A Vertical Timeline Visualization for the Exploratory Analysis of Dialogue Data

Paul Craig [1], Néna Roa-Seïler [1,2]
[1] Universidad Tecnológica de la Mixteca, [2] Edinburgh Napier University
{p.craig@mixteco.utm.mx, n.roa-seiler@mixteco.utm.mx}

Abstract

This paper presents a novel vertical timeline information-visualization technique developed to support the analysis of human-computer dialogue data. The technique uses combined linked views including distorted views to effectively communicate the timing of dialogue events while presenting text in such a manner that it is easily readable. A prototype has been implemented and tested to demonstrate the technique's effectiveness for supporting exploration and revealing previously unsuspected patterns.

Keywords--- **Information Visualization, Temporal Data, Spoken Language Dialogue Systems, Multiple Coordinated Views**

1. Introduction

A dialogue system is a computer system designed to converse with a human being. Spoken language dialogue systems (SLDSs) [1, 2] are dialogue systems in which spoken natural language plays an important part in the communication. The most common types of SLDS are task oriented systems designed to help the user towards some specific goal such as purchasing a train ticket or performing a simple bank transaction. These systems work with a limited vocabulary and are limited in the range of operations they are capable of performing. Non task oriented SLDSs are designed to be able to engage a human-being in a conversation that goes beyond the achievement of a simple task. These are becoming increasingly popular [2, 3] and present researchers with significant challenges over and above those posed by task-oriented systems due to the increased vocabulary and range of responses they require [4].

Broadly speaking there are three types of evaluation for spoken dialogue systems [5]. These focus on technical issues, usability of the system and customer evaluation. Factors important to the usability of a system are the semantics of the speech, the tone of voice and the timing. The timing of responses, overlaps in utterances and interruptions are considered particularly important for system usability [6-8]. For example, if a system appears to respond too slowly, it may appear inattentive or uninterested. If a system responds too quickly it may appear as sarcastic or as not giving enough thought to what is being said. Appropriate timing can also depend on the semantics of the conversation. For example, if a system is responding to serious news, long delays may be more appropriate. If the system is responding to informal questions relating to, for example, a cooking recipe, shorter pauses are likely to be more appropriate. In order to properly evaluate a system, testers need to take timing and speech text into account.

The raw-data produced by an SLDS experiment is the audio of a conversation between the system and a user. This is normally processed to generate a textual transcript. There are a number of ways to analyze this data. These include listening to the unprocessed audio recordings, reading transcripts of the audio and performing statistical analysis. The first of these options, listening to the unprocessed audio recordings, is perhaps the most effective for more focused analysis of small sections of dialogue but it becomes impractical for larger collections due to the time it takes to physically listen to each conversation [9]. Transcribed dialogues have the advantage that they can be read quickly (scanned) in order to locate and review sections of dialogue that may be more appropriate for the particular type of analysis being performed. The disadvantage of relying on transcribed dialogues is that plain text cannot communicate aspects related to the timing of the responses. These are key to the understanding of a spoken message [6-8] and proper analysis cannot be said to be done without considering them. Statistical methods have the potential to be able to account for more aspects of the data but have the disadvantage that they cannot decipher the semantics of words and phrases as effectively as a human being. Moreover the data often needs to be explored before an analyst can judge which statistical methods are appropriate.

2. Related work

A further option for the evaluation of SLDS data is to use software that supports exploration of the data using some visual representation. This allows analysts to explore the data and form casual hypothesis that can be confirmed later using statistical methods. Visualizations

1550-6037/12 $26.00 © 2012 IEEE
DOI 10.1109/IV.2012.22

of dialogue data can also provide analysts with a framework that allows them access the original audio or video data at a particular point for more focused analysis of shorter dialogue segments [9].

