Human Communication as Coupled Time Series: Quantifying Multi-Participant Recurrence

Article *in* IEEE Transactions on Audio Speech and Language Processing · August 2012

DOI: 10.1109/TASL.2012.2189566

Human Communication as Coupled Time Series: Quantifying Multi-Participant Recurrence

Daniel Angus, Andrew E. Smith, and Janet Wiles, *Member, IEEE*

***Abstract*—Human communication is more than just the transmission of information. It also involves complex interaction dynamics that reflect the roles and communication styles of the participants. A novel approach to studying human communication is to view conversation as a coupled time series and apply analysis techniques from dynamical systems to the recurring topics or concepts. In this paper, we define a set of metrics that enable quantification of the complex interaction dynamics visible in conceptual recurrence. These multi-participant recurrence (MPR) metrics can be seen as an extension of recurrence quantification analysis (RQA) into the symbolic domain. This technique can be used to monitor the state of a communication system and inform about interaction dynamics, including the level of topic consistency between participants; the timing of state changes for the participants as a result of changes in topic focus; and, patterns of topic proposal, reflection, and repetition. We demonstrate three use studies applying the new metrics to conversation transcripts from different genres to demonstrate their ability to characterize individual communication participants and intergroup communication patterns.**

***Index Terms*—Concept learning, discourse, recurrences and difference equations, text analysis, time series analysis.**

I. Introduction

A. Analysis of Human Communication

ANALYSIS of human communication is generally concerned with understanding orderliness, structure and general patterns of interaction, whether in a formal or informal communication context. Analyses can offer insight into patterns of topic recruitment between participants, and analysts may try to develop rules or models to explain these patterns. These models or rules can be used to inform participants about their own interaction dynamics, which may be useful for training purposes; or they can be used in a forensic context to assist in the identification of anomalies or patterns that deviate from a standard communication model. Human communication is incredibly complex and the majority of analysis methodologies rely on qualitative assessment performed by hand by a trained human operator, examples being Conversation Analysis [20], RIAS [19] and Bales' Interaction Process Analysis [5]. Some methods look specifically to conceptual content present in the discourse to guide the categorization of utterances [14], based on the assumption that effective discourse tends to have a coherent structure.

Computer-assisted qualitative data analysis systems (CAQDAS) can enhance traditional analysis methods through interactive and informative graphical interfaces. Example of such systems include cMap [6], Atlas.ti [1], and NVivo [3]. Such systems rely on user coding to build data relationships, but provide tools that let the user locate, code, and annotate findings in input data. These systems also allow users to interactively tune the importance of input variables, and to visualize complex relations between them.

Automated text analysis techniques attempt to accelerate and simplify the text analysis process while trying to maintain model fidelity and coding reliability, examples being Leximancer™ [23], WordSmith [22], and SPIRE™ [27]. These automated text analysis techniques create a language model to process and classify input text based on principles of word collocation implying connectedness between those terms [17], [21]. Given a suitable language model, these techniques can be used to mine text data for semantic content and provide a visual account of the relationships found; however, they tend to not take account of fine-grained user interaction dynamics or temporal qualities in the input data such as turn-taking.

Computational discourse analysis techniques use text, audio, and video data feeds to assist analyses, measuring features including voice pitch or stress, degree of motion, and percentage of time spent talking. As one example, the use of audio and video signals can be used to determine which participants are dominating a discourse [15].

As an alternative to the approaches listed above we propose an automated method for characterizing text-based multi-participant communication data, in particular for understanding the temporal structure and interaction dynamics of the conceptual content. The proposed technique consists of three major components: a text-based language model, a discourse visualization system, and a novel set of quantitative metrics. The language model and visualization of the temporal interactions using conceptual recurrence plots have been previously reported in Angus *et al.* [4]; a brief overview is provided in Section II (readers familiar with this work may omit this section). The remainder of the paper focuses on the development and demonstration of the quantitative metrics.

Manuscript received September 11, 2011; revised December 16, 2011; accepted February 25, 2012. Date of publication February 29, 2012; date of current version April 04, 2012. This work was supported by the Australian Research Council Thinking Systems under Grant TS0669699. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Haizhou Li.

D. Angus is with the School of Information Technology and Electrical Engineering and School of Journalism and Communication, The University of Queensland, Brisbane, QLD 4072, Australia (e-mail: d.angus@uq.edu.au).

A. E. Smith is with the Institute for Social Science Research, The University of Queensland, Brisbane, QLD 4072, Australia.

J. Wiles is with the School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, QLD 4072, Australia.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASL.2012.2189566

B. Coding an Utterance as a Vector of Concepts

Text-based language modeling is an extensive field with many methods for determining the semantic meaning of text segments; however, much of the sophistication is not needed for this study. The key requirements are a method that provides a model of the conceptual content of an utterance which can be used to calculate the similarity between pairs of utterances, and that the method be capable of processing small segments of text. In this section we provide an overview of a naive Bayesian co-occurrence metric [21, p. 275] which satisfies these requirements. The method for coding utterances presented here is a vector space model; however the vectors obtained for each utterance are based on concepts derived from the text itself, not individual terms.

1) Building Language Statistics: A document D is a collection of sentences, where each sentence is comprised of one or more words (called terms in this study). D is processed to remove stop terms (and, or, the, etc.) and all punctuation, the resulting document D' is then used to generate a single vector of all unique terms $\boldsymbol{T}$ of length $\boldsymbol{T}_{\text{size}}$. D' is then segmented into N blocks of text, each block containing w sentences such that the total number of sentences is $N \times w$ (in this study $w = 2$). These blocks are used to generate an occurrence vector $\boldsymbol{O}$ whose elements indicate how many blocks contain each individual term, and a co-occurrence matrix $\boldsymbol{C}$ whose elements indicate how many times any pair of terms co-occur in the same block. For occurrence and co-occurrence calculation, repeated terms within a block are treated as only occurring once within that block.

In determining the similarity of any two terms the naive Bayesian metric [21, p. 275] considers not only how frequently two terms co-occur $f(i,j)$, but also how often neither occur $f(\bar{i},\bar{j})$, and when they occur apart $f(\bar{i},j)$, $f(i,\bar{j})$. The metric is similar to an odds, or two-way contingency statistic. The similarity of term i to j can be determined as follows:

$$\text{similarity}(i,j) = \frac{f(i,j) \times f(\bar{i},\bar{j})}{f(\bar{i},j) \times f(i,\bar{j})}$$

$$f(i,j) = \frac{\boldsymbol{C}_{i,j}}{N}$$

$$f(\bar{i},\bar{j}) = \begin{cases} \epsilon, & \text{if } \boldsymbol{O}_i + \boldsymbol{O}_j = \boldsymbol{C}_{i,j} + N \\ \frac{(N - \boldsymbol{O}_i - \boldsymbol{O}_j + \boldsymbol{C}_{i,j})}{N}, & \text{otherwise} \end{cases}$$

$$f(i,\bar{j}) = \begin{cases} \epsilon, & \text{if } \boldsymbol{O}_i = \boldsymbol{C}_{i,j} \\ \frac{(\boldsymbol{O}_i - \boldsymbol{C}_{i,j})}{N}, & \text{otherwise} \end{cases}$$

$$f(\bar{i},j) = \begin{cases} \epsilon, & \text{if } \boldsymbol{O}_j = \boldsymbol{C}_{i,j} \\ \frac{(\boldsymbol{O}_j - \boldsymbol{C}_{i,j})}{N} & \text{otherwise} \end{cases}$$

where N is the number of text blocks, $\boldsymbol{O}_i$ is the occurrence of term i, $\boldsymbol{C}_{i,j}$ is the co-occurrence of terms i and j, and ϵ is an adjustment factor to avoid situations where the similarity would become undefined (in this study $\epsilon = 1$).

2) Coding Utterances With Concepts: An utterance is defined as a single turn in a conversation, or a single paragraph (or set number of sentences) in a report. The input document D contains multiple utterances which are stored in an utterance list $\boldsymbol{U}$, of length $\boldsymbol{U}_{\text{size}}$.

To code utterances according to their conceptual content a set of key terms $\boldsymbol{K}$ of size $\boldsymbol{K}_{\text{size}}$ are selected from the term list $\boldsymbol{T}$. In this study key terms ($\boldsymbol{K}_{\text{size}} = 50$) are chosen by selecting the terms with highest occurrence; however, alternative selection methods could be used. Each utterance is assigned a unique concept vector using these key terms as concept identifiers, coded according to the amount of evidence for each key term in that utterance.

