# Classification of Patient Case Discussions Through Analysis of Vocalisation Graphs

Saturnino Luz
Department of Computer Science
School of Computer Science and Statistics
O'Reilly Institute
Trinity College Dublin
Ireland
+353(1)869-3686
luzs@acm.org

Bridget Kane
Department of Computer Science
School of Computer Science and Statistics
O'Reilly Institute
Trinity College Dublin
Ireland
+353(1)869-2381
kanebt@cs.tcd.ie

## ABSTRACT

This paper investigates the use of amount and structure of talk as a basis for automatic classification of patient case discussions in multidisciplinary medical team meetings recorded in a real-world setting. We model patient case discussions as vocalisation graphs, building on research from the fields of interaction analysis and social psychology. These graphs are "content free" in that they only encode patterns of vocalisation and silence. The fact that it does not rely on automatic transcription makes the technique presented in this paper an attractive complement to more sophisticated speech processing methods as a means of indexing medical team meetings. We show that despite the simplicity of the underlying representation mechanism, accurate classification performance (F-scores: $F_1 = 0.98$, for medical patient case discussions, and $F_1 = 0.97$, for surgical case discussions) can be achieved with a simple $k$-nearest neighbour classifier when vocalisations are represented at the level of individual speakers. Possible applications of the method in health informatics for storage and retrieval of multimedia medical meeting records are discussed.

## Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing; H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems; I.5.2 [**Pattern Recognition**]: Classifier design and evaluation

## General Terms

Human Factors


*ICMI-MLMI'09,* November 2–4, 2009, Cambridge, MA, USA.


## Keywords

Patient Case Discussions, Medical team meetings, Electronic medical records, Language and action patterns

## 1. INTRODUCTION

Multidisciplinary medical team meetings have become an established practice in many hospitals. These are meetings in which a group of experts from various disciplines come together to discuss patient cases and make patient management decisions which have broader implications with respect to a number of processes. Such discussions generate a wealth of information that is not at present captured in traditional medical records. Since medical team meetings are increasingly taking place in rooms fitted with teleconferencing equipment and other audio-visual aids [23], digital recording of entire discussions is becoming a distinct possibility. However, the usefulness of audiovisual databases of meetings is dependent on how effectively their contents can be accessed. This has been evidenced by research in the area of "meeting browsing" [36, 37, 4, 3] and the scale of recent efforts to create meeting corpora [5, 31, 17]. This paper addresses an aspect of content retrieval that might help provide effective structured access to such resources, namely, classification of case discussions with respect to patient management categories. We introduce a classification technique based on "content-free" structures employed in social psychology and other disciplines [9, 16]. Although content-free analysis has been applied to the characterisation of dialogues in terms of descriptive statistics [9, 34] and, more recently, to meeting classification [18], this is, to our knowledge, the first time vocalisation graphs have been used in conjunction with a machine learning method to detect meeting type. We show that these structures can produce accurate classification results without requiring sophisticated speech processing or external sources of contextual information. We interpret these results to mean that there is a degree of autonomy between the structure of interactions at the meetings we studied and their linguistic content, both of which can be exploited in providing access to meeting records.

The paper is organised as follows. In section 1.1 we outline the nature of multidisciplinary medical team meetings, their role in the patient management process, their potential as information sources in educational and organisational

contexts, and the challenge of creating and integrating meeting records into existing systems. Section 1.2 introduces the theoretical background from which our approach to data representation derives, tracing its origins back to early research in social psychology, and reviews recent related work from the meeting analysis literature. This is followed by a formal statement of the classification problem and definition of the main concepts used in the construction of vocalisation graphs. The data gathering methodology is then presented, along with general statistics descriptive of the classes of case discussion targeted by our method. Section 4 describes the classification algorithm adopted and reports on evaluation results for two data representation schemes. The paper concludes with a discussion of these results, their practical implications to health informatics and, to a lesser extent, their theoretical implications with respect to the discipline of interaction analysis.

## 1.1 Multidisciplinary Medical Team Meetings

The need to develop more efficient and effective health services guided by evidence-based best practice is driving change in healthcare structures. Multidisciplinary medical team working that includes regular meetings (MDTMs) is believed to contribute to this goal. Thus, the practice of holding MDTMs has become established in healthcare, especially for cancer patient management. From their origins as educational and teaching fora MDTMs are becoming, for many, an important and integral part of the patient management process. In tandem with changing structures is recognition of the necessity for access to appropriate information for the development of best-practice guidelines. These developments pose new challenges for those designing technology support and information systems appropriate to the needs of healthcare staff.

Typically, several specialists (radiologists, pathologists, physicians, medical and radiation oncologists, and surgeons) bring their experience, knowledge and patient data to the meeting. An MDTM is structured as a sequence of patient case discussions (PCDs). A PCD involves questioning among the specialists, and explanation or elaboration of detail; opinions are exchanged and perhaps the initial findings revised on the basis of information presented and discussed. After the diagnosis and disease stage of the patient is clarified, a decision or recommendation is agreed on the best next step in the management of the patient. The potential value of records of these meetings has been recognised, from a number of perspectives. They could serve as a rich material for teaching and experience building purposes. They are useful venues for the collection of information for audit and evidence-based studies. They recommend patient management and ideally should be incorporated into the patient chart or record. This latter requirement is not easily achieved within our existing model of the patient centred record [13].