2.1 Timelines

In general, dialogue data is displayed in the form of a timeline [10, 11] with utterances (portions of dialogue) displayed as distinct entities in chronological order. The classic horizontal layout timeline representation of temporal data (see figure 1) displays events as colored labeled bars positioned according to a regular time-scale from left to right. The left-hand side of each event bar is aligned according to that event´s start time and the right hand side with its end time. Where events have no duration they are generally represented using a single tick. The vertical dimension is used to stack events and avoid labels overlapping. This particular type of timeline has been demonstrated with a number of applications including medical data [11, 12] (figure 1), research data [13], world history [14] and file system data [15].

While most timelines adhere to the classic form described above there are a number of variations worthy of note. One minor amendment is when callouts are used for text that can´t fit inside or around the bar that illustrates timing (e.g. [16] see figure 2). Other modifications avoid the use of bars or ticks to represent events altogether. For example, the themeRiver visualization [17] uses the metaphor of a river to represent the relative strength of themes in a document collection over time. Individual themes are represented as colored "currents" flowing within the river, narrowing or widening to indicate changes in theme strength at any point in time.

Timelines with time progressing on the vertical axis are relatively rare. Notable exceptions are the Google labs news timeline feature and the timeline view of user profiles in the social media site Facebook [18]. Google labs' news timeline presents news events in columns. Each column deals with a different time period (typically a month or a year) and events are ordered chronologically within their respective columns. The timeline view of user profiles in Facebook uses two columns with items ordered chronologically from bottom to top and sequential items alternating from the left to right column. Both of these layouts communicate the chronology of events rather than absolute timing. This makes it difficult to perceive patterns related to timing. Things like events clustered together, or periods of sparse activity, are hard to perceive using these types of view.

2.2 Visualizing dialogue data

The problem of displaying and time-aligning speech data is well known to conversation analysts [9] and a number of solutions have been devised. These range from more or less standard implementations of the classic horizontal layout timeline [9, 19] to solutions that

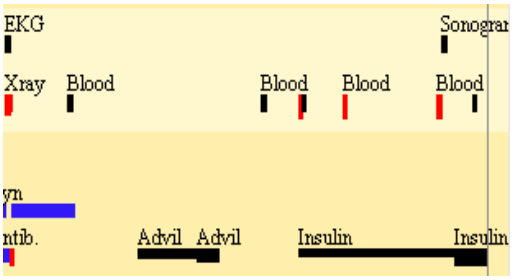

Figure 1: The classic horizontal layout timeline representation of temporal data [11]. Events as colored labeled bars positioned according to a regular time-scale from left to right.

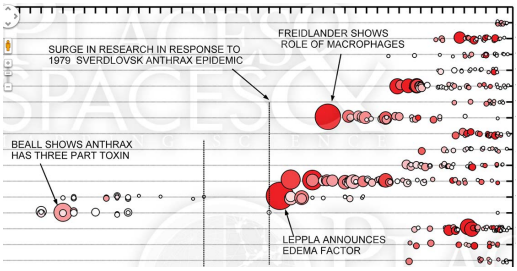

Figure 2: The classic horizontal layout timeline with callouts to accommodate the display of larger quantities of text [16].

Ava: He, he 'n Jo were like on the outs, yih know?
(0.7)
Ava: [So uh,
Bee: [They always are(hh)hhh

Figure 3: The timing of dialogue data represented symbolically [8] (numbers in brackets represent a delay of a number of seconds, squared brackets represent overlapping speech)

combine text with special symbols that communicate aspects of timing [7, 20]. All of these solutions have proven useful but they also have their limitations.

One proposed solution is to provide a horizontal layout timeline overview of utterance timings with textual content revealed when utterances are brushed [9]. Brushing to reveal utterance text has the advantage that it saves space allowing more timing data to be displayed effectively on the screen at the same time. Since the data is displayed using pre-attentive visual variables (color and displacement) it's easy to perceive patterns and gain a global perspective of the data. The problem with this approach is that it makes it difficult to read the text of the dialogue over a number of utterances. Any attempt to do this would involve moving the mouse over different consecutive utterance bars while continually readjusting to the different positions of detail-on-demand pop-ups.