To calculate the utterance concept vectors, a similarity matrix $\boldsymbol{S}$ is constructed such that $\boldsymbol{S}$ contains the similarity of every key term i to every term in the term list j using $\text{similarity}(i,j)$ ($\boldsymbol{S}$ has dimension $\boldsymbol{K}_{\text{size}} \times \boldsymbol{T}_{\text{size}}$). A Boolean matrix, $\boldsymbol{B}$, is constructed which indicates the presence of individual terms in each utterance (i.e., 1 if a term is present one or more times, 0 if the term is not present), $\boldsymbol{B}$ has dimension $\boldsymbol{T}_{\text{size}} \times \boldsymbol{U}_{\text{size}}$. The utterance feature matrix, $\boldsymbol{V}$, is generated as

$$\boldsymbol{V} = \boldsymbol{S}\boldsymbol{B}.$$

The dimensionality of $\boldsymbol{V}$ is $\boldsymbol{K}_{\text{size}} \times \boldsymbol{U}_{\text{size}}$, and each column of matrix $\boldsymbol{V}_{*,j}$ represents the weighted set of key terms (concepts) for utterance j.

3) Determining the Conceptual Similarity of Utterance Pairs: The similarity of any two utterances is determined as the cosine similarity of the two corresponding columns of $\boldsymbol{V}$; for example, $\boldsymbol{V}_{*,i} \cdot \boldsymbol{V}_{*,j} / |\boldsymbol{V}_{*,i}| \cdot |\boldsymbol{V}_{*,j}|$ is a measure of the similarity of utterances i and j.

C. Conceptual Recurrence Plots

The conceptual content of human communication can be viewed as a sequence of utterances by one or more participants, where utterances can be modeled as vector of concepts according to the process in the previous section. Each participant's contribution is a single time series over a conceptual space, and the communication as a whole forms a coupled time series. The interaction dynamics in such a time series can be viewed as the similarities between utterances at different points in time, and can be represented using a recurrence plot. The recurrence plotting technique was introduced by Eckmann [9] to display and identify trends within time series data from complex dynamical systems. Given n utterances, there are $n(n-1)/2$ recurrence values. Recurrence plots are visualized as a matrix of squares where each element on one half of the diagonal of a matrix is shaded according to the match between the corresponding horizontal and vertical elements, and where both the horizontal and vertical axes represent time series data.

Recurrence plotting has been applied to human communication data using terms, or symbols, measuring recurrence of single words or alphabetic characters [7], [8], [13], [18]. Conceptual recurrence plotting [4] extends these basic principles to allow comparisons to be made at a conceptual level, such that when utterances share conceptual meaning they are measured as being similar. Conceptual recurrence plots use a continuum of shading to indicate the strength of the conceptual match between terms, as well as changing the units of the time series from individual terms to many terms (sentences or paragraphs). Conceptual similarity measurements present a challenge to the creation of a recurrence plot as many of the existing techniques for

Fig. 1. Conceptual Recurrence Plot of 7 utterances from a television interview [28]. Each square on the diagonal represents one utterance, with color indicating the speaker (red is the interviewer, Andrew Denton; blue is the interviewee, Peter Singer). Shaded rectangles below the diagonal indicate that similar topics are present in the column and rows that intersect at that rectangle (a half/half colored rectangle indicates recurrence between the two speakers, and a single color indicates the speaker is speaking about one of their own earlier topics.

graphing recurrence information in a recurrence plot in both textual and numeric domains rely on the use of a dot-plot style recurrence plot. Such a dot-plot is comprised of individual pixels that, for text, indicate 1 if a word comparison is a match or 0 if it is a non-match. Dot plots are useful for identifying chains of repeated terms, however have limited utility when displaying the continuous values required for conceptual similarity. The use of color coding to display partial matches in matrices is a standard engineering technique; however, it appears to have rarely been applied to recurrence plots, an exception being music pattern identification [10], [16]. The music visualization techniques use shading and color to differentiate between channels of information and to highlight partial matches. A small section of a conceptual recurrence plot is included in Fig. 1.

The conceptual recurrence plots presented in Angus *et al.* [4] highlighted several stereotypical conceptual recurrence patterns. Patterns identified included vertical and horizontal lines stemming from individual utterances, sections of connected strong recurrence close to the diagonal, absence of recurrence for an extended period of time, and gradual weakening of recurrence values moving outward from the diagonal. The placement of these features in context suggested particular participant behaviors, for example strong vertical recurrence lines early in an interview indicated that those utterances set the context for the remainder of the interview.

D. Quantifying Interaction Dynamics

Inspiration for the metric approach outlined in this study stems from recurrence quantification analysis (RQA) which was introduced by Zbilut and Webber Jr. [25], [29] as a way to quantify stereotypical patterns of recurrence that may be present in recurrence plots. RQA is a family of metrics developed to study the phase space trajectory of dynamic systems, and formally RQA contains seven principled metrics: recurrence, determinism, divergence, entropy, trend, laminarity, and trapping time [25]. These metrics are measured using global recurrence information such as the degree to which recurrence is distributed into horizontal or diagonal lines of recurrence. The RQA metrics are useful for quantifying the temporal structure of data that contains repeated patterns, slow drift or rapid state changes.

RQA is widely applicable to time series data, particularly physiological systems data that has characteristic recurrent temporal dynamics [26]. A limiting factor for using RQA in human communication analysis is that RQA is defined using binary recurrence information and is generally applied across recurrence derived from an individual time series. The recurrence patterns identified in conceptual recurrence plots involve real-valued recurrence, and multiple coupled time series (participants). To study interaction dynamics, metrics are required that can make use of real-valued recurrence, and decouple recurrence patterns based on individual and joint participant topic usage over a variety of timescales.

II. Multi-Participant Recurrence Metrics

In this section, we propose a set of multi-participant recurrence (MPR) metrics based on real-valued recurrence information from multiple coupled time-series. The MPR metrics are designed to quantify topic usage patterns in human communication data. Primitives are defined that sum conceptual recurrence values for selected utterances belonging to particular conversation participants over different time scales. A variety of metrics are generated through different arithmetic combinations of these primitives. We define our MPR metrics in three stages:

Initially a set of underlying dimensions are identified which are derived directly from the conceptual recurrence plot (Section II-A). The first dimension of interest in a coupled time series is the proximity between utterances. In a conversation, engaging with a topic immediately signals different behavior from returning to an earlier topic after a long time. Hence, metrics are required to decouple time scales from immediate interactions to spanning the whole conversation. The second dimension decouples the direction of recurrences between those following an utterance (forward) from those preceding it (backward). The third dimension isolates the type of recurrence, which could be from within a single time series (which we term self-recurrence) or recurrence due to interaction with others (termed other-recurrence).

The underlying dimensions (timescale, direction, and type) are combined to define twelve primitives which are summarized in Table I and explained in detail in Section II-B. Each primitive sums a set of elements in a recurrence plot relative to a specific utterance.

A more complex set of eight metrics is then defined as combinations of one or more primitives in Section II-C. The metrics demonstrate several possible combinations of the primitives and are designed to quantify topic usage patterns that may be useful in one or more communication genres.

The primitives and metrics are demonstrated on published communication transcripts of a phone conversation, a television interview and an air emergency in Section III. Our findings demonstrate that the primitives and metrics are useful to quantify topic-based interaction patterns between participants for a range of conversation genres.

A. Dimensions of Recurrence

The primitives for a specific utterance are generated by summing one or more conceptual recurrence values selected according to rules defined over the three principal dimensions.

Time scale:

- Short: the closest single utterance forward or backward from the current utterance and within a short range t_{short}. In this study, t_{short} is set equal the number of conversation participants.
- Medium: utterances within a specified "medium" range t_{med} of utterances from the current utterance; the medium range is intermediate to the short and long ranges used, in this study $t_{\text{med}} = 10$.
- Long: utterances from the entire dataset.

Direction:

- Forward: utterances forward in time from the current utterance.
- Backward: utterances backward in time from the current utterance.

Type:

- Self: utterances by the same participant as the current utterance.
- Other: utterances by a different participant than the current utterance.

As described in detail below, all possible combinations of these three dimensions creates a set of twelve ($3 \times 2 \times 2$) primitives. Metrics are then generated by combining the primitives in various ways, as described in Section II-C.

B. Primitives

The 12 primitives correspond to all possible combinations of the recurrence dimensions (see Table I). For convenience of reference, each primitive is given a name and acronym corresponding to a particular combination of values. For example, the primitive for the parameter combination $\{\text{time scale} = \text{short}, \text{direction} = \text{forward}, \text{type} = \text{self}\}$, is referred to as "short-forward-self," or SFS. Variables and functions used in Table I and Sections III and IV include the following.

A visual depiction for each primitive is illustrated on a conceptual recurrence plot next to the corresponding formulae in Table I; white stars indicate the utterance for which the primitive is being measured, and black stars indicate the recurrence elements included in the calculation of each primitive.

1) Primitive Normalization: For the purposes of primitive combination, it is useful but not essential to be able to scale each primitive to the range $[0, 1]$. One method is based on the minimum and maximum possible values that can be obtained for each of the 12 primitives. If the similarity scores (conceptual recurrence values) obtained are within the range $[0, 1]$, the normalized value of each primitive is calculated by dividing the primitive by the normalization factors in Table II.

