Referring Expressions in Discourse about Haptic Line Graphs

Özge Alaçam
Department of Informatics
University of Hamburg
Hamburg/Germany
`alacam@informatik.uni-hamburg.de`

Cengiz Acartürk
Cognitive Science
Middle East Technical
University, Ankara /Turkey
`acarturk@metu.edu.tr`

Christopher Habel
Department of Informatics
University of Hamburg
Hamburg/Germany
`habel@informatik.uni-hamburg.de`

Abstract

Statistical line graphs are widely used in multimodal communication settings and they are crucial elements of learning environments. For visually impaired people, haptic-audio interfaces that provide perceptual access to graphical representations seem as an effective tool to fulfill these needs. In an experimental study, we investigated referring expressions used in a collaborative joint activity between haptic explorers of graphs and verbal assistants who helped haptic explorers conceptualize local and non-local second-order concepts (such as extreme values, trends, or changes of trends). The results show that haptic exploration movements evoke deictically referential links that are essential for establishing common ground between explorers and assistants.

1 Comprehending Graphs through Different Modalities

Data visualization aims at (re-)presenting data so that humans more easily access certain aspects of them (such as trends or anomalies) for thinking, problem solving and communication (Tufte 1983, Kosslyn 1989, 2006, Hegarty 2011, Alaçam, et al., 2013). Among many specific types of representational modalities (such as sketch maps, statistical graphs and schematic diagrams), statistical line graphs have found a widespread use in various daily life and professional settings. For making statistical graphs accessible to visually impaired people, technologies ranging from pure tactile graphs to verbal summaries (Demir et al., 2012) have been proposed. However, haptic presentations of graphs (henceforth, *haptic graphs*) provide a suitable means for visually impaired people to acquire knowledge from data sets, when they are integrated in hybrid systems that employ auxiliary modalities to the haptic-tactile modality, such as sonification and verbal assistance (Abu Doush et al., 2010; Ferres at al., 2013).

Users can explore haptic graphs by hand-controlling a stylus of a force-feedback device, for instance a Phantom Omni® (recently Geomagic® Touch^{TM}, see Figure 1.a), which yields information about geometrical properties of lines. Compared to visual graphs, one drawback of haptic graphs is the restriction of the haptic sense in simultaneous perception of spatially distributed information (Loomis et al, 1991). Comprehension of haptic line graphs is based on explorations processes, i.e. hand movements tracing lines, with the goal to detect shape properties of the graph line explored. The recognition of concavities and convexities, as well as of maxima and minima, is of major importance (see Figure 1.b for a sample haptic line graph).

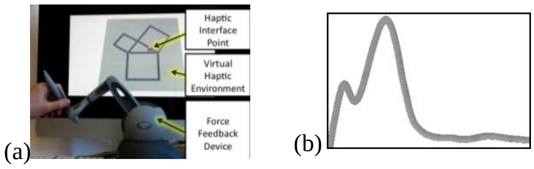

Figure 1. (a) Phantom Omni® device and visualization in a geometry domain (see, Kerzel & Habel, 2013, Fig. 1), (b) sample haptic graph

Although simple line graphs are often considered as a graph type easy to comprehend haptically, there are some critical problems about haptic representation of simple line graphs: Whereas it is only moderately difficult to comprehend the shape of a simple graph line with a single (global) maximum haptically, graphs with several local maxima require additional assistance for most users of haptic graphs. Providing additional information, such as aural assistance through the auditory channel, has been proved to be helpful for resolving some difficulties in haptic graph exploration (cf. sonification, Yu and Brewster, 2003). We propose to use speech utterances (i.e.

verbal assistance) to support—for example—the detection and specification of local and global extrema of graph lines, or other shape based concepts.

For designing haptic graph systems, which are augmented by computationally generated verbal assistance, it is necessary to determine which information, depicted by the graph or by its segments, are appreciated as important by haptic explorers. In this paper we focus on the use of referring expressions within dialogues in collaborative haptic-graph exploration-activities between blindfolded *haptic explorers* and seeing *verbal assistants*. The analyses of these joint activities provide crucial insight about how haptic explorers acquire high-level information from haptically perceived graphs. Moreover, they also provide the empirical basis (i.e. which spatial content should be verbalized) for our long-term goal: the realization of a cooperative system providing blind graph readers with verbal assistance (Habel et. al., 2013, Acartürk et. al, 2014).

1.1 Shape in Line Graphs: Perception, Cognition and Communication

Graph lines inherently convey shape information, namely information about convexities and concavities, about straightness, angles, and vertices. These are evoked in visual perception by visually salient graph-shape entities, in particular by curvature landmarks, positive maxima, negative minima, and inflections (Cohen & Singh, 2007).