Another solution uses classic horizontal layout timeline and nests textual information inside the bars that encode utterance timings [19]. Readability is also a problem for this type of approach since the user needs to readjust to accommodate for the different horizontal alignment of text when moving between utterances. This can have a significant effect on the users performance for the tasks of reading and scanning [21]. The same problem would be evident if text where to be contained in call out boxes horizontally aligned near the start of each bar (as is the layout mechanism of [16]).

The most legible representation of dialogue data is where timing data is represented symbolically together with the text of the dialogue [7, 20] (see figure 3). Here the use of symbols allows text to be displayed in more or less the standard left-aligned paragraph oriented format commonly found in books and periodicals. This allows the user to scan the text quickly in order to gain an overview of the semantic content or locate points of interest. The problem with these types of representation is that they require the additional cognitive step of decoding the symbolic representation. This makes it difficult for the user to perceive patterns or obtain a global perspective of the data since these types of tasks would involve the interpretation of larger quantities of symbolically presented data.

3. Dialogue Explorer

The design of the Dialogue Explorer application is based on the ideas that users should be able to benefit from the effectiveness of a graphical representation of timing data to communicate patterns, and that this can be combined with a natural left-aligned representation of the text content which is easy to read and scan. The requirements of a prospective users obtained during a series of interviews prior to the development of the first prototype largely concurred with these objectives. Although, the users gave an additional requirement which was that the application should be able to show the frequency of terms used in the dialogue.

The first version of the application was implemented as a paper prototype whose evaluation yielded nothing more than a few minor stylistic changes. This was followed by the development of a full software prototype. A screenshot of this prototype is shown in figure 4. The following sections describe various components of the display, explain how the user can interact with the interface and explain the rationale behind some of the design decisions.

3.1 Display

The display of Dialogue Explorer is divided into four separate component views arranged horizontally from left to right. These are the *search/term-frequency component*, the *timing overview*, *detail timing view* and the *distorted text detail view*.

The *search/term-frequency component* on the far left hand side of the interface allows the user to search for terms (words or phrases) or view term frequency. Results of searches are displayed with their frequency and ordered according to frequency. When no search is entered the component displays the most frequent terms in the dialogue. Clicking on a term in the results causes the term to be highlighted in all other views.

Immediately to the right of the *search/term-frequency component* is the *timing overview*. This component provides an overview of the entire data set using a regular scale and allows the user to navigate across time by clicking and dragging a rectangular box that represents the time range of the detail views.

To the right of the *timing overview* is the *timing detail view*. This uses a regular scale to display the timing of events over a limited time-range of the data with grid-lines at one second intervals. Each event in this view is linked by a curved line to the corresponding text in the *distorted text detail view* immediately to the right.

On the far right hand side of the interface the *distorted text detail view* displays the text for each event over the same time-range as *the timing detail view* and with the same horizontal grid-lines. In this view the scale is distorted, the size of text boxes is determined by the amount of text they contain (with a consistent font size that allows the text to be easily readable) and timing information between utterances is communicated using the horizontal grid-lines between text boxes.

3.2 Interaction

There are a number of ways to interact with the Dialogue Explorer application. These are; searching by keyword, navigating by clicking on the *timing overview*, navigating by dragging and zooming on or out on the time-scale.

The *search/term-frequency component* allows the user to search for terms by typing. Once a term is selected, instances are highlighted in each of the other views and the interface focuses onto the first instance. Focusing on a term instance involves the detail views scrolling until the utterance containing the term is positioned in the middle of the display. As the detail views scroll, the rectangular box that represents the time range of the detail views in the dialogue timing overview moves at the same time. Here the animated view transformation allows the user to keep track of their position in the data, helping to improve their perspective of the overall temporal context.

