Visualizing Spoken Interaction

Jens Allwood
Christopher Ahlberg

Abstract: This paper introduces and explores the use of IVEE, a technique for visualization and dynamic database queries applied to a database of transcriptions from various types of linguistic interaction. The use of IVEE is illustrated by presentation of data on vocabulary richness, verbal dominance, hesitance, uncertainty and liviliness,

1. Visualizing and exploring complex datasets with dynamic queries

1.1 Introduction - dynamic queries

Dynamic queries is a concept for information exploration and database querying based on visualization. A dynamic query is defined as
“a user controlled animated visualization of the results from a database query.”.

Users manipulate query widgets of various kinds to construct database queries, such as rangesliders, toggles, and alphasliders (Ahlberg, 1994, Figure 1), to incrementally update a visualization of the current query result. The visualization has an update rate of less than l00 ms, allowing for users immediate control of the result set.

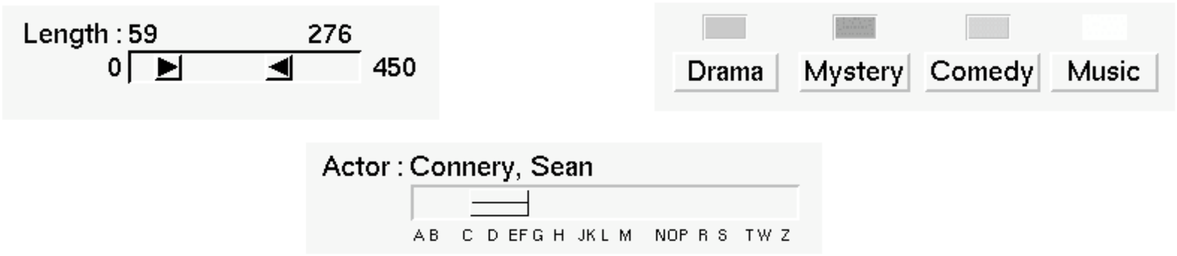


Figure 1. Query widgets in dynamic queries.

The benefit of static visualizations of data sets has long been known, visualizations of for example demographic, scientific, and economic data sets are commonplace today (Tufte, 1983). Static visualizations can be created with tools such as Excel, Lotus 1-2-3, and various statistical packages such as SAS and Statview. User controlled animated visualizations with a screen update rate of less than l00 ms have the additional benefit of allowing for easy exploration of relations between individual query result sets.

Users might increase the value of a query parameter to detect how the increase affects the size of the result set. The relevance of the result set can be judged from for example the color of the elements in a visualization, or from where the result set shows up in the visualization (indicating for example geographic proximity or some other high level semantic property). Users might also manipulate query widgets to explore trends and patterns in data, or to detect anomalies.

Visual representations of databases also allows for presentation of many more database elements in single screens. This in turn, allows for overviews in a way which is not possible with traditional textbased data exploration systems. A visualization might present thousands of elements simultaneously, while a text based system is limited to the magnitude of tens of elements.

The cognitive load while performing information exploration tasks is usually high. Finding patterns, exploring trends, and detecting anomalies is cognitively difficult. Allowing users to incrementally control the animated display of query result sets in visualizations, minimizes the mental effort needed to grasp relations between queries and query results and relations between individual query results. A lot of the benefits from animated visualizations arise from the fact that we can perform many perceptual tasks such as detection of trends, patterns, and anomalies in pictures with very little conscious mental effort. Not until recently has the use of these our remarkable perceptual abilities found its way into information retrieval systems (Ahlberg & Shneiderman, 1994; Robertson, Card & Mackinlay, l993).

We have built a number of interesting prototype dynamic queries applications for experimental purposes, such as a dynamic periodic table Ahlberg et.al, 1992 and a dynamic homefinder Williamson & Shneiderman, 1992. The prototypes used domain specific visualizations such as the periodic table of elements and geographic maps.

1.2 IVEE: An Information Visualization & Exploration Environment

The Information Visualization and Exploration Environment (IVEE) (Ahlberg & Wistrand, 1995) is an attempt at a system for creating visualization environments for nearly any database. IVEE can automatically create dynamic queries environments holding query widgets and visualizations. IVEE imports a relation with named attributes, given on a straight forward text format. The data in the relation is examined and classified into datatypes (integers, reals, and strings) and size (number of different values held in the attribute). Based on this examination, IVEE selects query widgets for each attribute. Existing widgets are: rangesliders, alphasliders 3, and toggles (see figure above).