The function $count(x)$ returns the total number of indices contained in the specified set x. As an example, the normalized long-backward-self primitive LBS', is given by

$$LBS'(t) = \frac{LBS(t)}{(count\{i \in [1, t-1] : A(t) = A(i)\})}.$$

Any of the primitives can be used in their normalized or un-normalized forms, depending on the context. Best practice in most contexts will be to use the normalized form (represented in this study by a prime symbol). A counterexample where un-normalized primitives would be used is in comparison between different datasets where differences in the average amount of utterance similarity (recurrence) is the measurement of interest.

2) Mean Expected Recurrence: In addition to normalization values which are based on maximum values, it can be useful to calculate the expected value for a primitive or more complex metric. The mean expected recurrence (MER) is the average recurrence calculated over all recurrences summed in that metric. It can be calculated for each time point to show the temporal structure of the conversation or averaged over whole conversations to give a single characteristic value. Note that MERs can be calculated for individual speakers or whole conversations.

For example, the mean expected backward recurrence, MEBR, and mean expected forward recurrence, MEFR, at time t, are given by

$$\text{MEBR}(t) = \frac{\sum_{i=1}^{t_{max}} \sum_{j=i+1}^{t_{max}} u_{ij}}{\frac{t_{max}(t_{max}-1)}{2}} \times t \tag{1}$$

$$\text{MEFR}(t) = \frac{\sum_{i=1}^{t_{max}} \sum_{j=i+1}^{t_{max}} u_{ij}}{\frac{t_{max}(t_{max}-1)}{2}} \times (t_{max} - t) \tag{2}$$

where t_{max} is the total number of utterances included in the recurrence plot, and u_{ij} is element i, j of the recurrence plot.

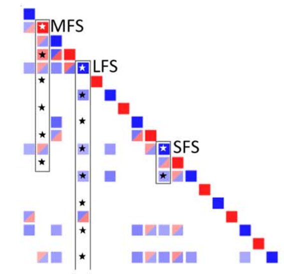
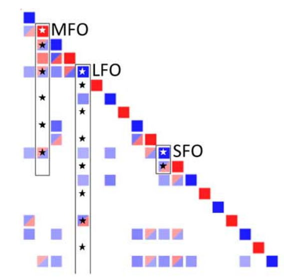
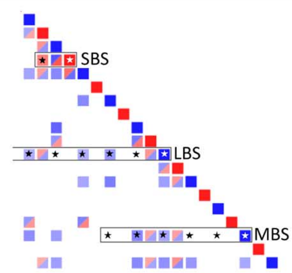
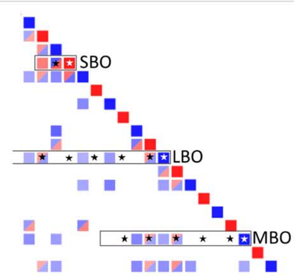
C. Metrics

The 12 primitives and corresponding normalization factors can be used as given or combined in various ways to generate metrics that provide quantitative measures of topic usage by one or more participants. The eight metrics described below are designed to quantify specific topic usage patterns, and while an exponential number of combinations are possible, these eight metrics explore a variety of arithmetic combinations of the primitives that are suitable for a wide range of analyses and are demonstrated in the use studies in Section III.

1) Short Time Constant Metrics: Short time constant metrics quantify short-term topic usage patterns in conversation, generally between one utterance and the next.

Immediate Topic Repetition (ITR): Topic repetition is a very common pattern in conversation and can indicate a range of behaviors. For example, in a learning task repetition may indicate adherence to a learning protocol, in a doctor-patient consultation it may indicate that a patient does not feel they have communicated their point, and in normal conversation it may indicate engagement between participants signalling comprehension and encouragement or the converse. Exact repetition (echolalia) can also be characteristic of communication disorders such as those exhibited by individuals with autism. The ITR metric measures the degree to which a person repeats concepts from the most recent utterance by another speaker. ITR is based on the normalized SBO primitive, which measures the similarity of the utterance under study to the previous utterance

TABLE I
MULTI-PARTICIPANT RECURRENCE PRIMITIVES

Primitive	Supporting Equation	Visual Depiction of Primitives
Short Forward Self	$SFS(t) = S(t, \text{next}(t))$	
Medium Forward Self	$MFS(t) = \sum_{i=t+1}^{t+t_{med}} \text{self}(t,i)S(t,i)$	
Long Forward Self	$LFS(t) = \sum_{i=t+1}^{t_{max}} \text{self}(t,i)S(t,i)$	
Short Forward Other	$SFO(t) = \text{other}(t,t+1)S(t,t+1)$	
Medium Forward Other	$MFO(t) = \sum_{i=t+1}^{t+t_{med}} \text{other}(t,i)S(t,i)$	
Long Forward Other	$LFO(t) = \sum_{i=t+1}^{t_{max}} \text{other}(t,i)S(t,i)$	
Short Backward Self	$SBS(t) = S(t, \text{last}(t))$	
Medium Backward Self	$MBS(t) = \sum_{i=t-t_{med}}^{t-1} \text{self}(t,i)S(t,i)$	
Long Backward Self	$LBS(t) = \sum_{i=1}^{t-1} \text{self}(t,i)S(t,i)$	
Short Backward Other	$SBO(t) = \text{other}(t,t-1)S(t,t-1)$	
Medium Backward Other	$MBO(t) = \sum_{i=t-t_{med}}^{t-1} \text{other}(t,i)S(t,i)$	
Long Backward Other	$LBO(t) = \sum_{i=1}^{t-1} \text{other}(t,i)S(t,i)$	

spoken. The metric is given the name ITR to maintain consistency of naming conventions with the medium and long time metrics defined below. ITR at time t is defined as follows:

$$\text{ITR}(t) = SBO'(t). \tag{3}$$

For (3) to (10) see Table I for the primitive definitions and Table II for their normalization factors.

2) Medium Time Constant Metrics: Medium time constant metrics measure patterns of topic use over intermediate time periods, longer than one exchange but shorter than a whole conversation. Medium term group oriented behaviors of interest involve aspects of topic introduction, consistency and repetition. The metric-relevant range, t_{med}, is selected to reflect the span of shared attention to a topic by a group, and can be measured by finding where the recurrence value drops below a set threshold, or can be set to a predetermined or default value ($t_{\text{med}} = 10$ utterances for the three use studies in this paper). For a single

TABLE II
PRIMITIVE NORMALIZATION FACTORS

Primitive	Normalisation Factor
SFS	1
MFS	$count\{i \in [t+1, t+t_{med}] : A(t) = A(i)\}$
LFS	$count\{i \in [t+1, t_{max}] : A(t) = A(i)\}$
SBS	1
MBS	$count\{i \in [t-t_{med}, t-1] : A(t) = A(i)\}$
LBS	$count\{i \in [1, t-1] : A(t) = A(i)\}$
SFO	1
MFO	$count\{i \in [t+1, t+t_{med}] : A(t) \neq A(i)\}$
LFO	$count\{i \in [t+1, t_{max}] : A(t) \neq A(i)\}$
SBO	1
MBO	$count\{i \in [t-t_{med}, t-1] : A(t) \neq A(i)\}$
LBO	$count\{i \in [1, t-1] : A(t) \neq A(i)\}$

focused conversation group, a constant value may be adequate, but for fragmented forums a temporal measure may be required, such as a running average.

Topic Introduction (TI): It can be useful to distinguish between the first use of a topic and subsequent repetition. The TI metric measures the degree to which a speaker contributes topics which are not direct repetition of the immediate previous utterance, and which are also repeated by subsequent utterances. The TI metric for a given utterance t is determined by summing the recurrence of every successive utterance by every other person within the metric-relevant range. This sum is then scaled by $1 - SBO'(t)$ to attribute higher TI scores to people who introduce the conceptual content that other people have subsequently talked about in the immediate conversation. TI at time t is defined as follows:

$$\mathrm{TI}(t) = MFO'(t) \times (1 - SBO'(t)). \tag{4}$$

Topic Reiteration (TR): The TR metric measures the degree to which an utterance contains previously mentioned topics, specifically those that were mentioned by another speaker in the immediately previous utterance. For example, if a person were to speak on topic but ignore the previous person's utterance it would score low TR; if they were to repeat only the previous person's utterance, and ignore the general topics of the last ten utterances that would also score low TR. To score highly, an utterance needs to contain topics raised by others in both the previous and earlier utterances. The TR metric is determined by summing the recurrence of every preceding utterance by other people within the metric-relevant range ($MBO'(t)$) and scaling this sum by $SBO'(t)$. TR at time t is defined as follows:

$$\mathrm{TR}(t) = MBO'(t) \times SBO'(t) \tag{5}$$

Topic Consistency Other (TCO): The TCO metric measures the degree to which the current utterance contains concepts mentioned by other people in the temporal vicinity (forward and backward), essentially measuring how on-topic the current utterance is in the context of the surrounding conversation. The TCO metric for an utterance is determined by summing the recurrence of every preceding and following utterance by other people within the metric-relevant range. TCO at time t is defined as follows:

$$\mathrm{TCO}(t) = MBO'(t) + MFO'(t). \tag{6}$$

Topic Consistency Self (TCS): The TCS metric measures the degree to which a person repeats their own concepts within a medium range conversational time frame. The TCS metric for an utterance is determined by summing the recurrence of every preceding and successive utterance by the same person within the metric-relevant range. TCS at time t is defined as follows:

$$\mathrm{TCS}(t) = MBS'(t) + MFS'(t). \tag{7}$$

3) Long Time Constant Metrics: Long time constant metrics measure patterns of topic use over entire conversations. These metrics quantify how a particular utterance or person introduces or reuses topics over an extended period of time, even after the conversation has had numerous topic changes.