From the perspective of a seeing human who describes a line graph, salient parts of the graph line are primary candidates to be referred to. In other words, referring expressions are evoked by visually salient graph entities. The conceptual inventory for verbalizing line-graph descriptions, as well as trend descriptions, has to fulfill requirements from language and perception. Since graph lines can be seen as a specific type of 2D-contours, we include some concepts proved as successful in visual shape segmentation into the inventory of spatial concepts, namely Cohen and Singh's curvature landmarks (2007). In addition to Cohen-Singh landmarks, the case of graph lines requires graph-line specific types of curvature landmarks: since graph lines are finite and not closed, two types of endpoints (left vs. right) have to be distinguished.

In haptic graph exploration the shape of the graph line is a major property for identifying referents by distinguishing it from its distractors. Additionally, certain aspects of graph segments (such as inflection points that show smooth change) are more difficult to acquire in the haptic modality than in the visual modality, largely due to the sequential and local perception with a narrow bandwidth of information in the haptic modality (Habel et. al., 2013). Finally, previous research has shown that not only saliency in the domain of discourse via the linguistic context but also saliency in the visual context influences humans' choice of referring expressions (Fukumura et al, 2010).

Haptic assistive systems that take shape properties of graphical representations into account in design process have been scarce except for a few instances (e.g. see Ferres et al., 2013; Wu et al., 2010). Additionally, there is still a lack of research on the role of shape comprehension in haptic graph exploration. Since the current state-of-the art haptic graph systems would benefit from providing verbal descriptions of shape properties and shape entities, we focus in this paper on the use of referring expression to these entities in collaborative graph explorations.

1.2 Assisted Haptic Graph Exploration: A Joint Activity Approach

Verbally assisted haptic graph exploration can be seen as a task-oriented collaborative activity between two partners, a (visually impaired) explorer (*E*) of a haptic graph and an observing assistant (*A*) providing verbal assistance (see Figure 2). Sebanz and colleagues (2006), who focus on bodily actions, describe joint actions as follows: "two or more individuals coordinate their actions in space and time to bring about change in the environment". In contrast to this characterization, the joint activities that we focus on shall bring about changes in *E*'s mental representations. To reach this goal, *E* and *A* have to establish common "understanding of what they are talking about" (Garrod & Pickering, 2004).

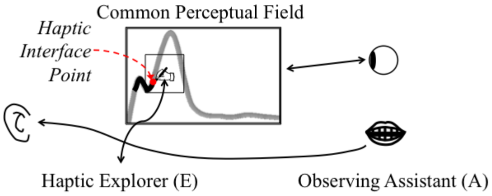

Figure 2. Assisted haptic graph exploration, a joint activity

A and *E* share a common field of perception, namely the haptic graph, but their perception and comprehension processes differ substantially. For example, while *E* explores the highlighted, black segment of the haptic graph, *A* perceives the global shape of the graph, in particular, *A* is aware of shape landmarks and line segments. For example, when *E* explores the first local maximum followed by a local minimum (see Figure. 2), *E* does not have information about the global maximum, which is already part of *A*'s knowledge. Therefore, *E* and *A* have different internal representations of the graph line, and A's referring to the graph could augment *E*'s internal model substantially. For example, uttering "Now you have reached the heights of the last peak" would provide *E* with additional information. Another suitable comment would be "You are in the increase to the population maximum", or even "You are in the increase to the population maximum of about 90, that was reached in 1985". Since verbal assistance is a type of instruction, overspecified referring expressions are adequate for our domain (see Koolen et al., 2011).

The success of the joint activity of explorer *E* and observing assistant *A* in general, and also the success of *A*'s utterances in particular, depends, on the one hand, on joint attention (Sebanz, et al., 2006), and on the other hand, on the alignment of the interlocutor's internal models, especially on building implicit common ground (Garrod & Pickering, 2004). Since *E*'s internal model of the activity space, i.e. the haptic graph and E's explorations, is perceived via haptic and motor sensation, whereas *A*'s internal model of the same space is build up by visual perception, similarities and differences in their conceptualization play the central role in aligning on the situation-model level.

The assisted haptic graph explorations we discuss in this paper can be conceived as an asymmetric joint activity: firstly, the participants have different activity roles (explorer vs. assistant), as well as different sensor abilities; secondly, the participants were told that *E* should initiate the help request and *A* should provide help based on explorer's need. Although the dialogues accompanying haptic explorations are—in principle—mixed-initiative dialogues, explorer-initiatives are the standard case.

Haptic explorers' contributions to the dialogue are given concurrently to their exploration movements. Thus, for the observing assistant, the referring expressions produced are accompanied with the current exploration point on the graph. In other words, *E*'s exploration movement evokes deictically a referential link—analogue to Foster and colleagues' (2008) haptic ostensive reference. And thus, common ground is established and the given-new contract between *E* and *A* is fulfilled (Clark and Haviland, 1977; Clark and Brennan, 1991). In the following turn, *A* is expected to provide most helpful and relevant information for *E* at that particular moment. In particular *A* should provide *E* with content that is difficult to acquire haptically, such as, information about whether a maximum is local or global. To maintain the common ground, *A* has to synchronize her language production with *E*'s hand-movements in a turn-taking manner, since the quality of verbal assistance depends on establishing appropriate referential and co-referential links.