This paper considers a proposed meeting record which would include the discussion of patient details, the considerations given to alternative or differential diagnoses, the weighing of options for the patient in the light of current evidence, annotation of artifacts, reasoning and the decisions reached (along with any dissent). Rather than look to traditional database models for storage and retrieval of these data, we consider an electronic capture of meeting proceedings – including audio, video and annotated clinical images – and the automatic indexing of PCDs for later analysis and review. Our focus in this paper is on technical issues of indexing and classification. Issues regarding the acceptability of this form of record of multidisciplinary medical team meeting proceedings by users and meeting participants, security, confidentiality and dependability are discussed elsewhere [20], and have been addressed in ongoing research. The broader goal of this ongoing work is to identify regularities in the way MDTM participants utilise the resources of the complex social and material environment within which they operate. In this paper, however, we specifically investigate how regularities in the participants' conversational behaviour might be exploited for classification of meeting segments with respect to type of patient case under discussion.

Patient cases under discussion at MDTMs can be generally categorised as either medical or surgical. In both cases, discussion typically opens with a presentation of the patient's symptoms and clinical findings (including endoscopy, for medical cases), followed by the demonstration of relevant radiological and pathological images by a radiologist and pathologist respectively. A discussion follows on the significance of the findings and, considering current clinical practice guidelines, participants reason and make a decision on the next step in the patient's management. From an organisational perspective, the main difference between medical and surgical cases is that surgical patients usually have a medical history. In a medical case discussion, clarification of the diagnosis is often protracted, and in surgical cases some level of medical assessment has already taken place.

## 1.2 Related Work

Our initial studies of MDTMs were influenced by ethnographic methods of interaction analysis and by social psychology research on group behaviour. Ethnographic methods and conceptual models of interaction have been increasingly employed in the area of human-computer interaction [8, 29], usually focusing on group characteristics and the context and nature of the tasks, processes and outcomes of cooperative activity. Although the origins of data-intensive studies of group interaction date back to a system of categories introduced by Bales [2], such studies have been greatly facilitated by the increasing availability of audiovisual recording technology which allows investigation of detail not possible with manual methods of data collection [19]. In this paper, we focus on paralinguistic features of group interaction, following a method based exclusively on the amount and structure of speech. This framework was initially developed for analysis of two-party dialogues in psychopathology research [16] and later extended to the study of interaction in small groups [9]. It builds on a technique of content-free analysis that uses turn-taking matrices as a way of summarising conversational history.

Since its introduction, the idea of abstracting away from content has attracted the attention of researchers due to a combination of practical and theoretical factors. From a practical perspective, content-free analysis can be reliably automated, requiring no transcription or any form of human annotation, thus enhancing the researcher's ability to collect substantial amounts of data [16]. From a theoretical perspective, its appeal includes the elimination of subjective factors necessarily associated with human annotation, and the possibility of describing conversation and meetings

as Markov processes [16, 9], whose mathematical properties are well understood. Furthermore, if as argued by Sacks et al. [33], turn-taking should be considered as a central phenomenon in its own right, a content-free approach might help identify those aspects of the system which operate independently of conversational context.

Due to the nature of its theoretical background, content-free analysis has been employed mainly in the study of social aspects of group interaction which, within computer science, have traditionally found application in the area of computer-supported cooperative work (CSCW). CSCW researchers use vocalisation statistics (mean duration, amount of simultaneous speech, length of pauses, entropy, etc) as dependent variables in assessing the effect of certain technologies (e.g. videoconferencing) on human communication [7, 34]. We depart from this tradition by employing content-free structures not simply as a source of descriptive statistics, but inductively, to categorise patient case discussions.

In terms of automatic analysis of speech and meeting data, the PCD categorisation task resembles a topic identification task. Topic identification based on recorded speech was initially investigated in connection with call routing [12] and classification of audio messages on the switchboard corpus [25]. More recently, several variants of topic identification methods have been applied to the analysis of meeting data. These include, among other things, detection of group actions [24], dialogue acts [10] and salient events such as "decisions" [15]. A related task is that of segmenting meetings recordings into topics [14, 28, 22]. In contrast to the method presented in this paper, many of the techniques employed in dealing with these issues rely on automatic speech recognition transcripts, usually combined with a text-tiling algorithm. A promising alternative to relying on lexical information is the use of phonetic transcriptions as the basis for (unsupervised) segmentation [14]. The importance of content-free features to information access has, however, been acknowledged in previous research on meeting browsing. Modelling of speaker activity changes and meeting structuring based on non-lexical features was proposed in [30] and, more recently, a technique for classification of meetings as "cooperative" or "competitive" based on acoustic features has been presented [18]. The latter is somewhat similar to the method described in this paper in that it is also inspired by social psychology research. However, it uses turn-structure statistics as its feature set, while our method employs a graph-based representation which retains more of the turn-structure itself. Furthermore, while [18] targets categories closely associated to group dynamics (competitiveness and cooperativity), our study aims to identify categories related to meeting content (type of medical discussion, etc).

