).

4.2. EFFECTIVENESS AND EFFICIENCY OF DRAWING

Response accuracy depends on the choices of both members of a pair and as a result it cannot be directly associated with individual drawings. In view of this, two measures were used to assess the effects of drawing type on accuracy of response. Firstly, accuracy of response was compared for cases in which both members of a pair chose the
same type of drawing (i.e. both Abstract, both Figurative or both Composite) with cases in which they selected different drawing types (i.e. Abstract–Figurative, Abstract–Composite or Figurative–Composite). A chi-squared analysis showed no reliable effect of coordination of drawing type on accuracy of response ($\chi^2(1) = 1.27, \ p = 0.26$). Secondly, to evaluate the contribution of drawing type, the accuracy of responses was compared for trials in which members of a pair both produced Abstract drawings or both produced Figurative drawings. Chi-squared analysis of the raw frequencies indicated no reliable difference in accuracy ($\chi^2(1) = 1.27, \ p = 0.25$). Accuracy in Figurative trials was 49%, in Abstract trials it was 55%.

The effort required to produce referring expressions in natural language is sometimes measured by the number of words and number of turns used (e.g. Clark & Wilkes-Gibbs, 1986). To provide analogous indices for graphical dialogue, two basic measures of drawing effort were used. The complexity of executing the drawings was indexed by the number of lines logged by the whiteboard. The overall amount of drawing was indexed by the amount of ‘ink’ or number of pixels logged by the whiteboard. In order to associate measures of drawing activity with individual drawings, only logs from the Matching task was used. The number of lines and number of pixels used for each drawing by each subject was calculated. For analysis, trials were grouped according to the type of drawing produced (Abstract, Figurative or Composite) and two levels of experience corresponding to the first and second half of the experiment (Early vs. Late).

Analysis of variance of the mean number of lines used with factors of drawing type (Abstract, Figurative or Composite) and experience (Early vs. Late) shows a significant main effect of drawing type ($F_{(2,83)} = 10.22, \ p = 0.00$) and a significant main effect of experience ($F_{(1,83)} = 7.93, \ p = 0.01$) but no interaction ($F_{(2,83)} = 0.15, \ p = 0.86$). Observed power of this analysis is high (0.98 for drawing type and 0.79 for experience, alpha = 0.05). Pairwise comparisons show that Abstract and Figurative drawings do not
differ in the number of lines used to produce them (Bonferroni, \( p = 0.83 \)) but both are reliably different from Composite drawings (Bonferroni: Abstract vs. Composite, \( p = 0.00 \); Figurative vs. Composite \( p = 0.02 \)). On average, Composite drawings require approximately 25% more lines to produce than Abstract or Figurative. As the means in Table 4 show, regardless of drawing type it took fewer lines to construct each drawing as experience increased.

Analysis of variance for mean total amount drawn shows the same pattern of results with a significant main effect of drawing type \( (F_{(2,83)} = 7.82, \ p = 0.00) \), a marginal effect of experience \( (F_{(1,83)} = 3.88, \ p = 0.05) \) and no interaction \( (F_{(2,83)} = 0.77, \ p = 0.46) \). The observed power of the analysis is high for drawing type (0.98) but low for experience (0.50). As before, pairwise comparisons show that Abstract and Figurative drawings are not reliably different (Bonferroni, \( p = 0.32 \)) but both are reliably different from the Composite drawings (Bonferroni: Abstract vs. Composite, \( p = 0.02 \); Figurative vs. Composite \( p = 0.00 \)). As Table 5 shows, Composite drawings require, on average, approximately 35% more drawing to produce than Abstract or Figurative drawings. There is also an overall reduction in the amount of drawing with experience across all drawing types.