1.3 Shape Concepts in Graph-Line Descriptions

Most qualitative approaches to shape representation focus on the shape of contours (see, e.g., Hoffman & Richards, 1984; Eschenbach et al., 1998), and on curvature landmarks of contours (Cohen and Singh, 2007), such as, positive maxima and negative minima, depending on the concepts of convexity and concavity of contours, and inflection points. However, graph lines require some additional shape representations and shape cognition characteristics beyond the characteristics of contours. In particular, graph lines are conventionally oriented corresponding to reading and writing direction and they are comprehended with respect to an orthogonal system of two axes. The haptic graphs we use in the experiments are realized in a rectangular frame that induces an orthogonal system of axes. The geometric shape concepts for describing graph lines are exemplified with a graph used in our experimental studies (see Figure 3).

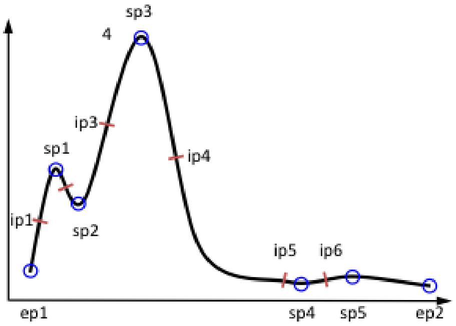

Figure. 3. Qualitative shape landmark ascription for a sample graph (augmented with orthogonal axes for making the reference frame in Table 1 explicit)

Table 1 gives a tabular summary of qualitative representations for selected shape landmarks and induced line segments. The functional character of statistical line graphs leads to the prominence of *value extrema* (in contrast to *curvature extre-*

ma of contours). Since we use in the experiments presented here *smoothed graphs,* these extrema are called *smooth points* (sp). *Inflection points* (ip) are depicted in Fig. 3 but not mentioned in Table. 1.)

Shape landmarks		
	Landmark characteristics	**Global properties**
ep1	left end pt., local min.	higher than sp4, ep2
sp1	smooth pt., local max.	higher than ep1, sp2, sp4, sp5, ep2
sp2	smooth pt., local min.	higher than ep1, sp4, sp5, ep2
sp3	smooth pt., local max.	global max.
sp4	smooth pt., local min.	same height as ep2
sp5	smooth pt., local max.	higher than sp4, ep2
ep2	right pt., local min.	same height as sp4
Shape segments		
	Shape characteristics	**Vertical orientation**
ep1–sp1	curved	steeply upward
sp1–sp2	curved	diagonally downward
sp2–sp3	curved	steeply upward
sp3–sp4	curved	steeply downward
sp4–sp5	curved	slightly upward
sp5–ep2	curved / nearly straight	slightly downward / nearly horizontal

Table 1. Qualitatively described shape landmarks and shape segments.

1.4 Referring to Shape Entities: Semantic Representations

Our long-term goal is to realize an automatic verbal assistance system that provides instantaneous support for haptic explorers during their course of exploration. Empirical studies are needed to study underlying principles of haptic graph exploration, and the effect of linguistically coded content in comprehension of second order entities, such as general and temporally restricted trends based on the recognition of global and local curvature landmarks.

The referring expressions produced by haptic explorers and verbal assistants during collaborative activity give insight about how graph readers comprehend graphs, which elements are mentioned most, and how they are referred to. The investigation of multimodal interactions (namely interaction by means of language, gesture and graph) requires systematic qualitative analysis, as well as quantitative analysis. We followed one of the widely accepted method developed by Dale and Reiter (1995), which addresses the generation of referring expressions, to characterize the semantic properties of graphical segments and the referring expressions produced during collaborative activity. In this paper, we do not aim to go into implementation level in detail, instead we used the method as a tool to make systematic mapping between semantic properties of graphical features and participants' referring expressions. According to Dale (1992), a system that generates referring expressions should at least satisfy Gricean-like conversational maxims targeting adequacy, efficiency and sensitivity. In more detail, a referring expression should contain enough information to allow the hearer to identify the referent, it should not contain unnecessary information and it should be sensitive to the needs and abilities of the hearer. They propose and implement a cost function that assumes (based on empirical research) people first and usually prefer to refer to type properties (zero cost), then to absolute properties. Relative properties and relations (the highest cost) follow them respectively. By following this method, we employed ⟨attribute, value⟩ pair representation to characterize the qualitative representations of graph shapes and landmarks. To illustrate, the attribute set which is available for the "*ep1-sp1*" shape segment (see Table 1) possesses the following properties: ⟨type, curved⟩, ⟨manner, steep⟩, and ⟨direction, up⟩. For the systematic data analyses, the verbal data produced in a joint activity were also characterized by using this method since it successfully foregrounds the common properties of multimodal data, see Table 2 for semantic attribute scheme for verbal data.