## 2. PROBLEM DEFINITION

Given a corpus of MDT patient case discussions annotated for vocalisation patterns and case types, our goal is to learn a classification function capable of assigning the right case type to a previously unseen case discussion described in terms of vocalisation patterns. The algorithm will operate on a set $\mathcal{C}$ of abstract descriptions of patient case discussions, which we will refer to simply as *PCD descriptions*. Each PCD description will be represented as a *vocalisation graph*, i.e. a directed graph $G = (V, E)$ where $V$ is a finite set of vertices or nodes and $E$ a binary relation on $V$. Elements of $V$ are labelled by pairs $(s, p(s))$ representing the probability $p_s$ that the dialogue is in state $s$ (e.g. a vocalisation or a silence) at any given instant. Edges are labelled by conditional probabilities. A probability $p(t|s)$ labelling an edge corresponds to the likelihood that a dialogue state $t$ (the terminal vertex) immediately follows dialogue state $s$ (the initial vertex).

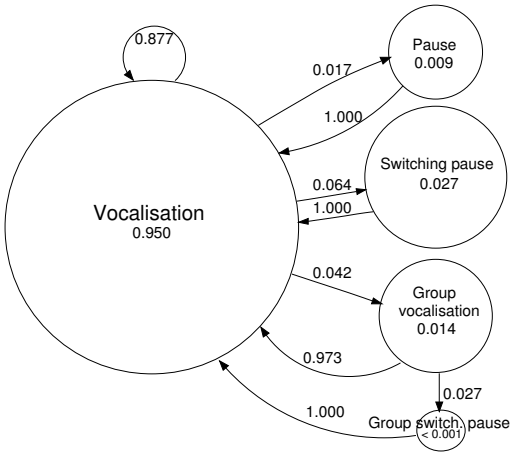


**Figure 1: Sample aggregated vocalisation graph**

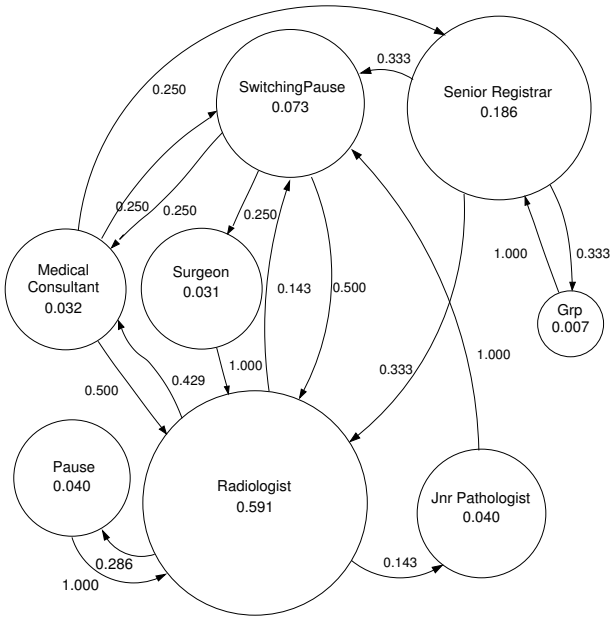


**Figure 2: Sample individuated vocalisation graph**

Case representations can take the form of *aggregated vocalisation patterns* or *individuated vocalisation patterns*. The former corresponds to the "GroupTalk" model proposed in [9], which records no distinctions between individual speakers. The latter is our augmented vocalisation graph model, where each state corresponds to the vocalisations of an individual speaker (or, in this paper, to their specialist roles at MDTMs). Examples of each representation scheme are shown in Figures 1 and 2. Both representations can be regarded as Markov chains, based on periodic sampling of the

**Table 1: Summary of conversational analysis statistics for surgical and medical case discussions**

| Parameter | $p <$ | medical | surgical |
|---|---|---|---|
| Entropy ($H$) | 0.02 | 2.420 | 2.029 |
| Length of vocalisations (in seconds) | 0.01 | 6.581 | 8.929 |
| No. of vocalisations/minute | 0.01 | 10.936 | 7.025 |
| Length group vocalisation (in seconds) | non-sig. | 1.764 | 1.455 |
| No. of Vocalisations per person | non-sig. | 3.270 | 3.618 |
| No. speakers per case | non-sig. | 8.333 | 7.250 |
| Pct. of silence per case | non-sig. | 4.883 | 3.371 |
| Participation ratio | non-sig. | 0.397 | 0.429 |

audio stream. Our representation scheme, however, differs from the GroupTalk matrices in one respect. In [9] transitions automatically triggered at the end of a 0.25-second interval are always represented, resulting in a large number of self-transitions, which tend to dominate the distributions. However, since all vocalisation graphs consist of single recurrent, aperiodic chains, aggregate probabilities for individual vocalisation events correspond to the node's steady state. Since these variables (steady state and self-transition probabilities) are strongly correlated, we unclutter our graphs by simply eliminating self-transitions, labelling nodes with the stationary distribution, and normalising the remaining transition probabilities. This approach reduces the undesirable effect of an arbitrarily chosen sampling interval on the probabilities. Furthermore, it provides a clearer distinction between *amount* and *structure* of talk, two concepts seen as fundamental in content-free analysis [16, 9]. Amount of talk is uniquely encoded by node labels, and conversation structure is represented by edges and their labels. Thus, Figure 1 represents a case discussion in which vocalisations take up 95% of the time, and most vocalisations (87.7%) are followed up by other vocalisations. Figure 2 shows the radiologist as the dominant speaker, holding the floor for nearly 60% of the time and interacting mainly with the medical consultant. The terms "vocalisation", "silence", and other terms that describe interaction states in vocalisation graphs are specified in Definition 1.

