Generating Verbal Descriptions from Medical Sensor Data: A Corpus Study on User Preferences

Dimitra Gkatzia
Heriot-Watt University
Edinburgh
EH14 4AS
dg106@hw.ac.uk

Verena Rieser
Heriot-Watt University
Edinburgh
EH14 4AS
v.t.rieser@hw.ac.uk

Alexander McSporran
BASP UK Ltd.
Grencoe
PH49 4HR
alexandermcsporran@hotmail.com

Alistair McGowan
University of Glasgow
Glasgow
G12 8QQ
Alistair.McGowan@glasgow.ac.uk

Alasdair Mort
University of Aberdeen
Inverness
IV2 3JH
a.mort@abdn.ac.uk

Michaela Dewar
Heriot-Watt University
Edinburgh
EH14 4AS
M.Dewar@hw.ac.uk

Understanding and interpreting medical sensor data is an essential part of pre-hospital care in medical emergencies, but requires training and previous knowledge. In this paper, we describe on-going work towards a medical decision support tool, which automatically generates textual summaries of underlying sensor data. In particular, we present results from a survey investigating the preferences of individual users and user groups when summarizing medical sensor data. We find that the users' preferences are not necessarily dependent on the user's training level, profession or gender. We therefore use cluster analysis to identify user groups with consistent preferences with regard to 4 different first aid scenarios and 3 types of physiological parameters. In future work, we will utilize these findings to automatically adapt the generated output to personal preferences.

User Preferences, Data-to-text, Decision Support Systems, Health data, Wearable tech

1. INTRODUCTION

In a medical emergency, a patient's survival often depends upon the prompt response and appropriate first aid given by the first person on scene, also known as "bystander". However, only 14% of potential UK bystanders would actually administer the required actions, if they found a person who was not breathing and had no other signs of life.

Up to 150k preventable deaths occur in the UK every year[1] — by proportion of population, 15k of such fatalities occur in Scotland — due in part to the lack of early intervention with appropriate first-aid.

The overall aim of this multidisciplinary project (funded by the Royal Society of Edinburgh) is to develop a medical decision support system which can guide bystanders in real time, helping them to take an active role at an incident, administer appropriate first aid more rapidly and more confidently and provide critical data for the handover of the casualty.

Clinical Decision Support Tools have been designed mostly to assist expert users such as nurses or doctors, e.g. [2]. So far, very few of these systems have focused on providing assistance for non-experts with limited training.

One notable exception is the system developed by the MIME (Managing Information in Medical Emergencies) project[2] [11], which uses non-invasive wireless sensors to monitor and record the patient's vital signs. This allows the bystander to respond more efficiently by focusing on patient care and to collect and save significantly more data than it would be possible when using repeated manual assessments.

---

[1] https://www.sja.org.uk/sja/about-us/latest-news/news-archive/news-stories-from-2010/april/lack-of-first-aid-costs-lives.aspx

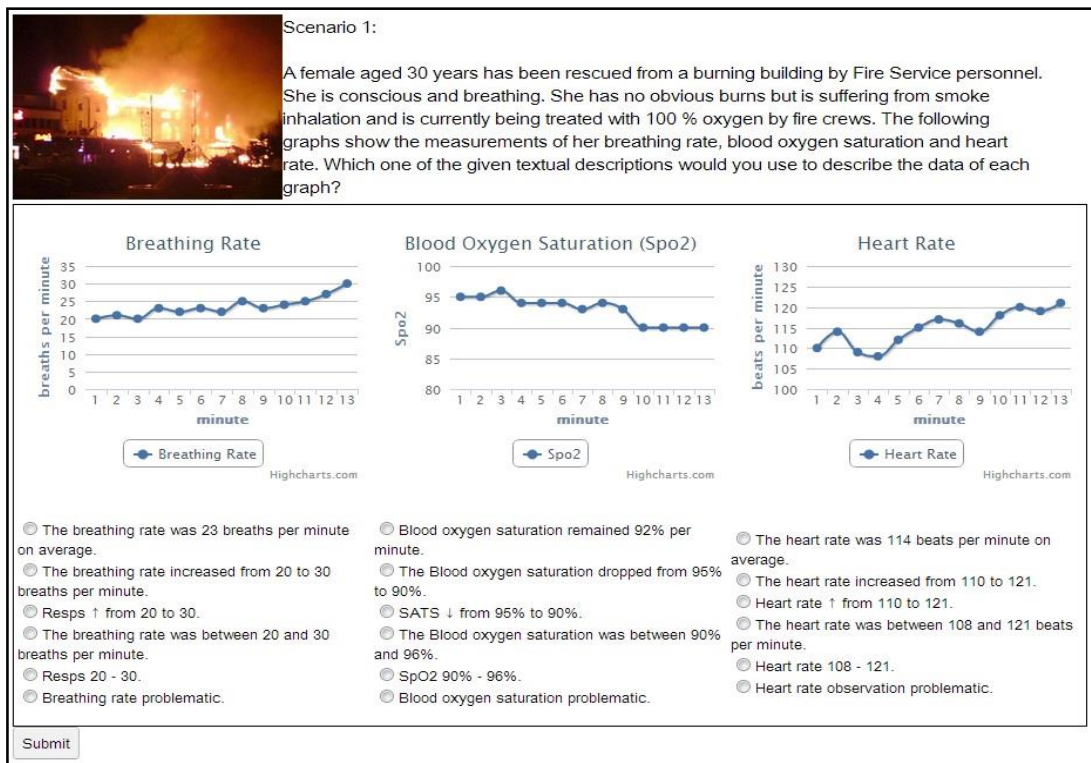[2] http://www.abdn.ac.uk/iahs/research/crh/projects/mime.php