Type Properties:
  Terms
    - ⟨term, peak⟩, ⟨term, something⟩
  Location
    - Frame of Reference Terms ("start point")
    - Haptic Ostensive Expressions

Absolute Properties:
  - ⟨value, 0⟩ for "it is 0"
  - ⟨count, 3 peaks⟩

Relative Properties:
  - ⟨size, small⟩ , ⟨manner, slowly⟩
  - ⟨direction, up⟩

Relations:
  - ⟨temporal relations, after the fall⟩
  - ⟨spatial relations, higher⟩

Others:
  - Interjections (hmm, ah…)
  - Affirmations/Negations

Table 2. Semantic attribute scheme

In addition to the attributes stated by Dale and Reiter (1995), we identified haptic ostensive expressions (*HOE*s). The haptic explorers produced *HOE*s that referred to the pointed locations, which are also accompanied by assistance request from the verbal assistant. Foster and colleagues (2008) define the *HOE* as a reference, which involves deictic reference to the referred object by manipulating it haptically. Since haptic explorer location is visible to verbal assistant during joint activity, haptic actions are useful to provide joint attention between E and A.

2 Experiment

2.1 Participants, Materials and Design

Thirty participants (fifteen pairs of sighted and blindfolded university students) participated in the experiment. The language of the experiment was Turkish, the native language of all participants. The experiment was conducted in single sessions and each session took approximately 1 hour (including warm-up & instruction sessions, exploration processes and post-exploration tasks). The sessions were audio/video recorded. Each participant pair was composed of a haptic explorer (*E*) and a verbal assistant (*A*). The participants were located in separate rooms so that they communicated through speakers without visual contact. During the experiment session, *E* explored the graph haptically and *A* was able to display the graph and the current location of *E*'s exploration, which was represented by an animated point marker on the visual graph presented at *A*'s screen. However, haptic pointing was possible only for E. The pairs explored informationally equivalent graphs, except for the difference in the modality of presentation (haptic and visual). Finally, *E* was instructed to explore the graph and ask for verbal assistance when needed by turning microphone on, whereas *A* was instructed to provide verbal assistance shortly and plainly, when requested by *E*. Before the experiment, a warm-up session was conducted to familiarize *E* with Phantom Omni® Haptic Device (Figure 1). After then, in the instruction session, the participants were informed that the graphs represented populations of bird species in a lagoon and also about post-exploration tasks detailed below. The graphs employed in this study were taken from a publicly available consensus report (*PRBO*, 2012). Each graph had a different pattern in terms of the number and polarity of curvature landmarks, length and direction of line segments. In the experiment session, each participant was presented five haptic line graphs in random order. Haptic graph exploration was performed by moving the stylus of the haptic device, which can be moved in all three spatial dimensions (with six degree-of-freedom). The haptic graph proper (i.e., the line of the line graph) was represented by engraved concavities on a horizontal plane; therefore haptic explorers perceived the line as deeper than the other regions of the haptic surface. The numerical labels were not represented. The participants did not have time limitation. After the experiment session, both participants (*E* and *A*) were asked independently to present single-sentence verbal descriptions of the graphs to a hypothetical audience. They also produced a sketch of the graph on paper. Two raters who are blind to the goals of the study scored the sketches for their similarity to the stimulus-graphs by using a 1 (least similar) to 5 (most similar) Likert Scale. The inter-rater reliability between the raters was assessed using a two-way mixed, consistency average-measures *ICC* (Intra-class correlation). The resulting *ICC* (=*.62*) was in the "good range" (Cicchetti, 1994).