*Case types* are selected from a set $\mathcal{T}$ of category labels. Presently, we distinguish two types of cases: *medical* and *surgical* cases. The learnt classification function $\widehat{\Phi} : \mathcal{C} \rightarrow \mathcal{T}$ will thus attempt to approximate a binary target function $\Phi : \mathcal{C} \rightarrow \mathcal{T}$ defined by human judgement codified through annotation.

## 3. DATA COLLECTION

Digital video recording was taken of meeting sessions at a major teaching hospital in a room equipped with specialised audio and video capture equipment for teleconferencing. A total of 19 MDTMs, or over 28 hours of meeting data (audio and video) have been collected, containing 346 PCDs, altogether. The original purpose of data collection was to investigate the diagnosis and decision making processes of multidisciplinary medical teams [20] within an interaction analysis framework [19]. For the study reported in this paper, a dataset of 54 PCDs were segmented and annotated using the ELAN Linguistic Annotator [26]. Definition 1 specifies the relevant vocabulary used in our annotation scheme.

DEFINITION 1. *We distinguish the following types of* dialogue states*:*

Vocalisation*: the length of time that a speaker "has the floor". A speaker takes the floor when they begin speaking to the exclusion of everyone else and speak uninterruptedly without pause for at least 1 second. The vocalisation ends when a silence, another individual vocalisation or a group vocalisation begins.*

Group vocalisation *occurs when two or more individuals are speaking together. The group vocalisation ends when any individual is again speaking alone, or a period of silence begins.*

Silence *represents quiet periods of over 0.9 seconds between vocalisations. Silences can further be classified as:* pauses*,* switching pauses*,* group pauses *and* group switching pauses*. Pauses are silences preceded and followed by the same speaker. Switching pauses are silences between two different speakers. A group pause is a silence between two group vocalisations. A group switching pause is a silence between a group vocalisation and an individual vocalisation.*

Although other meeting corpora exist [31, 17] which are larger and annotated to a finer level of detail, our corpus is unique in that it was collected *in situ*, under naturalistic conditions (one of the authors became a member of the multidisciplinary team and attended their MDTMs as part of an extensive ethnographic study conducted over a period of three years [20]) while the meeting participants were engaged in a complex real-world task.

Annotation followed the methodology described in [9, 34] and therefore focused mainly on the amount and structure of speech activity. The metadata created for this set of 54 PCDs are in fact much more detailed, containing information about artifacts employed during the meeting, use of informal language, roles and other annotation layers. For the purposes of this paper, however, only speaker activity is considered. The dialogue states specified in Definition 1 are similar to the ones used in [34], with an adjustment to the minimal duration of a vocalisation. Our definition of *silence* is similar to the concept of *switching pauses* described in [9]. One could also define simplified notions of *turns* as sequences of vocalisations and pauses, and analogously *group turns* as sequences of group vocalisations and group pauses. However, we chose to avoid the term "turn" altogether, as it is use in conversation analysis [33] in a different and more complex sense. Since a satisfactory account of turn-taking seems beyond what is currently achievable by automatic annotation, turns and group turns (in the sense of [33]) are

not employed in the above described dialogue representation scheme.

While the data used in our experiments have been hand annotated, all information needed to construct both kinds of graphs can, in principle, be automatically extracted from recorded audio through existing signal processing techniques [6, 32]. However, speech detection and, specially, diarisation techniques usually have high error rates, unless the audio is collected under favourable conditions, which is unlikely in MDTMs. An interesting alternative would be the use of "sociometric badges" [27]. Since full speech recording and transcription is not needed (nor, in fact desirable, due to privacy concerns), these badges could be worn individually by MDTM participants to generate the vocalisation patterns required by our method.