### 4.3. DIRECT GRAPHICAL INTERACTION

Subjects in the Music Drawing task generally tend to divide the screen space between them so that they can draw their target pieces without accidental overlap or interference from their partner’s drawing. However, as Figures 1, 2, 5 and 6 illustrate, subjects also sometimes annotate, modify and clarify each other’s drawings. A variety of devices are used in this way: bracketing, underlining and circling to highlight particular regions; bracketing, arrows and lines to identify possible correspondences or patterns of

<table>
<thead>
<tr>
<th>Task experience</th>
<th>Drawing type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Abstract</td>
</tr>
<tr>
<td>Low</td>
<td>21.5</td>
</tr>
<tr>
<td>High</td>
<td>14.7</td>
</tr>
</tbody>
</table>

### Table 4

**Mean number of lines used**

<table>
<thead>
<tr>
<th>Task experience</th>
<th>Drawing type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Abstract</td>
</tr>
<tr>
<td>Low</td>
<td>2728</td>
</tr>
<tr>
<td>High</td>
<td>2079</td>
</tr>
</tbody>
</table>

### Table 5

**Mean amount of drawing (pixels)**
alignment; and question marks and ticks to query or clarify particular regions and elements (as well being used as indicators of agreement and disagreement).

In order to assess the level of direct graphical interaction that occurred between the members of a pair, the whiteboard log files were analyzed for amount of spatial overlap in the drawing activity. The amount of overlapping drawing was calculated by placing an arbitrary $25 \times 25$ grid over each drawing and scoring the number of squares in which both members of a pair drew. In order to isolate the effects of drawing type on level of direct graphical interaction only trials where pairs had used drawings of the same type were included for analysis.

The overlap scores for each trial were entered in an analysis of variance with Task (Matching vs. Discrimination), Accuracy (Correct vs. Incorrect) and drawing types (Abstract vs. Figurative) as between-subject factors. This showed a main effect of drawing type ($F_{(1,529)} = 5.34, p = 0.02$) but no effect of task ($F_{(1,529)} = 0.71, p = 0.40$) or accuracy ($F_{(1,529)} = 1.06, p = 0.30$) and no interactions. Overall, trials in which subjects both produced Abstract drawings involved 21.6\% more overlap in their drawing than those in which both subjects produced Figurative drawings. Although the analysis shows no effect of task or accuracy on the amount of drawing overlap the observed power of the analysis for these factors is low (0.13 and 0.18, respectively) and no firm negative conclusion can be drawn.

5. Discussion

The Music Drawing task illustrates the potential power and flexibility of graphical dialogue. Although the task is designed to be unfamiliar, and subjects are initially hesitant, they found it engaging and laughter was often noted during the remote whiteboard exchanges. Subjects were able to carry out both versions of the task and they became quicker and more accurate with experience.

The Music Drawing task also fulfils the experimental aim of eliciting novel representational strategies from participants. Few subjects had attempted to draw pieces of music before and probably never as part of a graphical dialogue. Standard musical notation, the only conventional strategy available, could be used only if both members of a pair had the right expertise. Only two pairs attempted this approach in the present experiment and in both cases they abandoned it in favour of more ad hoc strategies. In general, subjects employed a range of idiosyncratic representations to carry out the task and no two pairs seemed to settle on exactly the same approach.

Although it obscures much interesting variation, the drawings subjects produce in the Music Drawing task can be reliably grouped into the three categories Abstract, Figurative and Composite. As noted, pairs tend to coordinate strongly on their choice of drawing type under this classification. However, coordination per se does not seem to be associated with task performance. In the present data neither coordination itself nor coordination on the use of particular drawing types was associated with accuracy of response.

In the introduction three hypotheses were outlined to account for the patterns of use of drawing types in the Music Drawing task: level of interaction, effectiveness for the task and efficiency of production. Taking these in reverse order, the present results do
not support an explanation based on differences in the effort required to produce drawings of each type. There is a reliable trend to use more Abstract drawings as task experience increases and subjects also become more efficient at drawing, as indicated by the reductions in the number of lines and amount of “ink” used to produce each drawing. However, these trends appear to be independent. No reliable difference was found between the Abstract and Figurative drawings for either of the measures of effort. In addition, all drawing types benefit equally from the reductions that occur with experience.

Additional evidence against an effort-based explanation comes from previous work, which has shown that manipulating ease of production, by switching between mouse and stylus as input devices, does not affect the frequency of use of Abstract and Figurative drawings (Healey et al., 2001). Considered together these findings count against an analysis of Abstract drawings as simplified or abbreviated forms of the Figurative drawings (cf. Bartlett, 1932; Clark & Wilkes-Gibbs, 1986; Tversky, 1995).