3 Results

The participants produced 75 dialogues (5 stimuli x 15 pairs). The data from two pairs were excluded since they did not follow the instructions. The remaining 65 dialogues were included into the analysis. The average length of a dialog was 103 seconds (*SD*=62 sec.). The results of this experiment, which focus on the role of taking initiative for assistance, were reported elsewhere (Alaçam et. al. 2014). In the present study, we focus on the semantic representation method and the production of haptic ostensive expressions during joint activity. Each utterance in the dialogues was transcribed and time-coded. The transcriptions were then annotated by the semantic attribute scheme presented in Table 2. The term "utterance" refers to speech parts produced coherently and individually by each participant. We classified the utterances into three categories; (i) Request-Response Pairs, (ii) Alerts initiated by *A* (but do not require response from *E*) and (iii) think-aloud sentences. In total, 1214 individual utterances were produced by the participants. 449 of them were initiated by the haptic explorers to communicate with their partners, 402 of them were produced by the verbal assistants as a reply to *E*. Those two types comprise 70.1% of all utterances. 65 utterances (5.35%) were initiated by *A*s. Utterances that were initiated by *A*s, without a request from *E* were mostly the utterances that alerted *E* when s/he reached to a start point or an end point. Although *E*s were not instructed to use the think-aloud protocol, self-talking during haptic exploration was observed in 10 of 13 haptic explorers. Those think-aloud sentences (i.e. the sentences without a communication goal with the partner since the explorers did not turn on microphone during self-talking) constituted 24.5% of all utterances (*N*=298). In this paper we focused on the communicative utterances, therefore we restricted our analysis to "Request-Response Pairs" and "Alerts" excluding "Think-aloud" sentences. The results pointed out that the most frequently observed assistance content was about information for positioning, such as being on a start point or end point, on the frame, or being inside or outside of the line. 72.4% of the utterances (341 utterances in total - 46 of them initiated by *A*) addressed this type of information.

*E*s showed a tendency to request assistance by directing "Yes/No Questions or Statements" to *A*s (*N*=418) instead of using open-ended questions (*N*=7). *A*'s contributions to the dialogue can be also classified as follows: (1) instructional, *N*=69 (i.e. navigational, such as 'go downward from there'), or (2) descriptional utterances, *N*=386. Descriptional utterances included, (2a) confirmative assistance, *N*= 342 (confirming the information which haptic explorer has already access), and (2b) additional assistance, *N*=44 (introducing new property or updating the value of already stated property). Below we present sample request-response pairs, which introduced new information or updated the value of the already introduced attribute.

- *E*: Is this the start point? *A*: Yes, it is also the origin (*A* updates ⟨type, start point⟩ as ⟨type, origin⟩ that emphasizes 2D frame of reference, and that implicitly carries over the value for the starting point)
- *E*: no request. *A*: You are at the first curve; ⟨type, curve⟩, ⟨relation, order, first⟩ (both type and relation attributes were introduced to the dialogue)

The non-parametric correlation analyses using Kendall's tau showed positive correlation between the existence of attribute update in the dialogue and higher sketching scores (N=62, τ=.46, p=<.01). Moreover, the number of attribute updates is positively correlated with higher sketching scores (N=62, τ=.45, p=<.01). As an illustration, consider one of the dialogues between *E* and *A*: *E* asked a question ("*Is this going perpendicular?*") to *A* by pointing "*ep1-sp1*" segment of the graph presented in Figure 3. As stated in Table 1, this shape segment can be labeled with ⟨type, curved⟩, ⟨manner, steep⟩, ⟨direction, up⟩ attributes. In his question, *E* addresses both manner and direction attributes. However, the word for "perpendicular" in Turkish can be used to refer to both being perpendicular and steep. Here *A*´s response ("*There is a slight slope*") updates *E*´s information and it also clarifies possible misunderstanding, since in statistical graphs in time domain, perpendicular lines are not allowed. The resulting request-response pair covers all attribute pairs for the particular graph shape (the region which *E* needs assistance) and the sketch was rated with 4.5 in average (in 1to5 Likert Scale). The parameters (Dale and Reiter, 1995) (i) the number of attributes that are available to be used in a referring expression and (ii) the number of attributes mentioned in the final expressions seem as a useful indicator to evaluate the successful communication.

Additionally, verbal assistants' expressions that referred to a point or a region on the graph, namely type property, were mostly graph-domain terms (such as "curve", "peak" etc.). On the other hand, haptic explorers showed a tendency to use simpler expressions such as "something", "hill", "elevation". This indicated that haptic explorers had difficulty to access graph-domain vocabulary to name the regions or the shape, so that they choose alternative ways to name it (including use of onomatopoeic words such as "hop hop").

The haptic ostensive actions and expressions performed to catch the attention of the assistant do not directly contribute to conceptualizing the graph shape; still their communicative role in the dialogues is important. 20.4% (*N*=247) of all the communicative utterances contained *HOE* that enhanced the reference resolution, therefore shorter descriptions could be produced instead of long descriptions. The analysis of verbal data revealed two major subcategories of *HOE*s: (i) Demonstrative Pronouns (*DP*s) such as "This/Here" or "like this"), and (ii) temporal pointings (*TP*s) such as "Now". Table 3 illustrates the frequency values for each *HOE* category. Non-parametric Wilcoxon Signed-Rank tests were conducted to investigate the use of different *HOE* types. The results showed that the haptic explorers produced more *DP*s (z=-4.88, p=<.001) and *TP*s (z=-3.75, p=<.001) than the assistants produced. While there is no significant difference in the number of *DP*s and *TP*s produced by *E*s (z=-.50, p=>.05), *A*s preferred to use *TP*s rather than *DP*s. Only a few instances (*N*=5) of *DP*s uttered by *E* was responded by *A*'s use of *DP*s. The instances that illustrate *A*´s responding to *E* by using different *HOE* category than the one used by *E* were not observed at all.