Longterm Topic Novelty (LTN): The LTN metric measures the degree to which concepts contained in an utterance are used forward in time when these same concepts have not been used previously. The LTN metric for an utterance is an asymmetric metric that is determined by summing the recurrence of every preceding utterance by a different person and subtracting this sum from the summed recurrence of every subsequent utterance by a different person within the entire conversation. A largely positive LTN implies that the utterance contains concepts that have been repeated by other people strongly in the future; if largely negative, the utterance uses concepts mentioned by other people in the past that do not recur strongly in the future. The asymmetry of this metric introduces a positional bias in a time series which can be normalized if needed. LTN at time t is defined as follows:

$$\mathrm{LTN}(t) = LFO'(t) - LBO'(t). \tag{8}$$

Longterm Topic Consistency Other (LTCO): The LTCO metric measures the degree to which the current utterance contains concepts mentioned by other people, like the TCO metric, however unlike the TCO metric the LTCO metric is concerned with topics that appear throughout a conversation. The longterm topic consistency metric is determined by summing the recurrence of every preceding and subsequent utterance by a different person within the entire conversation. High LTCO values indicate that a person is referring to concepts mentioned by other speakers (from all time). Lower LTCO values indicate that an utterance is unique in the concepts it contains, or is comprised mostly of stop words. LTCO at time t is defined as follows:

$$\mathrm{LTCO}(t) = LFO'(t) + LBO'(t). \tag{9}$$

Longterm Topic Consistency Self (LTCS): The LTCS metric measures the degree to which a person repeats their own concepts throughout a conversation (extending the TCS metric to the full conversation). The LTCS metric is determined by summing the recurrence of every preceding and subsequent utterance by the same person within the entire conversation. LTCS at time t is defined as follows:

$$\mathrm{LTCS}(t) = LFS'(t) + LBS'(t). \tag{10}$$

The four metrics TCO, LTCO, TCS, and LTCS systematically measure patterns of topic reuse over intermediate to long periods of time by self and others. Total topic recurrence can be calculated by summing corresponding pairs over the medium term ($\mathrm{TCO} + \mathrm{TCS}$) and long term (LTCO and LTCS) time scales.

III. Use Case Analysis

The major challenge for conversation data is seeing the global structure of the conversation and identifying important points that signal features of interest or major changes in the dynamics in participant interactions. We are particularly interested in patterns of topic use by individuals and patterns of interaction, and ways in which quantification of those patterns can be used to identify regions of interest for further study, including detailed examination of the text.

In analysis of communication data, it is important to note that features of a conversation are not in themselves necessarily good or bad. Hence, the interpretation and behavioral correlates of a particular metric need to be made with respect to the context of the conversation, and verified by examination of the text itself. In the sections above we have defined a comprehensive set of metrics which are intended to quantify a wide variety of patterns of topic use.

In this section, the metrics are demonstrated in use studies in which the metrics enable informed analysis of interaction patterns in the context of specific genres. However, note that the use studies are illustrative of the metrics only, as a complete empirical validation is beyond the scope of this paper.

Three use cases of the proposed primitives and metrics are presented to illustrate how the metrics can be used in a range of communication genres where analysts face a variety of challenges and goals. The use cases demonstrate how features identified using the recurrence plotting process can be quantified using the multi-participant recurrence metrics.

A. Use Case 1: Telephone Conversation

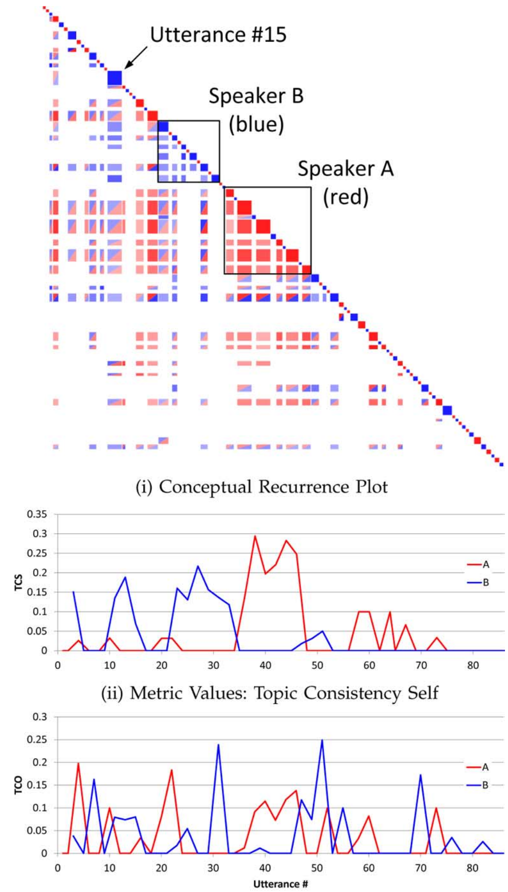
1) Aim: The aim of the first use case (UC1) was to demonstrate the metrics on natural conversation in which the average utterance size is small. In particular, this first use study shows how qualitative features of the recurrence plots are reflected by the MPR metrics. The data set is from a standard corpus, the Switchboard Dialog Act Corpus [12], which contains telephone conversations between unacquainted adults.

2) Data: A transcript of a conversation (#4325) from the Switchboard Dialog Act Corpus [12] was analyzed. This conversation is between two female callers discussing issues relating to child care. The complete conversation consisted of 43 utterances by each speaker, 86 utterances in total.

3) Observations: A conceptual recurrence plot of the entire SDAC_4325 conversation was created and is reproduced in full in Fig. 2(i). The squares along the diagonal of the recurrence plot show turn taking by the speakers, with the size of each square representing the number of concepts in that utterance and the color representing the speaker. As the squares are mostly small, the visualization shows that utterances were typically short, with a few longer ones by both speakers. The triangular region below the diagonal shows the recurrence between utterances.

In this conversation, there is a large amount of red recurrence (indicating speaker A talking about their own concepts), a medium amount of blue recurrence (indicating speaker B talking about their own concepts) and a moderate amount of two-color recurrence (red-blue and blue-red, indicating that both speakers referred to each others' topics). A variety of interesting features can be identified in the recurrence plot. For example, at utterance 15 there is a large blue square on the diagonal with the stripe to the left empty (i.e., no recurrence) and a multicolored vertical stripe below it, indicating that the utterance introduced novel topics which then recurred throughout the rest of the conversation. Examination of the utterance (reproduced below) indicates the topics of interest related to child care:

> B: No, I don't, have any kids. I, uh, my sister has a, she just had a baby, he's about five months old and she was worrying about going back to work and what she was going to do with him and, Uh-huh. the different.



(i) Conceptual Recurrence Plot

(ii) Metric Values: Topic Consistency Self

(iii) Metric Values: Topic Consistency Other

Fig. 2. Features of a phonecall. (a) Recurrence plot of the entire conversation. The boxed sections show two periods where speakers have high self-recurrence. (b) The medium term metric TCS shows high self-recurrence values for these periods, contrasted by low recurrence for topics raised by the other speaker, as measured by the metric TCO (c). Data source: Switchboard Dialog Act Corpus [12] #4325.

As a second example, two periods of strong single color recurrence followed shortly after this statement, which on examination proved also to be on the topic of child care (see the boxed sections in Fig. 2(a). The original text from eight utterances in the first of the boxed sections is reproduced here:

> B: Uh, yeah she lives, it's a fairly large community. She, uh, got real lucky, though. He had a boss who, uh, moved into a larger office
> A: Uh-huh.
> B: and she's able to take her baby to work with her.
> A: Oh, really?
> B: And it's a small office that she works in.
> A: Uh-huh.
> B: and, uh, it's a legal firm, office, and it's just one lawyer.
> A: Um.