The user can also focus on different utterances by clicking on their representations in the dialogue timing overview or detail views. Clicking and dragging on the overview detail box of either of the detail views also allows the user to move through time. Here the interface operates using kinetic flick and break scrolling [22] so as to increase responsiveness and reduce the effort required of the user during the interaction. Another way to move through time is to use the play, stop and rewind buttons in the toolbar. These allow the user to play the audio of the dialogue. Here, the point in time from which the audio starts can be changed by clicking to move a red bar

that passes through the overview and detail views. This red bar also moves to mark the current position during audio playback.

A slightly unusual artifact of the visualization is that the length of the detail view changes as it is moved. This is due to the fact that the *distorted detail view* has an irregular time scale. When the detail view covers periods of time with larger amounts of speech, its duration is shorter so that all the speech can fit into the fixed display space. Conversely, when the detail view covers periods of time with smaller amounts of speech, the view can cover longer periods of time. Changes in the duration of the *distorted detail view* can be observed in the timing detail view (which always covers the same time-period as the *distorted detail view*) where the spacing between grid-lines becomes more or less. It can also be seen in the overview where the detail view rectangle becomes larger or smaller as the focuses on different time-points.

Turning the mouse wheel allows the user to zoom in or out of the time-scale. Here, zooming adjusts the amount of space between gridlines in the distorted detail view. The gridlines always mark a one second interval. Providing more space between gridlines reduces the amount of time that can be displayed in the detail view to have the effect of zooming in on the time-scale. Providing less space between gridlines has the opposite effect. Zooming in and out of the time-scale also affects the legibility of the text and the visibility of the timing. When the display is zoomed-out, with a smaller amount of space between grid-lines, it is easier to read the text since there is less vertical space between lines of text. A smaller amount of space between grid-lines also makes it marginally more difficult to relate vertical displacement to temporal displacement. Hence, there is a tradeoff between communicating temporal information (when the interface is zoomed-in) and communicating more textual information more legibly (zoomed-out).

3.3 Design Rationale

The principle goal of the design of the Dialogue Explorer application is to support the analysis of dialogue data through the combination of an effective graphical representation for timing data and an easily readable textual representation of speech.

To ensure the readability of speech we use a vertical timeline layout in preference to a standard horizontal layout. This allows the text to be displayed in a single long column that can be read without the user having to readjust to different horizontal alignments as they move their focus from the text of one utterance that of another (a similar rotation of a standard visualization layout to improve the readability of text has previously been applied for hierarchical data in the Concept Relationship Editor application [23]). The text of the dialogue is contained in the *distorted text detail view* of the application (see figure 4d). This assigns enough vertical space as is required for each utterance to be easily