	Only by *E*	Only by *A*	Both *E & A*
Demonstrative Pronoun-*DP*	99	6	5
Temporal Pointing-*TP*	67	27	19

Table 3. The number of *HOE*s for each category

We performed a further analysis on salient graph parts by focusing on in which area of the graph the participants preferred to use one of the two *HOE* categories (demonstratives and temporal pointing) for referring. For this, the accompanying content (location being referred to) were classified into three groups, (i) reference to start points and end points, (ii) reference to intermediate points or regions on the graph and (iii) reference to frame (such as being on the frame, or being outside of the line). The results of the analysis showed a significant association between the referred location and the *HOE* preference, $X^2(2)=38.2$, $p<.001$. The results (the standard residuals for each combination) indicated that when the participants referred to a start/end point of the graph line, they used *DP*s (N=48, z=-.6) and *TP*s (N=48, z=-.7). However, for referring to any particular point or any region on the graph, they preferred *DP*s (N=59, z=2.8) rather than *TP*s (N=16, z=-3.1). Moreover, when they mentioned about the events related to the reference frame, they preferred *TP*s (N=29, z=3.3) rather than DPs (N=6, z=-3, *all p values are smaller than .05*). However no main association was found between *HOE* types (*DP*s or *TP*s) and whether the referred region is a point or area. This indicates that both specific points (i.e. landmarks) and broader regions (i.e. line segments) haptically highlighted by *E* were accompanied by any of *HOE* types; however the position of the point or region on the graph (i.e. at the beginning or at the intermediate region on the line) has effect on which *HOE* type is preferred.

4 Discussion

In an experimental setting, which employed a joint-activity framework, pairs of participants (haptic explorers and verbal assistants) explored the graphs and they exchanged verbal information when necessary. Following Dale and Reiter (1995), we categorized graph shapes (segments/landmarks) and verbal data as attribute pairs such as ⟨type, maximum⟩. When *E* needs assistance about a segment, or global shape, her/his question was modeled as a specification of the choices of some of the attributes. As a response to the request for assistance, the description of *E* may be complete, lacking or partially or completely inaccurate. In order to have successful communication, verbal assistant should provide lacking information or correct the incorrect interpretation to complete the coverage of attributes in “target set” of attributes. Within this framework then, we assume that successful communication is achieved when *E* requests assistance (initiated by haptic explorer w.r.t. his needs to avoid over-assistance) and *A* updates the attribute pairs or introduces new attributes. Moreover, since *E* already has access to basic spatial properties, a useful solution would be to provide information with graph-domain terms, and relative terms (since absolute terms are difficult to implement), as well as relational terms that emphasize size and manner gradually (w.r.t. haptic explorer´s needs and current knowledge). The results of the experiment also showed that *A*’s role in *E*´s comprehension is critical. First, *A* has a more complete mental representation of the graph starting from the onset of haptic exploration due to spontaneous visual exposure to both global and local information on the graph. Their guidance on salient points with additional attributes or their aligning the instructions w.r.t haptic explorer´s current understanding of the graph enhances the comprehension of *E*. Moreover, the verbal assistants introduced more graph domain oriented concepts to dialogues, while haptic explorers tended to use simpler daily terms or even onomatopoeic words. This information is important when forming attribute set for graph shapes.

Our focus was to investigate the content that needs additional assistance but our results also pointed out the information that can be provided more effectively by a different modality than verbal modality. The research by Moll and Sallnäs (2009) and Huang et al (2012) suggest audio-haptic guidance for visually impaired people to enhance navigational guidance in virtual environments so that the participants focus on communication at a higher level. Their results indicated that "by using haptic guiding one can communicate information about direction that does not need to be verbalized" (Moll and Sallnäs, 2009, p.9) and "sound provides information that otherwise has to be conveyed through verbal guidance and communication" (Huang et al., 2012, p.265). Considering that 72.4% of the utterances in our experiment contained information about positioning (being on the start point, or on the line etc.), providing this information to the explorer seems crucial for the assistive system; however delivering this information verbally would yield continuously speaking assistance, therefore sonification can be a good candidate to carry this message. Additionally, haptic exploration allows haptic ostensive actions that highlight the attended location. The location attribute has different characteristics than other attribute pairs. It grounds joint attention between partners by pointing where the assistance is needed, then other attributes provide additional information about what the graph shape means. As for *HOE*s, the type of referring expressions (demonstrative pronouns or temporal pointing) seems affected by the referred location (start/end points, intermediate regions or graph frame). The results also indicated that the explorers produce significantly more *HOE*s during joint activity compared to the verbal assistants. In the collaborative activity settings that allow both users (the human explorer/learner and human or robot assistant) to manipulate the environment haptically (Foster et al., 2008; Moll and Sallnäs, 2009), the assistants' haptic ostensive actions have salient communicative function. However, in our assistance setting, only haptic explorers have active role in the haptic exploration. Even after requesting assistance from A regarding specific point or region by pointing with *HOE*, *E* may still continue to explore. Therefore verbal assistants tend to omit uttering *HOE* and when necessary, they use temporal indicators to relate a previously mentioned expression to currently explored region. This preference of verbal assistants may be due to prevent explorers' incorrect reference resolution.