Table 1 summarises the conversational analysis statistics for the two types of patient case discussions in our dataset. We used Welch's $t$-test for all comparisons. The first row reflects the degree of "disorder" in PCDs, as expressed by they *entropy*. Entropy is calculated for a probability distribution $P$ of vocalisations by $n$ speakers, i.e. the steady state probabilities for the nodes of a vocalisation graph, where each $p_i$ corresponds to the probability that speaker $s_i$ is speaking at a given time during a case discussion, as set out in equation (1).

$$H = \sum_{i}^{n} p_i \log \frac{1}{p_i} \tag{1}$$

Rows two through to seven show the mean values for a number of parameters typically considered in CSCW studies [34, 20]. These include: mean duration and frequency of vocalisations, mean number of vocalisations per participant during a case discussion, typical number of speakers participating in each case discussion and mean duration of intervals of silence. The last row contains the mean values for a metric we call *participation ratio*. The participation ratio of a meeting participant is defined as the ratio between the number of case discussions they took active part in and the total number of cases discussed. The figures for mean participation ratio in Table 1 were calculated according to equation (2), where $\mathcal{C}_i$ represent the set of cases in which speaker $s_i$ produced at least one vocalisation and the sets of case descriptions $\mathcal{C}$ is appropriately restricted according to case type.

$$\sum_{i}^{n} \frac{|\mathcal{C}_i|}{n|\mathcal{C}|} \tag{2}$$

Participation ratio figures are meant to summarise variability in the composition of the groups across case discussions. Table 1 indicates a high degree of variation in both medical and surgical meetings, showing that a speaker will on average take part in only around 40% of all case discussions. The slightly higher participation ratio for surgical cases reflects the fact that more specialities may have input into a PCD on a surgical patient than on a medical patient.

# 4. CLASSIFICATION EXPERIMENTS

The statistics in Table 1 seem to indicate a qualitative difference between surgical and medical patient case discussions. In particular, speech distribution patterns, vocalisations and turn taking frequency differ significantly between the two types of cases. Surgical case discussions appear to be slightly more structured, with less turn taking and longer vocalisations suggesting a predominance of case *presentation*, as opposed to *discussion*. However, these differences are small, and there is considerable indeterminacy with respect to the other conversational parameters. The question therefore is: can case discussions be automatically classified based only on vocalisation data of the sort used in interaction analysis and social psychology? In order to investigate this question we have performed a few experiments employing a simple machine learning methods, namely instance-based learning [1], for classification. These experiments made use of the two kinds of data representation depicted in Figure 1.

## 4.1 Classification with Aggregated Vocalisation Graphs

A $k$-nearest neighbours ($k$-NN) classifier [1] was employed. Classifier training consisted simply of storing instances and their true classification values $\Phi(c)$. The classification function $\widehat{\Phi}$ was implemented as shown in Algorithm 1. The distance metric $dist(c_i, c_j)$, in line 6, can be implemented in different ways. Graph matching scores would perhaps be the most appropriated approach. However, for simplicity, we chose Euclidean distance for identifying the $k$ nearest neighbours to the query instance. Thus, in this first experiment we encode instances of aggregated vocalisation graphs, $\mathcal{C}$, as 42-tuples of the form $c = (a_1, \ldots, a_{42})$ where the first six attributes correspond to probabilities associated with the dialogue states identified in Definition 1 (e.g. the probability of occurrence of a group vocalisation) and the remaining attributes represent transition probabilities between pairs of dialogue states. The value of $k$ in all experiments was automatically determined through hold-one-out cross validation.

ALGORITHM 1. *k-NN classification*

```
1  input: c_q     // a query instance
2         Tr      // set of training instances
3         k       // number of neighbours
4   var:  Knn // a set
5   for ( i in 0..k )
6     if ( c_i == arg min_{c_j ∈ Tr} dist(c_i, c_j) )
7       Tr ← Tr \ c_i
8       Knn ← Knn ∪ c_i
9   return
10    Φ̂(c_q) ← arg max_{v∈V} w_i Σ_{i=1}^{k} δ(v, Φ(c_i))
11  // where w_i is a weight assigned c_i,
12  // and δ(a,b) the Kronecker delta.
```

A 3-NN classifier with all weights set to 1 was tested in a ten-fold cross validation experiment. That is, for each fold, the classifier was trained on nine tenths of the case base and tested on the remaining one tenth. Each fold was selected by stratification so as to reflect the balance of target classes in the dataset. This process was repeated ten times. Results were then compiled and averaged. Contrary to our initial expectations, classification results were rather poor. For medical case discussions, precision (the ratio of true positives to the number of selected cases) was approximately 0.57 and recall (the ratio of true positives to the number of target cases) was approximately 0.5. The classifier performed slightly worse on surgical cases, with precision and recall figures of 0.37 and 0.44 respectively.

In order to improve on these results we supplemented the PCD feature set with individuated vocalisation data,

as described in Section 2. This consisted basically in breaking down vocalisation nodes into their constituent speakers and encoding the respective vocalisation transitions, as illustrated in Figure 2.

## 4.2 Classification with Individuated Graphs

Speaker information added to the representation consisted of unique identifiers assigned to each meeting participant. If a graph matching approach were adopted, one might be able to do without speaker identifiers and still employ the individuated representation. This would be similar to treating the graphs as social networks, as used for speaker role identification in [35], for instance. However, since the number of active participants varies from PCD to PCD this approach would involve substructure search, which would result in a computationally costlier classification algorithm.