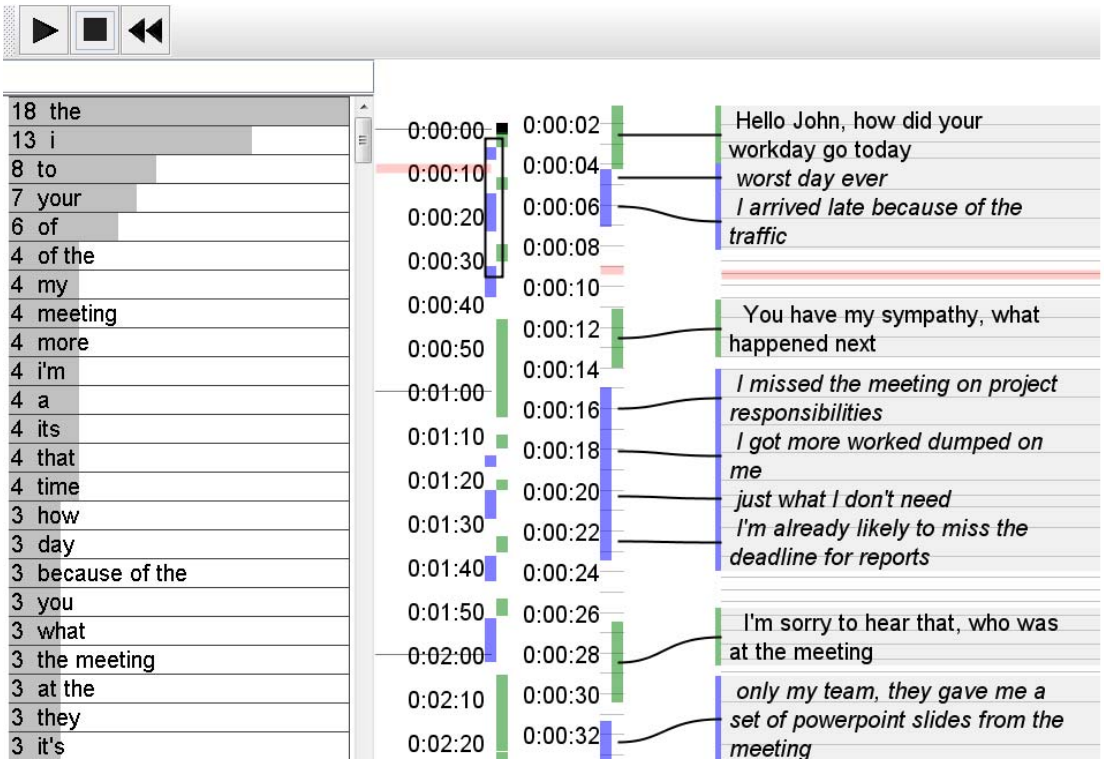

Figure 4: The Dialogue Explorer prototype interface. a) *Search/term-frequency component*, b) *timing overview* c) *timing detail view* and d) *distorted text detail view*.

readable and uses a regular time-scale for the pauses between utterances. This allows for the length of pauses to be communicated alongside the text of the utterances themselves.

Both the *timing overview* (figure 4 b) and *timing detail view* (figure 4 c) provide the user with an effective graphical representation for timing data. The use of a separate overview and detail view allows the user to view their selection within the context of the timing of the data as a whole. This can be classified as a Focus + Context [24] view of the data, which is preferable to alternate overview and detail views since parts of a dialogue are heavily influenced by both the timing within the dialogue as a whole and the content of previous parts of the dialogue.

Other design decisions are the use of color to encode the identity of participants, the use of curved lines to link utterance timings and text and the use of animation to smooth view transitions. Color is chosen to encode the identity of participants since both spatial dimensions are already used (y for time and x to organize views) and color is the next most effective pre-attentive visual variable [25]. The colors used (purple/dark-blue and green) are selected using guidelines developed to maximize accessibility for the visually impaired [26]. To further improve accessibility color is never used on its own encode information. In the dialogue timing overview and detail timing view horizontal displacement is used together with color to encode the identity of participants. In the distorted detail text view we use italics. Curved lines are used to link timings in the detail timing view to utterance text in the distorted detail text view. This design decision was based on the gestalt law of connection [27, 28] which tells us that the physically connected items appear as a unit. This is appropriate since the timing information and text refer to the same utterance. Animation [23, 29, 30] is used to smooth the transition between subsequent detail views when the user interacts to move through time. This also helps enforce the relationship between views which are coordinated as they move.

4. Evaluation

The Dialogue Explorer prototype was evaluated with dialogue data generated for the analysis of the Embodied Conversational Agent (ECA) [31] Samuella as part of the Companions project [32]. This involved two experienced conversational agent researchers using the application to perform exploratory analysis of the data with the objectives of gaining a general overview of the data and revealing previously unsuspected patterns.