2.3 Query widgets in IVEE

Rangesliders are useful for selecting a ranges criteria for integer attributes, or other attributes with an order relation, and where it is meaningful to specify ranges. Increasing or decreasing the range of an attribute allows for powerful exploration of trends and anomalies.

The alphaslider is a widget for selecting items from long lists of for example strings. The rationale of the alphaslider is its small size, i.e., it allows for selection from lists of thousands of elements in a small screen area. The alphaslider can not only be used for selection of specific strings, it is also very useful for browsing categorical variables while observing the query result set in a visualization. Finally IVEE utilizes so called toggles when only a few alternatives exist for an attribute and these alternatives should be presented on the screen explicitly.

2.4 Visualizations in IVEE

IVEE offers a number of visualizations. The basic visualization is an interactive scatterplot, called a starfield - see Figure 1 for a starfield visualization. Two ordinal variables from a database relation are chosen as the axes in a scatterplot. Users can perform a number of operations on the starfield: zooming, panning, filtering, and selection of details-on-demand. IVEE allows for other visualizations as well, such as geographic maps and other spatial visualizations, and cluster visualizations.

The starfield is simple, yet powerful as it allows for display of many database elements simultaneously. Each element is represented by a small square, which can be coded by color, brightness, shape, size, etc. The spatial location of an element also effectively encodes two properties of an element. The starfield can be used for quick judgements of result relevance, exploration of trends, detection of anomalies, etc.

An obvious advantage of the starfield compared to many other databases visualizations is that it is very easy to create and works for many databases. The only criteria is that two ordinal variables that are meaningful to visualize should exist in the database to be queried.

3. Data and Measures

The interactive visualization methods described above have been tried on a database of spoken interaction consisting of 171 transcriptions involving 120 speakers (59 men and 61 women). Each activity has been subsectioned into between 1 and 26 subsections. There are a total of 302 subsections. The transcriptions contain 26,778 utterances and 216,240 words.

The transcriptions are transcribed in "Modified standard orthography" cf Allwood & Hagman 1994 - a format of transcription for Swedish which is identical to standard orthography except that words which have several pronunciations in spoken language are given separate representations, eg written language *jag* (I) can be written *jag* or *ja* depending on pronunciation.

The transcriptions otherwise indicate speakers, utterances, word and subsections due to change of topic or activity. Over and above standard orthography the transcriptions contain indications of pauses, overlaps between speakers, stressed words or segments of words, inaudible stretches of speech as well as comments on gestures or non-linguistic activity when this is necessary for comprehension of the interaction.

On the basis of the data directly available, a number of measures derived from this data can be defined cf Allwood & Hagman, 1994. Below we will exemplify how dynamic queries provide a new way of exploring the data with these measures.

The measures we will consider are:

1) Four measures of **vocabulary richness** (Van Hout, cf. Broeder, Extra and Van Hout, 1986)

   - **Type - token ratio** = types/tokens
   - **Herdan** = log (types)/log(tokens)
   - **Guiraud** = $\text{types} / \sqrt{\text{tokens}}$
   - **Theoretical Vocabulary** = The expected number of types for a transcription reduced to a given number of tokens

2) We will also consider the following complex measures defined in Allwood & Hagman 1994.

   - **Verbal dominance** = % of tokens/expected % tokens

This measure takes the actual % produced words for a certain speaker and divides it by the % words speakers would have if all speakers had the same share. For 5 speakers this would mean that each speaker spoke 20% of the total produced words. This makes the measure comparable across transcriptions with different numbers of speakers.

- **Hesitance** = (pauses/tokens) + % OCM

  This measure correlates hesitance with share of pausing and share of OCM. OCM abbreviates Own Communication Management and includes hesitation morphemes and other devices for change and management of one's own speech

- **Uncertainty** = % OCM + % epistemic tokens

  Uncertainty is by this measure related to share of OCM and share of epistemic tokens (primarily verbs and adverbs with epistemic attitudinal content, eg think, believe, perhaps, maybe, probably etc).