In terms of the Euclidean distance approach adopted in the previous experiment, adding speaker information caused a dramatic increase in dimensionality, as each speaker who participated in at least one case discussion and their vocalisation transitions had to be represented. This time a 4-NN classifier was used and the contribution of each neighbour was weighted according to the inverse square of its distance to the query. The experimental setup otherwise remained the same. The results obtained this way were markedly better. All surgical cases were classified as such (recall = 1) and precision was approximately 0.73. Conversely, medical cases had maximum precision and 0.75 recall. This level of accuracy, however, is still unsatisfactory for most applications in this domain.

Nearest-neighbour classifiers are known to be very sensitive to irrelevant attributes (the so-called *curse of dimensionality* problem). With as large a feature set as the one used in the previous experiment, one has good reasons to suspect that classification accuracy is being negatively affected by the presence of irrelevant features. Although inverse distance weighting helps alleviate the problem, the adverse effect of high dimensionality cannot be fully eliminated through weighting. Feature selection (FS) techniques were therefore applied in order to reduce dimensionality. The feature selection method chosen for our next experiment consisted of discretizing [11] the attribute set $A = \{a_1, ..., a_{|A|}\}$ and ranking its attributes according to their chi square score, $\chi^2_a$, as defined in equation (3), where $m$ is the number of intervals resulting from discretization, $n$ is the number of classes (2, in this case) and $E_{ij}$ the expected frequency of $a_{ij}$. The attribute set was thus reduced to 34 attributes.

$$\chi^2_a = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{(a_{ij} - E_{ij})^2}{E_{ij}} \quad (3)$$

A ten-fold cross validation experiment similar to the ones described above was performed, and this time high accuracy case type classification (F-score above 97%) was achieved with a distance-weighted 5-NN classifier. All surgical cases were classified as such (100% recall) and no medical cases were misclassified (100% precision). Results are summarised in Table 2, where the last column records the F-score for each classification task with equal weight assigned to precision and recall (i.e. their harmonic mean). The values of $k$ for all experiments were selected through cross validation so that $k = 4$ and $k = 5$ represent the optimal values for the individuated data set with and without FS, respectively.

# 5. DISCUSSION

Content-free analysis has been advocated by social psychologists as a way to investigate the dynamics of interaction in dialogues [16] and collaborative meetings [9]. The results above show that content-free vocalisation graphs can also be useful as a means of data representation for content indexing and classification. As can be seen in the last two rows of Table 2, both medical and surgical cases can be accurately classified once feature selection has been applied. Also noteworthy is the fact that classification performance for individuated diagrams is robust to speaker variation. The low participation ratio reported in Table 1 suggests that the composition of the groups discussing each case varies considerably. The classification mechanism, however, is able to generalise over group membership differences, which indicates that the generalisation is over patterns of participation by specialist roles, rather than individuals.

The first experiment, in which we attempted to classify patient case discussions based on aggregated diagrams of the kind used in the GroupTalk model introduced in [9], indicates that vocalisation and silence patterns alone do not suffice in distinguishing between medical and surgical case discussions. However, the second and third experiments, which used individuated vocalisation graphs, showed that accurate classification can be achieved at little extra cost. These findings have theoretical implications with respect to the methodology used in interaction analysis as well as practical implications for medical informatics.

An in-depth discussion of the implications of the results above to interaction analysis methodology would fall outside the scope of this paper. We will, however, make some general observations in that regard. First, the low discriminating power exhibited by aggregated vocalisation diagrams appears to call into question their usefulness as interaction analysis tools. The fact that, in our experiments, aggregated diagrams failed to even indicate differences partially captured by descriptive statistics (Table 1) suggests that these models do not provide the level of detail required by typical interaction analysis problems. In [9], an alternative proposal which consists of assigning each speaker an individual diagram is briefly mentioned in an apparent attempt to address this issue. However such a strategy would not yield easily comparable interaction summaries, except for very small meetings. In contrast, the representational strategy proposed in this paper combines an account of the distribution of speech to the idea of vocalisation patterns as a stochastic process, producing a single, scalable conceptual entity which, in addition to its usefulness to machine learning applications, seems to provide a more appropriate basis for theoretical analysis.

From a medical informatics perspective, the results reported in this paper point to practical applications in the area of production, indexing and retrieval of information from patient case discussion records. As pointed out in Section 1, although the information generated at MDTMs could constitute valuable resources for a number of processes in healthcare, ranging from patient management to teaching, incorporating audio-visual recordings of these meetings into existing models of patient-centred record is far from straightforward. Research in the area of CSCW has recognised the need for information systems in healthcare to take more account of informal discussion, differential and provisional diagnoses for the effective operation of multidisciplinary teams

**Table 2: Summary of $k$-NN classification results for different case representations**

| Representation | FS? | type | $k =$ | precision | recall | $F_1$ |
|---|---|---|---|---|---|---|
| aggregated | no | medical | 3 | 0.57 | 0.5 | 0.53 |
| aggregated | no | surgical | 3 | 0.37 | 0.44 | 0.4 |
| individuated | no | medical | 4 | 1 | 0.75 | 0.86 |
| individuated | no | surgical | 4 | 0.73 | 1 | 0.84 |
| individuated | yes | medical | 5 | 1 | 0.96 | 0.98 |
| individuated | yes | surgical | 5 | 0.94 | 1 | 0.97 |

[21]. Research has also highlighted discrepancies between the presumed role of the electronic medical record (EMR) in achieving service integration and the ways in which medical workers actually use and communicate patient information [13].