The results of the evaluation where generally positive. After a short time (less than a minute) familiarizing themselves with the interface, the researchers were able to explore the data with relative ease. They were able to navigate, relate the contents of each view and move their focus between views effortlessly. Patterns in the data started to become apparent after a very short period of time (less than 4 minutes). Firstly, the researchers became aware of a break down in the pattern of turn-taking that one might normally expect in standard person-to-person dialogue [7]. As Samuella sometimes took a longer time to formulate a response the user would appear to become impatient and deliver shorter utterances more quickly over time. This pattern could be perceived in the overview and inspected more closely in the *timing detail view*. Reading text in the text *distorted detail view* and listening to the audio help confirm this theory. Occasionally, however, the pattern would change and the user would appear to take a long time to respond to Samuella. Here, reading the text in the *distorted text detail view* would normally give the reason. This was normally that Samuella had said something confusing and the user needed to take time to think about their response. The researchers were also able to identify some areas where the interface could be improved. They felt that the interface should support collaborative working [33] and that they should be able to annotate the dialogue. They wanted their annotations to be editable, easily accessed and visualized together with the dialogue. They also wanted to be able to view video data of results within the application. With regard to scalability, we found that it started to become difficult to navigate using the overview with around 100 utterances. This equates to around 4.5 hours of conversation and is far beyond the duration of conversations that are generally analyzed in spoken language research.

Overall the researchers where very encouraged by early results and are planning to use the application for the analysis of the results of two upcoming experiments. The first of these will record dialogue between a human and a human-controlled wizard-of-oz interface [34] in order to investigate user reaction to an ECA interface without timing and speech errors. The next experiment will look at communication patterns of children with severe motor disabilities.

Conclusions

We have developed a novel visualization technique for the exploratory analysis of dialogue data. This uses a vertical layout timeline with coordinated linked views that combine legibility with an effective representation of timing data. A prototype has been developed and tested to demonstrate the technique's effectiveness for supporting exploration and revealing previously unsuspected patterns. Overall results of the evaluation were positive and we plan to further develop the initial prototype to support data annotation and the display of video data.

References

[1] G. Tur and R. D. Mori, Spoken Language Understanding: Systems for Extracting Semantic Information from Speech: John Wiley & Sons, 2011.

[2] D. Suendermann, Advances in Commercial Deployment of Spoken Dialog Systems: Springer, 2011.

[3] L. Dybkjær, et al., Evaluation of Text and Speech Systems vol. 38: Springer, 2007.

[4] J. Edlund, et al., "Two faces of spoken dialogue systems," presented at the INTERSPEECH 2006, ICSLP SATELLITE WORKSHOP DIALOGUE ON DIALOGUES: MULTIDISCIPLINARY EVALUATION OF ADVANCED SPEECH-BASED INTERACTIVE SYSTEMS, Pitsberg, PA., 2006.

[5] L. Dybkjæra, et al., "Evaluation and usability of multimodal spoken language dialogue systems," Speech Communication, vol. 43, pp. 33-54, 2004.

[6] S. E. Brennan, "Processes that shape conversation and their implications for computational linguistics," presented at the Proceedings of the 38th Annual Meeting on Association for Computational Linguistics, Hong Kong, 2000.

[7] H. Sacks, et al., "A Simplest Systematics for the Organization of Turn-Taking for Conversation," Language, vol. 50, pp. 696-735, 1974.

[8] N. Campbell, "On the use of nonverbal speech sounds in human communication," presented at the Proceedings of the 2007 COST action 2102 international conference on Verbal and nonverbal communication behaviours, Vietri sul Mare, Italy, 2007.

[9] N. Campbell, "Tools and Resources for Visualising Conversational-Speech Interaction Multimodal Corpora," vol. 5509, pp. 176-188, 2009.

[10] [S. B. Cousins and M. G. Kahn, "The visual display of temporal information," Artificial Intelligence in Medicine, vol. 3, pp. 341-357, 1991.

[11] C. Plaisant, et al., "LifeLines: Visualizing Personal Histories," in ACM CHI'96 Conference, 1996, pp. 221-227.

[12] R. Bade, et al., "Connecting time-oriented data and information to a coherent interactive visualization," presented at the Proceedings of the SIGCHI conference on Human factors in computing systems, Vienna, Austria, 2004.

[13] S. A. Morris, et al., "Time line visualization of research fronts," Journal of the American Society for Information Science and Technology, vol. 54, pp. 413-422, 2003.