- **Liveliness.** = 2 x % overlapped words +
  % overlaps utterances +
  2 x % stressed tokens
  2 x (pauses/tokens) - MLU (mean length per utterance)

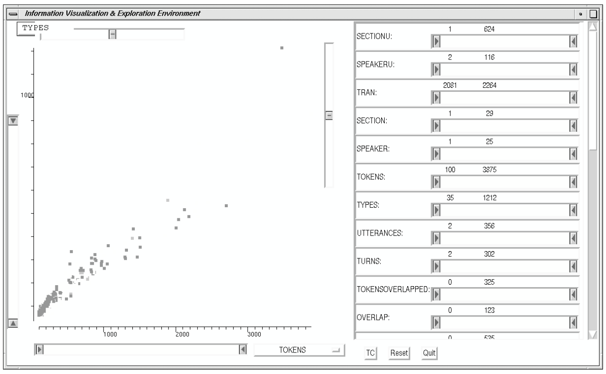
This measure views overlaps and stress as factors adding to liveliness while pauses and long MLU are seen as diminishing liveliness.

4. Some examples of spoken interaction visualized

4.1 Vocabulary richness

A desideratum among researchers who develop measures of vocabulary richness is that the measure should be independent of text length and sampling contingencies. The measure should give a fairly stable result independently of number of tokens.

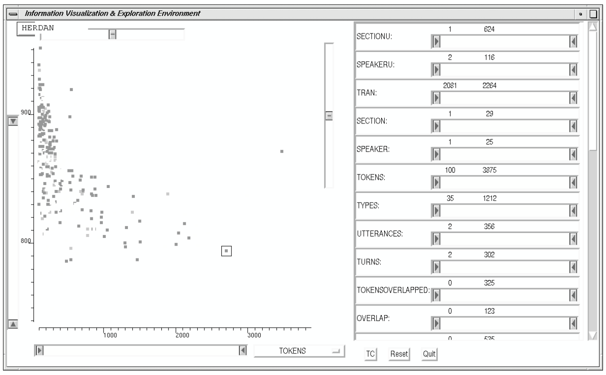
The IVEE visualizer now gives us a nice and direct way of exploring the four measures defined above, with regard to this desideratum. Compare the four figures below for the four measures.



SECTIONU: 1 624
SPEAKERU: 2 116
TRAN: 2081 2264
SECTION: 1 29
SPEAKER: 1 25
TOKENS: 100 3875
TYPES: 35 1212
UTTERANCES: 2 356
TURNS: 2 302
TOKENSOVERLAPPED: 0 325
OVERLAP: 0 123

Figure 2. The type token ratio

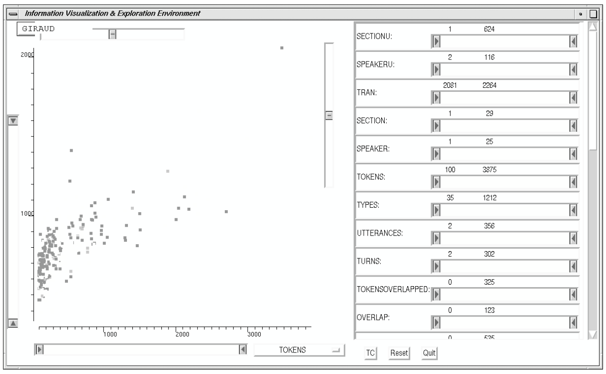
As we can see, the measure is not linear which would mean that it was independent of number of tokens. Rather it tends to give lower values for higher number of tokens.



SECTIONU: 1 624
SPEAKERU: 2 116
TRAN: 2081 2264
SECTION: 1 29
SPEAKER: 1 25
TOKENS: 100 3875
TYPES: 35 1212
UTTERANCES: 2 356
TURNS: 2 302
TOKENSOVERLAPPED: 0 325
OVERLAP: 0 123

Figure 3. Herdan

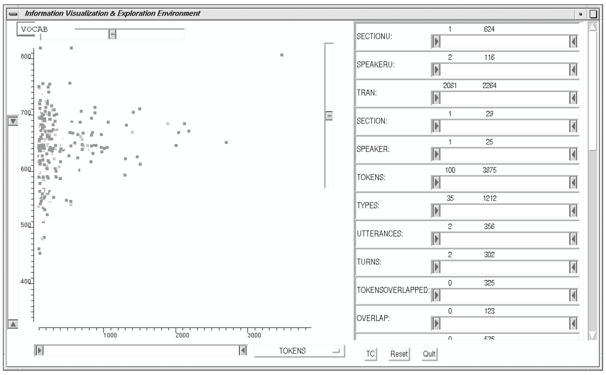
The visualization shows that Herdan' s measure of vocabulary richness has the same problem as the type/token ratio. A large number of tokens tends to give a low value.