Given that MDT meeting participants work under tight time constraints, automatic recording seems to be the only viable approach to data gathering. Recording and storage of multimedia meeting data in digital form have become relatively commonplace in recent years. The challenge consists in finding effective ways to structuring and providing effective access to these data. There has been considerable research interest in the topic of information retrieval from meeting recordings. Most approaches, however, build on speech recognition (see [4] for a survey of recent work in this area) or improved user-interface support for annotation of content. The work presented in this paper is an initial attempt attempt to use content-free interaction analysis techniques as a basis for classification of recorded meeting contents. Although the techniques presented still need to be extended and tested on larger datasets, they show promise as a way of complementing and enhancing existing approaches. In demonstrating that case discussion for medical and surgical cases can be classified in this way, we suggest that new paradigms can be seriously considered for electronic patient medical records that would help support the highly interactive and complex nature of medical work.

# 6. CONCLUSION

The content-free interaction analysis method has yielded promising results when applied to patient case discussions at MDTMs. Representation structures derived from simply tracking duration and source of vocalisations can support accurate classification of medical and surgical case discussion types. Although it is necessary to distinguish among speakers or specialist roles in order for classification to be accurate, the technique is robust to speaker variation.

The results reported here are part of a larger on-going study aimed at understanding the task and process at MDTMs. Our ultimate objective is to identify how technology might be applied in such settings, including the possibility of generating an electronic meeting record with automatic indexing of cases. Indexing of cases discussed at the meeting potentially would allow users to easily retrieve PCDs for teaching and business purposes, including the development of tools for analysis. To this end, we are currently tackling the issue of automatic segmentation of MDTMs into PCDs [22]. We are also carrying out further annotations on the MDTM corpus so that finer distinctions of patient case discussion types can be investigated, and the method presented above can be validated on more challenging classification tasks.

Future work will employ individuated vocalisation diagrams for classification of medical and surgical cases into subtypes according to the time within the patient care pathway when the PCD takes place, whether the the PCD is on a referred case from another hospital or not, and the nature of the patient's clinical presentation. Detection of specific events, such as intervals during which discussion of TNM (Tumour, Nodes, Metastases) disease stage (categorisation) takes place is also been investigated. Finally, we are also experimenting with graph matching approaches to nearest neighbour selection which, as discussed in Section 4.2 might eliminate the need for speaker identification.

# 7. ACKNOWLEDGEMENTS

Our thanks to all the members of the multidisciplinary team at St James's Hospital, Dublin, for their cooperation in this study. This research is funded by Science Foundation Ireland under the Research Frontiers program.

# 8. REFERENCES

[1] D. Aha and D. Kibler. Instance-based learning algorithms. *Machine Learning*, 6:37–66, 1991.

[2] R. F. Bales. *Interaction Process Analysis: A Method for the Study of Small Groups*. Addison-Wesley, Cambridge, Mass., 1950.

[3] S. Banerjee, C. Rose, and A. I. Rudnicky. The necessity of a meeting recording and playback system, and the benefit of topic-level annotations to meeting browsing. In *Proceedings of the 10th International Conference on Human-Computer Interaction (INTERACT'05)*, pages 643–656, 2005.

[4] M.-M. Bouamrane and S. Luz. Meeting browsing. *Multimedia Systems*, 12(4–5):439–457, 2007.

[5] J. Carletta. Unleashing the killer corpus: experiences in creating the multi-everything ami meeting corpus. *Language Resources and Evaluation*, 41(2):181–190, 2007.

[6] S. Chen and P. Gopalakrishnan. Speaker, environment and channel change detection and clustering via the bayesian information criterion. In *Proc. of DARPA Broadcast News Transcription and Understanding Workshop*, 1998.

[7] K. M. Cohen. Speaker interaction: video teleconferences versus face-to-face meetings. In *Proceedings of Teleconferencing and Electronic Communications*, number 189–199, 1982.

[8] A. Crabtree. *Designing Collaborative Systems: A Practical Guide to Ethnography*. Springer-Verlag, 2003.

[9] J. J. Dabbs and R. Ruback. Dimensions of group process: Amount and structure of vocal interaction.

*Advances in Experimental Social Psychology*, 20(123–169), 1987.

[10] A. Dielmann and S. Renals. Recognition of dialogue acts in multiparty meetings using a switching dbn. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(7):1303–1314, 2008.

[11] U. M. Fayyad and K. B. Irani. Multi-interval discretization of continuous-valued attributes for classification learning. In *Proceedings of the 13th International Joint Conference in Artificial Intelligence*, pages 1022–1026. Morgan Kaufmann, 1993.

[12] J. Golden, O. Kimball, M.-H. Siu, and H. Gish. Automatic topic identification for two-level call routing. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, volume I, pages 509–512. IEEE, 1999.

[13] M. Hartswood, R. Proctor, M. Rouncefield, and R. Slack. Making a case in medical work: Implications for the electronic patient record. *Computer Supported Cooperative Work*, 12:241–266, 2003.