Finally, in addition to attribute-set approach of Dale and Reiter (1995), a more context sensitive version that implemented salience weights was proposed by Krahmer and Theune (2002). The comparative study between visual and haptic perception of graphs indicated that haptic readers tend to overestimate small variations on the graph shape due to haptic salience induced by haptic friction and to underestimate smooth regions that can be useful for segmentation (Habel et. al, 2013). Choosing appropriate attribute value enhanced with salience weights for this kind of haptically problematic regions might overcome this problem in the implementation level.

5 Conclusion

Graphs are one of the efficient ways of visual communication to convey the highlights of data, however visual perception differs from haptic perception; therefore the highlighted piece of information in visual modality can be hidden when it is converted to haptic modality. Hence, investigation of differences in two modalities is necessary to detect and close the informational gap. The current study that explores on-line haptic graph comprehension in the presence of verbal assistance contributes our understanding about haptic graph comprehension by investigating dialogues between haptic explorer and verbal assistant as a collaborative activity.

Taking the Gricean Maxims into account in the generation of referring expressions (careful selection of the information provided in "attribute pairs", updating attributes gradually and being sure that at the end of the communication target attribute set is covered) seems useful in enhancing the conversational success of the communication (Grice, 1975; Dale, 1992; Dale & Reiter, 1995). In contrast to providing all likely information to the graph reader all together, the detection of what s/he wants to know at a particular time would yield a more effective design of the (learning) environment for the graph reader when we take into account his/her current position, previous haptic exploration movements and utterances (the referred locations and how these regions were referred), thus addressing adequacy, efficiency and sensitivity criteria. For this reason, semantic mapping needs to be accomplished in multimodal data. Following Dale and Reiter´s approach, we represented graph shapes and verbal data as attribute pairs in the present study. The empirical results revealed that a more successful communication was observed when the attributes used by haptic explorers were enriched by means of specific, graph-domain terminology. Accordingly, building up a multimodal system based upon this approach looks promising. Future work will address designing the generation of verbal assistance based on the experimental findings.

Acknowledgments

The study reported in this paper has been supported by DFG (German Science Foundation) in ITRG 1247 'Cross-modal Interaction in Natural and Artificial Cognitive Systems' (CINACS) and by METU BAP-08-11-2012-121 'The Study of Cognitive Processes in Multimodal Communication'.

References

Abu Doush, I., Pontelli, E., Simon, D., Son, T.C., & Ma, O. (2010). Multimodal Presentation of Two-Dimensional Charts: An Investigation Using Open Office XML and Microsoft Excel. *ACM Transactions on Accessible Computing, 3,* 8:1–8:50.

Acartürk, C., Alaçam, Ö., & Habel, C. (2014). Developing a Verbal Assistance System for Line Graph Comprehension. In A. Marcus (Ed.): *Design, User Experience and Usability* (DUXU/HCII 2014), Part II, (pp. 373–382). Berlin: Springer-Verlag.

Alaçam, Ö., Habel, C. & Acatürk, C. (2014). Verbally Assisted Haptic Graph Comprehension: The Role of Taking Initiative in a Joint Activity. To be published in the Proceedings from the 2st European Symposium on Multimodal Communication, University of Tartu, Estonia, August 6-8, 2014.

Cicchetti, D.V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment,* 6(4), 284–290.

Clark, H., & Haviland, S. (1977). Comprehension and the Given-New Contract. In: R. O. Freedle (ed.), *Discourse Production and Comprehension* (pp. 1–40). Erlbaum, Hillsdale, NJ.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared* cognition (pp. 127-149). American Psychological Association, Washington, DC.

Cohen, E., & Singh, M. (2007). Geometric Determinants of Shape Segmentation: Tests Using Segment Identification. *Vision Research, 47,* 2825-2840.

Dale, R. (1992). *Generating Referring Expressions: Constructing Descriptions in a Domain of Objects and Processes*. MIT Press, Cambridge, MA.

Dale, R., & Reiter, E. (1995). Computational Interpretations of the Gricean Maxims in the Generation of Referring Expressions. *Cognitive Science*, 19(2), 233-263.

Demir, S., Carberry, S., & McCoy, K.F. (2012). Summarizing Information Graphics Textually. *Computational Linguistics, 38,* 527–574.