4) Metrics: The interaction dynamics between the speakers around shared topics is reflected in the color and structure of the patterns in the recurrence plot, and these patterns contribute to a qualitative understanding of the dynamics. To compare the speakers' contributions to the conversation we quantified the use and reuse of topics by each speaker.

TABLE III
AVERAGE TOPIC CONSISTENCY VALUES PER SPEAKER FOR SDAC DATASET #4325

	TCS	TCO
A	0.043	0.033
B	0.037	0.036

Metrics for this conversation were selected to illustrate the time varying nature of topic consistency by each speaker, both with reference to their own and the other speaker's utterances. To quantify the boxed sections of self recurrence, medium time scales were considered appropriate, and two metrics TCS and TCO were calculated for each utterance in the conversation (see Fig. 2(ii) and (iii), respectively). The time series for the TCS metric quantifies the regions where each speaker is dominating the conversation. It also indicates speaker B (blue) had another long contribution earlier in the conversation, speaker A had most to say at the midpoint, with a number of smaller peaks in the metric later in the conversation.

Comparison of the TCS and TCO metrics provides a method for quantifying and contrasting how much each speaker revisits their own compared to the other speaker's topics. The total TCS and TCO metrics were compiled into a summary table to illustrate the relative contributions of the two speakers. Speaker A revisited her own topics more frequently than the other speaker's topics $(\mathrm{TCS} > \mathrm{TCO})$ whereas speaker B referred to both equally see Table III). The summary metrics also indicate that both speakers referred to each other's topics at a similar rate (0.033 and 0.036). An analyst interested in specific issues could use additional metrics to provide insight into more detailed aspects of the conversation, or to quantify other comparisons of interest.

UC1 demonstrates how the recurrence plot can be used to identify the global structure of the conversation, and find utterances and periods of interest for further analysis. These sections can then be quantified using the recurrence metrics. The medium term metrics enable comparison of self and other recurrence for topic use throughout the conversation, and the time series plot of the metrics provided an overall perspective on the interactions between the participants in their use of their own and the other speaker's topics.

B. Use Case 2: Television Interview

1) Aim: The aim of the second use case (UC2) was to evaluate the use of the MPR metrics in identifying regions of interest in a conversation in which the participants have distinctly different roles. To assess the usefulness of the MPR metrics in performing such a task the metric values were compared to observations and predictions made by an expert communication analyst (Section III-B5 and Table IV). In particular, do the metrics predict complementary and conflicting patterns of topic use, topic changes, and flag other points of interest within a conversation for further analysis.

The second use case data set is from a television talk show which necessarily involves distinct roles for the interviewer and interviewee. Talk show hosts can be politically and socially influential [24] and the host is typically perceived to have a powerful position in setting the agenda during a conversation, as well as editing after the recording session prior to broadcast. Television talk show interviews are typically semi-structured with the host aiming to present an entertaining event. Interviewees frequently have a reason for participating in the interview and hence a set of topics that they intend to discuss. Of particular interest are the patterns of topic use by both speakers, including the asymmetric discussion of topics by the host and interviewee and topic transition points. In a longer interview, such as a televised interview for entertainment, the recruitment of topics is likely to also contribute to a global structure.

2) Data: The interview analyzed in this use case is from the Australian television series *Enough Rope* [28]. The transcript is of an interview with Jeff Kennett, a former Australian politician and the spokesperson for a depression awareness and assistance organization called Beyond Blue. This particular interview was described by the interviewer, Andrew Denton, post-interview as being particularly difficult. In a post-interview reflection, Denton indicated that he had wanted to discuss elements of Kennett's political career, whereas Kennett was only comfortable discussing his position as spokesperson for Beyond Blue. The complete conversation consisted of 60 utterances by each speaker, 120 utterances in total.

3) Observations: A conceptual recurrence plot of the entire Denton/Kennett interview was created and is reproduced in full in Fig. 3. Uniform sizing of the utterance squares on the diagonal enables a clear view of the interactions between participants around topics of interest. The recurrences (shown as colored squares below the diagonal) are primarily blue, indicating that Kennett largely spoke on his own topics. The low level of red recurrence indicates that Denton rarely repeated his own topics. A complementary recurrence plot using variable sized recurrence squares was also created, showing that Kennett's utterances were much longer than Denton's. As the same global structure of the conversation was visible in both the fixed and variable sized recurrence plots, only the fixed size is shown.

Examination of the transcript indicates that the Denton/Kennett interview was characterized by periods where Denton tried unsuccessfully to discuss Kennett's political career or private life. Strong engagement by both participants was elicited only around concepts related to depression. Five key utterances by Kennett early in the interview recurred with a major section later in the conversation and showed strong self-recurrence by Kennett (Fig. 3, see the large boxed sections on the left and in the center of the recurrence plot). The later section of strong recurrence was eventually interrupted by Denton who tried, unsuccessfully, to segue away from depression towards politics. Within four utterances the conversation returned to talking about depression, with Kennett being the one initiating the switch back to this topic, indicated by the blue horizontal recurrence stripe towards the end of the conversation.

The conversation provides good examples of an interviewer raising issues and an interviewee refusing to discuss them and only engaging with topics of his own choosing.

4) Metrics: The recurrence plot suggests that the Denton/Kennett conversation contained periods characterized by distinctly different interaction dynamics. Given the focus by Denton on politics and Kennett on depression awareness, the question of interest for the metric analysis is to what

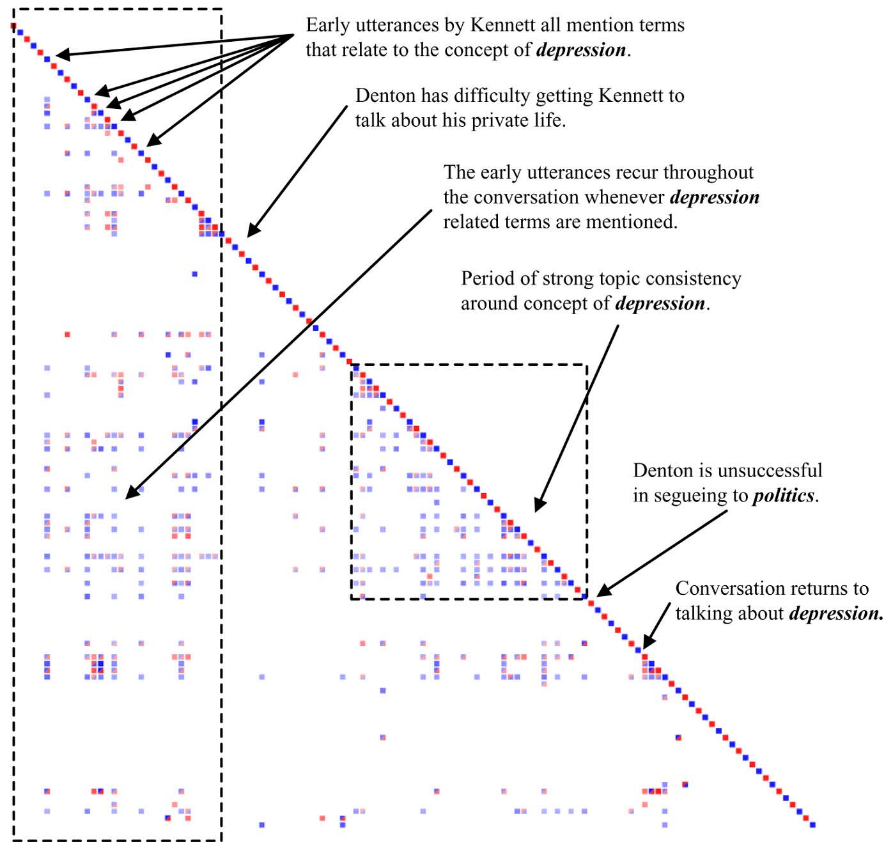


Fig. 3. Conceptual recurrence plot for a televised conversation between Andrew Denton and Jeff Kennett on the ABC Enough Rope program. Due to the length of the conversation, a fixed size square is used for each utterance along the diagonal (Kennett in blue and Denton in red). Below the diagonal, the recurrence is mostly blue indicating Kennett repeating his own topics. The large dashed boxes indicate two major sections of recurrence, both around the concept of depression (see text for details). Data source: ABC Enough Rope [28].

degree different periods of the conversation could be identified and the features of the different periods characterized. Long term metrics were selected to quantify the topic-based interactions, including both long term time-oriented primitives and symmetric topic consistency metrics.

The metric values for LTCO and LTCS were combined to give a value of the total recurrence for each of Denton and Kennett, and the time series for each agent were also combined to give a total topic consistency value [see Fig. 4(i)]. The non-normalized long-forward-other, LFO, metric primitive was calculated for the entire duration of the interview [see Fig. 4(ii) and (iii)], for utterances by Denton and Kennett, and graphed using drop lines against the MEFR, calculated using (2). Similarly, The non-normalized long-backward-other, LBO, primitive was calculated for the entire duration of the interview [see Fig. 4(iv) and (v)], again for utterances by Denton and Kennett, and graphed using drop lines against the MEBR, calculated using (1).