[14] P.-Y. Hsueh. Audio-based unsupervised segmentation of multiparty dialogue. In *Procs. of the International Conference on Acoustics, Speech and Signal Processing*, pages 5049–5052, April 2008.

[15] P.-Y. Hsueh and J. D. Moore. Automatic decision detection in meeting speech. In A. Popescu-Belis, S. Renals, and H. Bourlard, editors, *Machine Learning for Multimodal Interaction (MLMI '07)*, volume 4892 of *Lecture Notes in Computer Science*. Springer, 2007.

[16] J. Jaffe and S. Feldstein. *Rhythms of dialogue*. Academic Press, New York, 1970.

[17] A. Janin, D. Baron, J. Edwards, D. Ellis, D. Gelbart, N. Morgan, B. Peskin, T. Pfau, E. Shriberg, A. Stolcke, and C. Wooters. The ICSI meeting corpus. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03)*, volume 1, pages 364–367, 2003.

[18] D. B. Jayagopi, R. Bogdan, and D. Gatica-Perez. Characterising conversationsal group dynamics using nonverbal behaviour. In *Proceedings ICME 2009*, 2009.

[19] B. Jordan and A. Henderson. Interaction analysis: Foundations and practice. *The Journal of the Learning Sciences*, 4(1):39–103, 1995.

[20] B. Kane and S. Luz. Multidisciplinary medical team meetings: An analysis of collaborative working with special attention to timing and teleconferencing. *Computer Supported Cooperative Work (CSCW)*, 15(5):501–535, 2006.

[21] B. Kane and S. Luz. Achieving diagnosis by consensus. *Computer Supported Cooperative Work (CSCW)*, 18(4):357–391, 2009.

[22] S. Luz. Locating case discussion segments in recorded medical team meetings. In *SSCS '09: Proceedings of the ACM SIGMM Workshop on Searching Spontaneous Conversational Speech*, New York, NY, USA, Oct. 2009. ACM Press.

[23] J. McAleer, D. O'Loan, and D. Hollywood. Broadcast quality teleconferencing for oncology. *Oncologist*, (6 (5)):459–462, 2001.

[24] I. McCowan, D. Gatica-Perez, S. Bengio, G. Lathoud, M. Barnard, and D. Zhang. Automatic analysis of multimodal group actions in meetings. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):305–317, March 2005.

[25] J. McDonough, K. Ng, P. Jeanrenaud, H. Gish, and J. Rohlicek. Approaches to topic identification on the switchboard corpus. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, pages 385–388. IEEE Press, 1994.

[26] MPI. ELAN: Eucido Linguistic Annotator. Max Planck Institute for Psycholinguistics, March 2005. http://www.mpi.nl/tools/elan.html.

[27] D. Olguin, B. Waber, T. Kim, A. Mohan, K. Ara, and A. Pentland. Sensible organizations: Technology and methodology for automatically measuring organizational behavior. *IEEE Transactions on Systems, Man, and Cybernetics*, 39(1):43–55, Feb. 2009.

[28] M. Purver, T. L. Griffiths, K. P. Körding, and J. B. Tenenbaum. Unsupervised topic modelling for multi-party spoken discourse. In *ACL '06: Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the ACL*, pages 17–24, Morristown, NJ, USA, 2006. Association for Computational Linguistics.

[29] D. Randall, R. Harper, and M. Rouncefield. *Fieldwork for design: theory and practice*. Springer-Verlag New York Inc, 2007.

[30] S. Renals and D. Ellis. Audio information access from meeting rooms. In *Procs. of the Internacional Conference on Acoustics, Speech, and Signal Processing*, volume IV, pages 744–747. IEEE, 2003.

[31] S. Renals, T. Hain, and H. Bourlard. Recognition and interpretation of meetings: The AMI and AMIDA projects. In *Proc. IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU '07)*, 2007.

[32] D. Reynolds and R. Rose. Robust text-independent speaker identification using gaussian mixture speaker models. *IEEE Transactions on Speech and Audio Processing*, 3(1):72–83, 1995.

[33] H. Sacks, E. A. Schegloff, and G. Jefferson. A simplest systematics for the organization of turn taking in conversation. *Language*, 50(4):696–735, 1974.

[34] A. J. Sellen. Remote conversations: The effects of mediating talk with technology. *Human-Computer Interaction*, 10(4):401–444, 1995.

[35] A. Vinciarelli. Speakers role recognition in multiparty audio recordings using social network analysis and duration distribution modeling. *IEEE Transactions on Multimedia*, 9(6), 2007.

[36] A. Waibel, M. Brett, F. Metze, K. Ries, T. Schaaf, T. Schultz, H. Soltau, H. Yu, and K. Zechner. Advances in automatic meeting record creation and access. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 1, pages 597–600. IEEE Press, 2001.

[37] P. Wellner, M. Flynn, and M. Guillemot. Browsing recorded meetings with Ferret. In S. Bengio and H. Bourlard, editors, *Proceedings of Machine Learning for Multimodal Interaction (MLMI '04)*, volume LNCS 3361, pages 12–21. Springer, 2004.