[14] R. B. Allen and S. Nalluru, "Exploring History with Narrative Timelines," presented at the Proceedings of the Symposium on Human Interface 2009 on ConferenceUniversal Access in Human-Computer Interaction. Part I: Held as Part of HCI International 2009, San Diego, CA, 2009.

[15] J. Rekimoto, "Time-machine computing: a time-centric approach for the information environment," presented at the Proceedings of the 12th annual ACM symposium on User interface software and technology, Asheville, North Carolina, United States, 1999.

[16] S. A. Morris and K. W. Boyack, "Visualizing 60 Years of Anthrax Research," presented at the 10th International Conference of the International Society for Scientometrics and Informetrics, Stockholm, 2005.

[17] S. Havre, et al., "ThemeRiver: Visualizing Thematic Changes in Large Document Collections," IEEE Transactions on Visualization and Computer Graphics, vol. 8, pp. 9-20, 2002.

[18] (2012, March). Facebook. Available: www.facebook.com

[19] T. v. d. Malsburg, et al., "TELIDA: a package for manipulation and visualization of timed linguistic data," presented at the Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue, London, United Kingdom, 2009.

[20] J. Local, "Phonetic Detail and the Organisation of Talk-in-Interaction," presented at the XVIth International Congress of Phonetic Sciences., Saarbruecken,Germany, 2007.

[21] J. Ling and P. van Schaik, "The influence of line spacing and text alignment on visual search of web pages," Displays, vol. 28, pp. 60-67, 2007.

[22] M. Baglioni, et al., "Flick-and-brake: finger control over inertial/sustained scroll motion," presented at the Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems, Vancouver, BC, Canada, 2011.

[23] P. Craig and J. Kennedy, "Concept Relationship Editor: A visual interface to support the assertion of synonymy relationships between taxonomic classifications," presented at the Visualization and Data Analysis 2008, San Jose, CA, 2008.

[24] J. Lamping, et al., "A focus + context technique based on hyperbolic geometry for visualizing large hierarchies," in ACM CHI '95, Denver, Colorado, USA, 1995, pp. 401-408.

[25] A. Treisman, "Preattentive processing in vision," Comput. Vision Graph. Image Process., vol. 31, pp. 156-177, 1985.

[26] C. Brewer, "Guidelines for use of the perceptual dimensions of color for mapping and visualization," in Color hard copy and graphic arts III, Proceedings of the international society for optical engineering (SPIE), San Jose, February 2004., 1994, pp. 54-63.

[27] M. Wertheimer, "Laws of Organization in Perceptual Forms," in A source book of Gestalt pyschology, W. Ellis, Ed., ed London: Routledge & Kegan Paul, 1938, pp. 71-88.

[28] P. Rookes and J. Willson, Perception: theory, development, and organisation. London, UK: Routledge, 2000.

[29] M. Graham and J. Kennedy, "Combining linking & focusing techniques for a multiple hierarchy visualisation," in IV 2001, London, UK, 2001, pp. 425-432.

[30] P. Craig, et al., "Animated Interval Scatter-plot Views for the Exploratory Analysis of Large Scale Microarray Time-course Data," Information Visualization, vol. 4, pp. 149-163, Sept. 2005.

[31] J. Cassell, et al., "Embodiment in conversational interfaces: Rea," presented at the Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit, Pittsburgh, Pennsylvania, United States, 1999.

[32] M. Cavazza, et al., "How was your day?' An Affective Companion ECA Prototype," presented at the Proceedings of the SIGDIAL 2010 Conference, Tokyo, Japan, 2010.

[33] P. Craig, et al., "Pattern browsing and query adjustment for the exploratory analysis and cooperative visualisation of microarray time-course data," presented at the Proceedings of the 7th international conference on Cooperative design, visualization, and engineering, Calvia, Mallorca, Spain, 2010.

[34] J. Edlund, et al., "Towards human-like spoken dialogue systems," Speech Communication, vol. 50, pp. 630-645, 2008.