Separating the individuals made it possible to observe how the deviation from the expected recurrence differs for each participant, particularly Kennett's large period of backward recurrence from utterance 35 to 61.

The time series plots of the metric values for total topic consistency [Fig. 4(i)] were used to divide the conversation into three distinct periods (marked by vertical lines in Fig. 4). The text for each of these periods was then subjected to a more detailed examination. It was found that each period was strongly driven by the topics that Denton raised, and the reaction by Kennett:

- Period 1 ($t = 1 \rightarrow 33$): Initial greeting and discussion by Denton. Kennett outlined how he wanted to talk about his role with the depression awareness group Beyond Blue, with four key statements indicating this conceptual position. These four statements are observed as peaks of high backward and forward recurrence by Kennett in Fig. 4(ii). Denton then tried, unsuccessfully, to get Kennett to talk about his private life. Kennett's responses were short and contained few concepts that were mentioned elsewhere in the conversation. Denton's low forward recurrence in this section [Fig. 4(iii)] indicates that he did not repeat these concepts later in the interview.
- Period 2 ($t = 34 \rightarrow 63$): Kennett's high backward and forward recurrence [Fig. 4(iv)] for this period was due to the focus of the interview returning to the concept of depression.
- Period 3 ($t = 64 \rightarrow 83$): At the beginning of this period Denton asked Kennett a series of questions about his political career. Kennett's responses turned back to his role with the depression awareness group Beyond Blue and these utterances can be seen as having higher than average backward recurrence [Fig. 4(iv)]. The lack of recurrence for the rest of the period suggests that Kennett did not reflect on previous concepts in answering these questions, which is

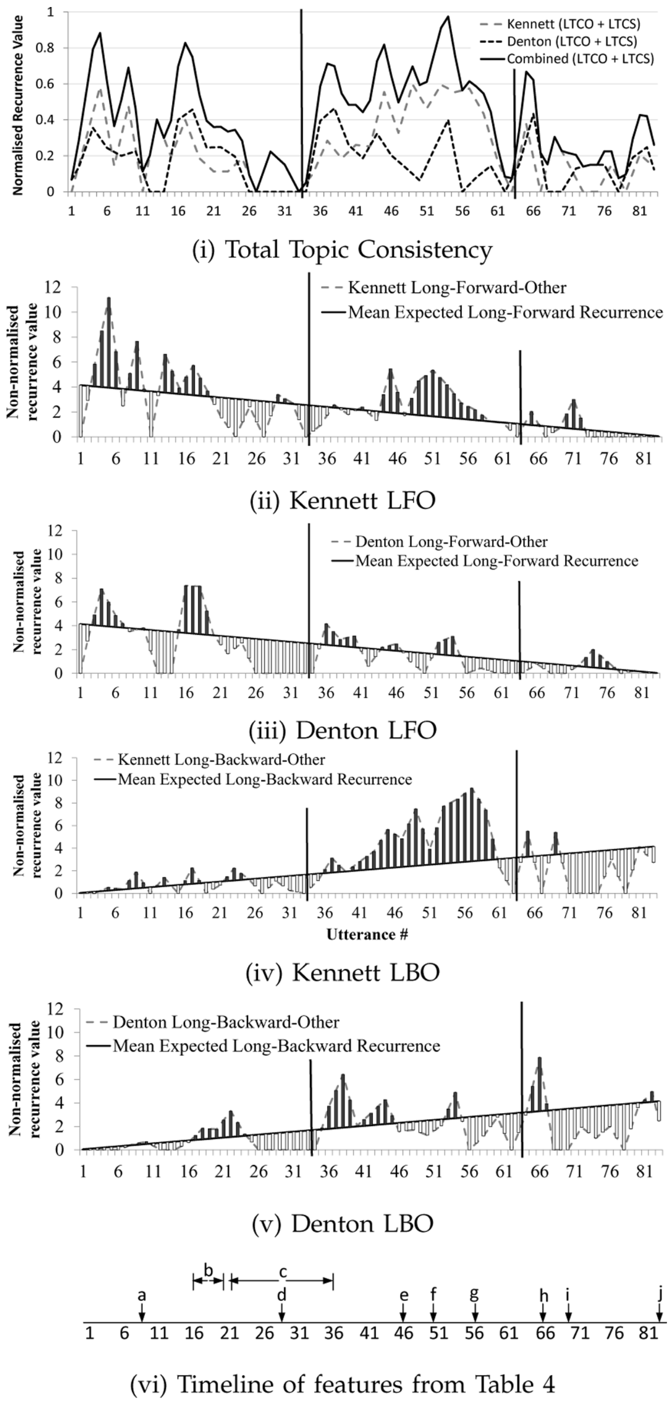


Fig. 4. Graphs of long term asymmetric primitives and symmetric metric values showing key points in the ebb and flow of the conversation. (a) The metric values for LTCO and LTCS combined give the total recurrence for each of Denton and Kennett, further combined they give a total topic consistency value. (b) and (c) Long forward other primitive, LFO, and mean expected forward recurrence, $MEFR$, shown for both speakers. (d) and (e) Long backward other primitive, LBO, and mean expected backward recurrence, $MEBR$, shown for both speakers. By separating the individuals and directions of recurrence we observe how the deviation from the expected recurrence sum is different for each participant, particularly how Kennett has a large period of backward recurrence after utterance 37.

TABLE IV
INDEPENDENT EXPERT'S COMMENTS ON DENTON/KENNETT CONVERSATION (SEE FIG. 4)

Tag (Time)	Comments
a (8)	Kennett is clearly setting the agenda and Denton is floundering. He has hardly said anything and Kennett is moving the interview along not Denton.
b (16 → 20)	This part of interview looks good and is more engaging and on topic for both participants.
c (21 → 36)	Minimal topic repetition here with Denton doing all the work.
d (28)	Some minimal engagement here because Denton gets facts right.
e (46)	Kennett talks about personal experience with depression and Denton has a chance to pick up on this topic.
f (50)	This is more promising even though again it is Kennett driving the the interview and explaining what depression is.
g (56)	This section of the conversation is much better because the topic is what Kennett wants to talk about. This is the best segment in the interview because Kennett has engaged with Denton but again it is about to fail.
h (66)	No real topic building in this section, just more examples of how if the interview had gone well there could have been good discussion on the passion that Kennett feels for the subject of depression and how there could have been rapport building between the two speakers. This however does not happen.
i (69)	This is going nowhere. They really are talking at cross purposes even though both are staying on topic.
j (End)	Denton was desperate to reconcile Kennett so that he could move on with the interview, instead the interview ends flat.

TABLE V
AVERAGE TOPIC CONSISTENCY VALUES PER SPEAKER FOR DENTON/KENNETT DATASET

	TCS	TCO
Denton	0.071	0.099
Kennett	0.145	0.098

supported by analysis of the raw text. This period of the interview mostly contained closing remarks and a final unsuccessful effort by Denton to get Kennett to talk openly about his personal and political life, consequently the total recurrence for both participants is lower than average.

UC2 demonstrates the use of metrics to identify periods of interest in a conversation and data guided exploration of the dynamics around particular topics. Denton's utterances are on average more topically consistent with Kennett's utterances (Denton's TCO $>$ TCS), while Kennett's utterances tend to recur mostly with his own (Kennett's $\mathrm{TCO} < \mathrm{TCS}$).

The total TCO and TCS metrics were compiled into a summary table (Table V) to compare the relative topic reuse of the two speakers. The summary table shows a clear asymmetry between the interviewer and interviewee. Both speakers engaged similarly with each other's topics ($\mathrm{TCO} = 0.099$ and 0.098) but differed markedly in repeating themselves, with Denton having a lower self-recurrence ($\mathrm{TCS} = 0.071$), and Kennett much higher ($\mathrm{TCS} = 0.145$). This example highlights the advantages of being able to divide recurrence values along the type dimension (self or other), as it clearly highlights how Denton spent time repeating concepts mentioned by Kennett and tried to also steer the conversation through the introduction of different topics, while Kennett exhibited a strong self-monologue by mostly repeating his own concepts.

5) Expert Analysis and Prediction: An independent (human) expert was approached to assess the Denton/Kennett interview. The inclusion of an independent assessment of this dataset allows observations made using the MPR metrics to be compared with a ground truth provided using an alternative and widely accepted communication analysis methodology. Dr. Bernadette Watson, a clinical psychologist and expert in Communication Accommodation Theory (CAT) [11], was provided with video and audio of the original conversation (not used in the metric analysis), and the transcript, but no information on what conclusions had been reached either qualitatively, or with the MPR metrics. The expert's independent findings (shown in Table IV) are a useful ground truth for comparing how well the MPR metrics capture critical time points of the conversation related to topic consistency (self and other).