The hypothesis that the drawings differ in their effectiveness for carrying out the task gains more support. It was argued in Section 2.2 that the different drawing types are associated with different underlying semantic models of the pieces of music; in particular, that Abstract drawings support quasi-compositional or “structural” inferences about each piece and generalize across instances more readily than Figurative drawings. These considerations inform the prediction that Abstract drawings should be better suited to more comparative tasks than Figurative drawings. The finding that the more comparative Discrimination task promotes greater use of the Abstract drawing types provides support for this prediction. The interpretation of this result is complicated by the fact that the Discrimination task is also simply more difficult than the Matching task. It requires the production of more drawings in the same time and subjects are slower and less accurate. However, given the finding that Abstract and Figurative drawings are not distinguishable in the amount of effort involved in producing them, it is difficult to explain the difference in use across the two tasks by appeal to time pressure or difficulty alone.

The results of the present study thus provide qualified support for the hypothesis that the Abstract and Figurative drawings are associated with different semantic models of the music. The findings also suggest that the use of these drawing types is sensitive to level of interaction. As Figure 8 illustrates, the difference between the Matching and Discrimination tasks has a relatively small impact on the distribution of drawing types: in both cases Abstract drawings predominate. The difference between Interactive and Non-interactive tasks is more marked: the preference for Abstract and Figurative drawings reverses. However, the influence of level of interaction and level of comparison in the present tasks cannot be clearly separated. Although the Matching task and the Non-interactive task both require each subject to draw one piece of music, the Non-interactive task does not involve direct comparison of drawings. More work is required to resolve this issue.

A second piece of evidence for the influence of level of interaction derives from the finding that participants’ drawings overlap more often during the coordinated use of Abstract drawings than during the coordinated use of Figurative drawings. Although it is possible that this is because subjects are less careful when producing Abstract drawings this does not seem likely. Differences in the effort involved in executing both
types of drawing were not found in the present study. Also the pressure to maintain legibility and avoid accidental overlap should be broadly the same in both cases. It appears more likely that Abstract drawings involve intentionally closer coordination: possibly through more careful alignment of the contours of each drawing, or possibly through greater use of devices such as direct annotation, clarification, circling and underlining to localize and elaborate problems.

6. Conclusion

This study is part of a programme to investigate the character of graphical dialogue and the influence of distinctively interactional factors on the form of graphical representations. With respect to the Music Drawing task, the basic proposal is that the sustained, coordinated use of the Abstract drawings depends on the ability to interact directly. It was claimed that this is due to two things: the relatively high level of semantic coordination required for the use of Abstract drawings and the mechanisms of interaction required to establish and maintain this coordination.

The results of the present study provide some support for this position. The type of abstraction involved in the Abstract drawings is more than abbreviation or reduction of a representation, it appears to reflect abstraction in an ontological sense. Abstract drawings invoke a relatively systematic semantic framework for the pieces of music that draws on regularities such as pitch, intensity or rhythm. Because these regularities allow generalization across instances, the Abstract drawings are more effective for certain tasks but they also require more sustained and precise semantic coordination between individuals to be used effectively. The intuition is that the coordinated use of Abstract drawing types consequently requires more direct interaction to develop and maintain this level of coordination. The current results provide weak support for this conclusion. Direct interaction appears to favour the use of Abstract drawings and where they are used more overlap occurs. However, this provides only indirect evidence of higher levels of semantic coordination. More work is required to test this claim adequately.

If this explanation is correct, the question arises of what the mechanisms of interaction are that mediate these effects. Some of the continuities between the present observations and the findings for natural language dialogue were touched on above. This opens up the possibility that the same generic mechanisms of interaction, such as entrainment, grounding or repair, may operate across modalities. There are also interesting potential discontinuities between graphical and verbal interaction. For example, graphical dialogue can exploit the spatial layout and persistence of drawings to manage topic shifts and turn taking. These interactional devices are unavailable in verbal exchanges. Much of the interest of the present programme depends on the potential (dis)analogies between graphical and verbal dialogue and what this might ultimately indicate about general mechanisms and processes of communicative coordination.

This work was generously supported by ATR Media Information Sciences Laboratory, Kyoto, Japan and the ESRC/EPSRC PACCIT programme under the grant MAGIC: Multimodality and Graphics in Interactive Communication (L328253003). We gratefully acknowledge Prof. R. Bailey and Dr S. Gilmour for advice on the statistical analysis.
References