SECTIONU: 1 624
SPEAKERU: 2 116
TRAN: 2081 2264
SECTION: 1 29
SPEAKER: 1 25
TOKENS: 100 3875
TYPES: 35 1212
UTTERANCES: 2 356
TURNS: 2 302
TOKENSOVERLAPPED: 0 325
OVERLAP: 0 123

Figure 4.   Guiraud

For Guirand the visualization shows that the problem is the opposite. A high number of tokens tends to give a high vocabulary richness score.



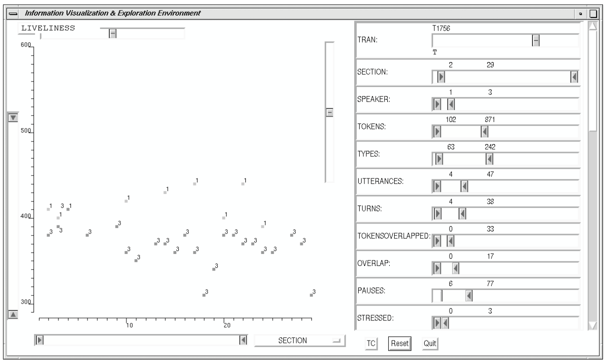
SECTIONU: 1 624
SPEAKERU: 2 116
TRAN: 2081 2264
SECTION: 1 29
SPEAKER: 1 25
TOKENS: 100 3875
TYPES: 35 1212
UTTERANCES: 2 356
TURNS: 2 302
TOKENSOVERLAPPED: 0 325
OVERLAP: 0 123

Figure 5.   Theoretical Vocabulary (Van Hout)

The theoretical vocabulary measure gives the desired stability. High numbers of tokens do not affect the measure in either an upwards or downward direction.

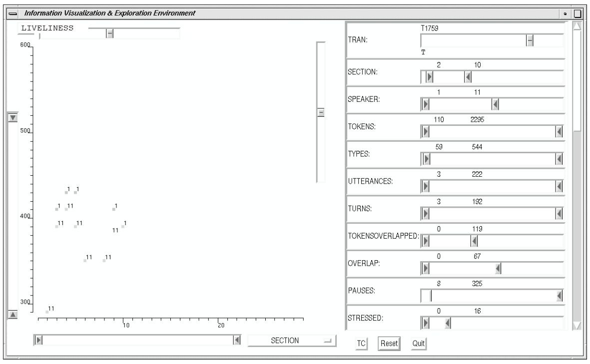
4.2 Activity dynamics and personal variation

The IVEE alphaslider query device also gives us a way to sequentially inspect the dynamics and personal variability in different activities. In figures 6, 7 and 8 we show how liveliness can vary in three different activities. The dots represent different speakers and the location of the dot represents a liveliness measure for a specific speaker in a specific subsection of an activity.



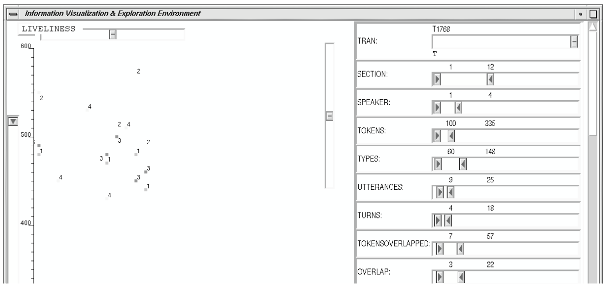
TRAN: T1756 T
SECTION: 2 29
SPEAKER: 1 3
TOKENS: 102 871
TYPES: 63 242
UTTERANCES: 4 47
TURNS: 4 38
TOKENSOVERLAPPED: 0 33
OVERLAP: 0 17
PAUSES: 6 77
STRESSED: 0 3

Figure 6. Liveliness variation in an activity with 2 speakers and 28 sections



TRAN: T1759 T
SECTION: 2 10
SPEAKER: 1 11
TOKENS: 110 2295
TYPES: 59 544
UTTERANCES: 3 222
TURNS: 3 192
TOKENSOVERLAPPED: 0 119
OVERLAP: 0 67
PAUSES: 8 325
STRESSED: 0 16

Figure 7. Liveliness variation in an activity with 2 speakers and 9 sections



TRAN: T1768 T
SECTION: 1 12
SPEAKER: 1 4
TOKENS: 100 335
TYPES: 60 148
UTTERANCES: 9 25
TURNS: 4 18
TOKENSOVERLAPPED: 7 57
OVERLAP: 3 22