Findings from the independent assessment agree well with the findings from the MPR metrics. The rise in topic consistency between utterance 16 and 20 agree with the expert's finding that this period of the interview was more engaged and on topic. The rise in topic consistency by Kennett between utterances 34 and 63 is reflected in comments by the expert about Kennett driving the interview and speaking at length around the topic of depression. The dip in topic consistency after utterance 56 is matched by the expert's observations of the interview failing after this point, and beyond this point she said the interview is "going nowhere" which agrees with the low topic consistency for the remainder of the interview.

C. Use Case 3: Aircraft Transcript

1) Aim: The aim of the third use case (UC3) was to demonstrate the use of the MPR metrics for analysis of a dataset containing multiple (more than 2) conversation participants, and to compare different metrics for one speaker and each metric across multiple speakers. The interactions concerned an air emergency, analyzed using a transcript of the black box recorder. Issues of interest in such cockpit recordings include characterization of the interaction dynamics associated with different roles during the emergency.

2) Data: The third conversation transcript is of communications from the famous United Airlines flight 232 [2]. On July 19, 1989, this Douglas DC10 suffered an uncontained failure of its number 2 engine. The actions of the crew and a DC-10 instructor pilot (jumpseat training pilot) who happened to be traveling on this flight at the time, are credited with a reduction in the total number of casualties following a crash landing. This is a good example of communication management in a critical situation. The transcript contains 372 utterances by 11 participants.

3) Observations: An important aspect of an emergency situation such as that of Flight 232 is how the captain/leader/manager controls the discourse. In the case of Flight 232 the captain was observed to poll the flight and ground crew for information, listen to the response and then order a directive based on the information obtained. The effective use of time by the crew is also notable as they are engaged in topically consistent communication most of the time, particularly around topics such as suitable landing sites, aircraft control, and mechanical issues (hydraulics).

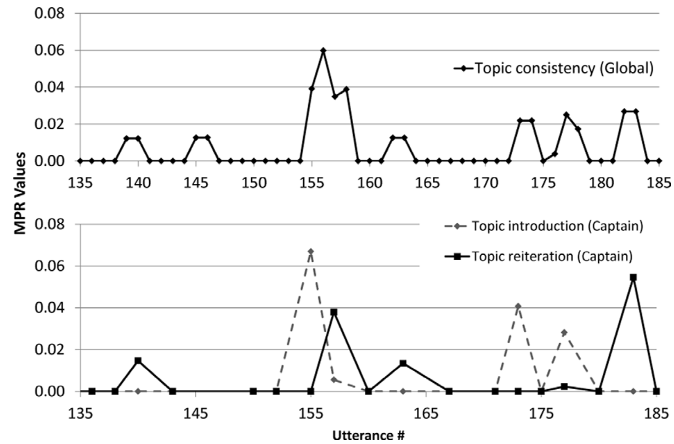


Fig. 5. Topic introduction/reiteration pattern of the captain in the Flight 232 transcript. The upper graph indicates the LTCO metric averaged over all participants, and the lower graph indicates the TI and TR metric values for the captain alone. Data source: [2].

The following text illustrates particular roles within the conversation. It shows how the jumpseat[1] captain offered assistance by gathering information as directed by the captain and first officer, an example being:

> Captain: What the hell. Let's do it. We can't get any worse than we are …
> First Officer: Slats are out?
> Jumpseat Captain: No, you don't have any slats.

Ground crew, such as Sioux City Approach also fulfil an information response role offering answers to requests for information by the flight crew, two examples being:

> United two thirty two heavy, there are a couple of really small airports out in the vicinity here, and Storm Lake is four thousand two hundred feet by seventy five. That's about fifteen miles east of your position.

> United two thirty two heavy, there is a small airport at twelve o'clock and seven miles. The runway is four thousand feet long there.

4) Metrics: Metrics were calculated for global TCO (the TCO of all participants combined), TCO values for each individual utterance by all participants, and the captain's TI and TR scores for his individual utterances (an example is shown in Fig. 5). The captain's TI values are shown to increase sharply immediately prior to many of the large TCO blocks, while his TR increases immediately following these periods of strong topic consistency. These metric patterns are consistent with observations that the captain was engaged in interactions that involved both raising issues and actively listening to responses, using the topics included by respondents in acknowledging their response to his questioning.

The summed totals for each of six metrics for a subset of the most prominent participants for the transcript were calculated. These summed metric values can be used to suggest the degree of adherence to specific roles by the crew members (see Fig. 6).

[1]In aircraft, jump seats are provided for employees who are not operating the aircraft, often used by trainee pilots or off-duty crew members who are in transition to another airport.

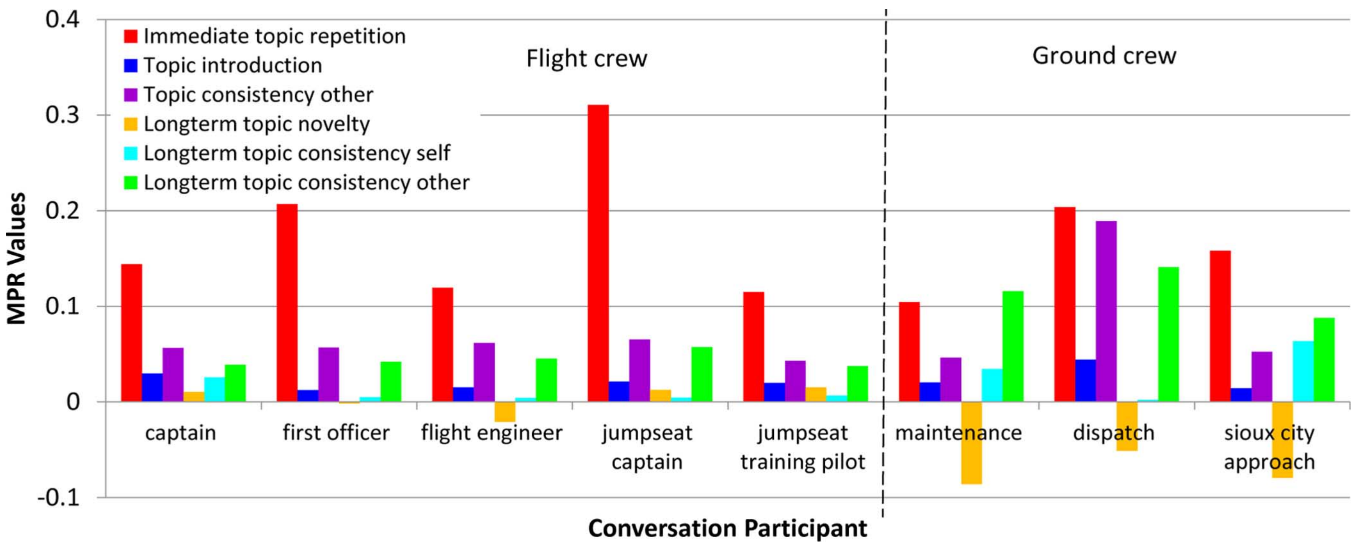


Fig. 6. Participant scorecard for Flight 232 transcript. The bar length reflects the averaged metric totals from the entire transcript. Note the very high scores for ITR (red bars) shown by all participants, an important feature of air emergency procedures.

Not surprisingly, the captain scored highly on TI as he was introducing topics for the length of the conversation. The ground crew tended to have low LTN while the flight crew (including the captain) had high LTN.

The values for LTN reflect the directionality of the communication, particularly as the role of the ground crew in this situation was to respond to questions by the flight crew, such as requests for a suitable landing site. Sioux City Approach has the highest TCS which reflects repetition of information regarding airport locations and heading directions to the flight crew. The high value for ITR by the jumpseat captain is due to this person repeating topics from directives provided by other members of the flight crew.

UC3 demonstrates that the averaged metrics provide a useful way to quickly gain a overview of the topic usage dynamics of participants in a conversation, particularly when there are more than two participants. Further analyses could examine finer-grained aspects of interaction dynamics between the participants or the time series of the interactions throughout the emergency.

IV. Conclusion

In this paper, we proposed a set of primitives and metrics to identify and measure the structure and general patterns of topic usage in both formal and informal communication contexts. The set of multi-participant recurrence (MPR) metrics can be used to decouple recurrence patterns based on participant, proximity to nearby utterances, and direction in time. As demonstrated in the three use cases, the application of the MPR metrics to a conversation provides an analyst with a range of measures to help them identify critical points in the conversation, build an understanding of topic use by specific participants within a single conversation, or generate informative summaries for individual conversation participants.