Eschenbach, C., Habel, C., Kulik, L., & Leßmöllmann, A. (1998). Shape nouns and shape concepts: A geometry for ‚corner‘. In C. Freksa, C. Habel, & K. Wender (eds.), *Spatial Cognition*. (pp. 177–201). Springer, Heidelberg

Ferres, L., Lindgaard, G., Sumegi, L., & Tsuji, B. (2013). Evaluating a tool for improving accessibility to charts and graphs. *ACM Transactions on Computer-Human Interaction ,20(5),* 28:1–28:32.

Foster, M.E., Bard, E.G., Hill, R.L., Guhe, M., Oberlander, J., & Knoll, A. (2008). The Roles Of Haptic-Ostensive Referring Expressions in Cooperative, Task-based Human-Robot Dialogue. In *Proceedings of ohe 3rd ACM/IEEE International Conference on Human-Robot Interaction,* pp. 295-302. Amsterdam, March 12-15, 2008.

Fukumura, K., van Gompel, R., & Pickering, M. J. (2010). The Use of Visual Context During the Production of Referring Expressions. *Quarterly Journal of Experimental Psychology 63,* 1700–1715.

Garrod, S., & Pickering, M. J. (2004). Why is Conversation so Easy? *Trends in Cognitive Sciences, 8,* 8-11.

Grice, H.P. (1975). Logic and conversation. In P.Cole & J. Morgan (Eds.), *Syntax and Semantics: Vol 3, Speech acts* (pp.43-58). New York: Academic.

Habel, C., Alaçam, Ö., & Acartürk, C. (2013). Verbally assisted comprehension of haptic line-graphs: referring expressions in a collaborative activity. *In Proceedings of the CogSci 2013 Workshop on Production of Referring Expressions,* Berlin.

Hegarty, M. (2011). The Cognitive Science of Visual-spatial Displays: Implications for Design. *Topics in Cognitive Science, 3,* 446–474.

Hoffman, D. & Richards, W. (1984). Parts of recogni*tion. Cognition, 18,* 65–96.

Huang, Y. Y., Moll, J., Sallnäs, E. L., & Sundblad, Y. (2012). Auditory Feedback in Haptic Collaborative Interfaces. *International Journal of Human-Computer Studies, 70(4),* 257-270.

Koolen, R., Gatt, A., Goudbeek, M., & Krahmer, E. (2011). Factors Causing Overspecification in Definite Descriptions. *Journal of Pragmatics, 43,* 3231-3250.

Krahmer, E., & Theune, M. (2002). Efficient Context-sensitive Generation of Referring Expressions. In: K. van Deemter & R. Kibble (Eds.) *Information Sharing: Reference and Presupposition in Language Generation and Interpretation.* (pp. 223-264). CSLI Publications, Stanford.

Kerzel, M. & Habel, C. (2013). Event Recognition During Exploration of Hapic Virtual Environment Line-based Graphics. In T. Tenbrink, J. Stell, A. Galton & Z. Wood (eds.) *Spatial Information Theory, 11th International Conference, COSIT 2013.* (pp. 109–128). Berlin: Springer-Verlag.

Loomis, J., Klatzky, R., & Lederman, S. (1991). Similarity of Tactual and Visual Picture Recognition with Limited Field of View. *Perception,* 20, 167-177.

Moll, J., & Sallnäs, E. L. (2009). Communicative Functions of Haptic Feedback. In: M. E. Altinsoy,

U. Jekosch, & S. A. Brewster (Eds.), *Haptic and Audio Interaction Design*. (pp. 1-10). Springer, Berlin Heidelberg.

PRBO. Waterbird Census at Bolinas Lagoon, MarinCounty, CA. Public report by Wetlands Ecology Division, Point Reyes Bird Observatory (PRBO) Conservation Science. (2012) http://www.prbo.org/cms/366, retrieved on January 29, 2012.

Sebanz, N, Bekkering, H., & Knoblich, G. (2006). Joint Action: Bodies and Minds Moving Together. *Trends in Cognitive Sciences, 10*, 70-76.

Spanger, P., Yasuhara, M., Iida, R., Tokunaga, T., Terai, A., & Kuriyama, N. (2012). REX-J: Japanese Referring Expression Corpus of Situated Dialogs. *Language Resources and Evaluation, 46,* 461-491.

Wu, P., Carberry, S., Elzer, S., & Chester, D. (2010). Recognizing the Intended Message of Line Graphs. In: Goel, A.K., Jamnik, M., & Narayanan, N.H. (eds.) *Diagrammatic Representation and Inference*. (pp. 220–234). Springer, Heidelberg.

Yu, W., & Brewster, S.A. (2003). Evaluation of Multimodal Graphs for Blind People. *Journal of Universal Access in the Information Society 2,* 105-124.