The MPR metrics can be used for exploratory data analysis, or to confirm topic use patterns by specific conversation participants given prior knowledge or expectations. The value of the metrics is first and foremost their ability to characterize conversation periods and time points of interest. In use case 1, the SDAC phone conversation showed how the two callers shared the conversation with each having relatively equal amounts of OTC. In use case 2, the participant roles were known, and the conversation was known to be a difficult interview; however, specific time points where this difficulty was visible was unknown ahead of time. For use case 2 the MPR metrics were able to highlight the points where the interview became difficult, or where topic agreement was reached, and an independent expert assessment agreed that the features identified by the MPR metrics were noteworthy. For use case 3, the Flight232 transcript, the MPR metrics aligned with prior expectations for participant roles in a high-pressure situation, for example the captain exhibited high TI and LTN, while a lower-placed flight crew member (jumpseat captain) showed high values for ITR in agreement with close following of orders. In future work the metrics could be used in a population-wide study where metric scores from multiple conversations could be compared. For example, a comparison of conversations by a control group with conversations from a group with a known language disorder could generalize how the language disorder affects topic use patterns within the conversations as a group.

The eight MPR metrics discussed in this paper were chosen for their ability to span a wide space of possible combinations relating to topic use, and the analyses highlighted their correspondence with known qualities in the three transcripts under study. These metrics could be extended to analysis of other transcripts to obtain similar insights. Additionally, primitives can be combined in an unlimited number of ways to create new metrics that highlight other topic usage patterns. Although not demonstrated here, the metrics could enable both real-time monitoring and post-hoc analysis of human communication from many sources, including web forums, email, command and control situations, planning meetings, negotiations, and training scenarios.

ACKNOWLEDGMENT

The authors would like to thank Dr. B. Watson, The University of Queensland, for providing expert assessment of data used in the paper.

REFERENCES

[1] Atlas.ti Qualitative Data Analysis Software, [Online]. Available: http://www.atlasti.com/
[2] Cockpit Voice Recorder Transcript of the July 19, 1989 Emergency Landing of a United Airlines DC-10-10 at Sioux Gateway Airport, IA (SUX), USA [Online]. Available: http://www.tailstrike.com/190789.pdf
[3] NVivo Qualitative Data Analysis Software [Online]. Available: http://www.qsrinternational.com/
[4] D. Angus, A. Smith, and J. Wiles, "Conceptual recurrence plots: Revealing patterns in human communication," *IEEE Trans. Visual. Comput. Graphics*, 2012, to be published.
[5] R. F. Bales, *Interaction Process Analysis*. Cambridge, MA: Addison-Wesley, 1950.
[6] A. J. Cañas, G. Hill, R. Carff, N. Suri, J. Lott, T. Eskridge, G. Gómez, M. Arroyo, and R. Carvajal, "Cmaptools: A knowledge modeling and sharing environment," in *Proc. Concept Maps: Theory, Methodol., Technol., Proc. 1st Int. Conf. Concept Mapping*, A. Cañas, J. Novak, and F. González, Eds., 2004, pp. 125–133.
[7] K. W. Church and J. I. Helfman, "Dotplot: A program for exploring self-similarity in millions of lines of text and code," *J. Comput. Graph. Statist.* , vol. 2, no. 2, pp. 153–174, 1993.
[8] R. Dale and M. J. Spivey, "Unraveling the dyad: Using recurrence analysis to explore patterns of syntactic coordination between children and caregivers in conversation," *Lang. Learn.*, vol. 56, no. 3, pp. 391–430, 2006.
[9] J.-P. Eckmann, S. O. Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," *Europhys. Lett.*, vol. 5, pp. 973–977, 1987.
[10] J. Foote, "Visualizing music and audio using self-similarity," in *Proc. 7th ACM Int. Conf. Multimedia*, 1999, pp. 77–80.
[11] C. Gallois, T. Ogay, and H. Giles, "Communication accommodation theory: A look back and a look ahead," in *Theorizing About Intercultural Communication*, W. Gudykunst, Ed. Thousand Oaks, CA: Sage, 2005, pp. 121–148.
[12] J. Godfrey, E. Holliman, and J. McDaniel, "Switchboard: Telephone speech corpus for research and development," in *Proc. ICASSP '92*, 1992, pp. 517–520.
[13] J. I. Helfman, "Similarity patterns in language," in *Proc. IEEE Symp. Vis. Lang.*, 1994, pp. 173–175.
[14] J. R. Hobbs, "On the coherence and structure of discourse," Center for the Study of Language and Information, Stanford Univ., 1985, Tech. Rep. No. CSLI-85-37.
[15] H. Hung, Y. Huang, G. Friedl, and D. Gatica-perez, "Estimating the dominant person in multi-party conversations using speaker diarization strategies," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, 2008, pp. 2197–2200.
[16] V. K. J. Wolkowicz and S. Brooks, "Midivis: Visualizing music pieces structure via similarity matrices," in *Proc. Int. Comput. Music Conf. (ICMC '09)*, 2009.
[17] J. Morris and G. Hirst, "Lexical cohesion, the thesaurus, and the structure of text," *Comput. Linguist.* , vol. 17, no. 1, pp. 21–48, Mar. 1991.
[18] F. Orsucci, K. Walter, A. Giuliani, C. L. W. Jr., and J. P. Zbilut, "Orthographic structuring of human speech and texts: Linguistic application of recurrence quantification analysis," *Int. J. Chaos Theory Applicat.*, vol. 4, no. 2–3, pp. 21–28, 1999.
[19] D. Roter and S. Larson, "The roter interaction analysis system (rias): Utility and flexibility for analysis of medical interactions," *Patient Educat. Counsel.*, vol. 46, no. 4, pp. 243–251, 2002.
[20] H. Sacks, *Lectures on Conversation*. Oxford, U.K.: Blackwell, 1992.
[21] G. Salton, *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Reading, MA: Addison-Wesley, 1989.
[22] M. Scott, *WordSmith Tools Version 4*. Oxford, U.K.: Oxford Univ. Press, 2004.
[23] A. E. Smith and M. S. Humphreys, "Evaluation of unsupervised semantic mapping of natural language with leximancer concept mapping," *Behav. Res. Meth.*, vol. 38, no. 2, pp. 262–279, 2006.
[24] B. Timberg, *Television Talk: A History of the TV Talk Show*. Austin, TX: Univ. of Texas Press, 2002.
[25] C. L. Webber Jr. and J. P. Zbilut, "Dynamical assessment of physiological systems and states using recurrence plot strategies," *J. Appl. Physiol.*, vol. 76, no. 2, pp. 965–973, 1994.
[26] C. L. Webber Jr. and J. P. Zbilut, *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences Web Book*. Arlington, VA: National Science Foundation (U.S.), 2005, ch. Recurrence quantification analysis of nonlinear dynamical systems, pp. 26–94.
[27] J. Wise, J. Thomas, K. Pennock, D. Lantrip, M. Pottier, A. Schur, and V. Crow, "Visualizing the non-visual: Spatial analysis and interaction with information from text documents," in *Proc. Inf. Visual.*, Oct. 1995, pp. 51–58.
[28] Zapruder's Other Films Pty. Ltd. ABC Enough Rope Program, 2009 [Online]. Available: http://www.abc.net.au/tv/enoughrope/
[29] J. P. Zbilut and C. L. Webber Jr., "Embeddings and delays as derived from quantification of recurrence plots," *Phys. Lett. A*, vol. 171, no. 3–4, pp. 199–203, 1992.

Daniel Angus received the B.S./B.E. double degree in research and development, and electronics and computer systems and the Ph.D. degree in computer science from Swinburne University of Technology, Melbourne, Australia, in 2004 and 2008, respectively.

He is currently a Lecturer in the School of Information Technology and Electrical Engineering and School of Journalism and Communication, University of Queensland, Brisbane, Australia. His research focuses on the development of text analytic techniques, with a specific focus on techniques for analyzing human discourse.

Andrew Smith received the B.Sc. (Honors) and Ph.D. degrees in physics and the M.Phil. degree in information science in 2002, from the University of Queensland (UQ), Brisbane, Australia, in 1987, 1993, and 2002, respectively.

He is the creator of Leximancer, a software application for text analysis. He is currently Leximancer Chief Scientist, Adjunct Researcher within the School of Information Technology and Electrical Engineering at UQ, and Senior Research Officer within the Institute for Social Science Research (ISSR) in the UQ Faculty of Social and Behavioral Sciences. His research focuses on developing new ways to visualize and quantify the temporal dynamics of communication.

Janet Wiles (M'11) received the B.Sc. (Hons I) and the Ph.D. degrees in computer science from The University of Sydney, Sydney, Australia, in 1983 and 1989, respectively.

She is a Professor of Complex and Intelligent Systems in the School of Information Technology and Electrical Engineering at The University of Queensland, Brisbane, Australia, and Director of the Thinking Systems Project. Her research program involves using computational modeling to understand complex systems with particular applications in biology, neuroscience, and cognition.