Figure 1: *In the top, a textual description of the event is given. In the middle, the graphs present the processed physiological data and in the bottom 6 different phrases that describe each parameter are given.*

Depending on technology, the saved data can then be passed on to paramedics and medical staff at handover. We consider that such sensors will become commonplace in future healthcare, e.g. placed in rural village medical centers and post offices. In the future, mobile phones will also assist the monitoring of physiological data (a few devices already measure the heart rate).

However, if these sensors are to aid bystanders, then it is vital that the sheer volume of data generated does not become an additional burden for them, i.e. the sensors should reduce information load rather than increase it.

This project investigates the use of Natural Language Generation to automatically translate sensory data into user-adaptive textual summaries. Research with clinicians shows that this type of textual summaries can help to reduce information load on the users [2], and that verbal descriptions of sensor data are more beneficial to decision making than graphical representations of the same data [8]. We extend this previous work in producing summaries which are on the one hand easily understandable by bystanders without any previous first aid training, but at the same time also reduce the cognitive load by adapting to the individual user's preferences.

As a first step towards such a system, we conduct an online survey, collecting data on preferred textual realizations (phrases) for a given set of sensor data within a first aid scenario. Section 2 summarizes previous work in this area. Section 3 describes the data collection framework. Section 4 presents results analyzing the data in terms of user preferences for individual users and user types. Section 5 concludes by describing how the presented results will be used in future work.

2. BACKGROUND & PREVIOUS WORK

Data-to-text generation is a sub-field of Natural Language Processing (NLP) and Natural Language Generation (NLG) in particular, that copes with the task of automatically generating text from non-linguistic input, such as time-series data [10].

Time-series data such as sensor data, often display a complex structure and the identification of useful information is often domain dependent. Humans are able to describe time-series data fluently by using natural language. Their descriptions are a result of a combination of decisions, for instance, they can decide to refer to only unusual fluctuations, to a trend or even adapt their language to the background of their interlocutor.

However, for machines, the task of determining the content to effectively summarize time-series data in natural language remains challenging. In addition, time-series data are usually presented with visualization techniques, such as graphs. Early research has shown that the interpretation of graphs requires expertise [9].

Summarizing time-series data in the medical domain is not a new task. Babytalk [2] is a family of systems developed to support clinicians (BT-nurse) and parents (BT-family) in the Neonatal Intensive Care Unit (NICU). BT-nurse produces textual summaries in the end of a nurse shift in order to assist nurses when changing shifts at NICU [6]. BT-family produces textual reports for parents that their infant is under intensive care, using affective natural language generation [7]. This system is able to adapt its output to the parents' stress level.

As discussed in the previous section, the MIME project produces textual handover reports from physiological sensors, which are passed to paramedics from the first bystander on the scene [11]. Our work extends this previous work, in that we suggest a system that adapts its summaries to user groups with different backgrounds simultaneously, by employing Multi-adaptive Natural Language Generation techniques [4]. Previous data-to-text systems adapt their output to specific user groups, therefore each system can only be used by a particular user group, for instance parents or clinicians, such as BT-family or BT-nurse. In this work, we suggest a methodology that is able to take into account the preferences and the background of different user groups when summarizing data.

Multi-adaptive Natural Language Generation refers to the task of automatically adapting the generated text to preferences of potential users or user groups simultaneously. This approach is able to address the preferences of unknown users, as for instance in a medical emergency, where the level of training of the first person on the scene is unknown. For example, the first person on scene might only have received first aid at work training or s/he might be a medical doctor.

In this work, we investigate potential users of our system and we explore their preferences when summarizing time-series data, with a goal to simultaneously adapt to their preferences. Related work, on a different domain, shows that it is possible to simultaneously adapt to two user groups [3], by modeling their preferences as objective functions.

3. DATA COLLECTION

The aim of the data collection is to create a corpus, which will allow us to study and define the preferences of different users. Here, we analyze user preferences when describing sensor data that measure 3 physiological conditions: Breathing Rate, Blood Oxygen Saturation (SpO2) and Heart Rate. The scenarios are based on data from the MIME project that are generously shared with us.

A first aid trainer of the British Association of Ski Patrollers (BASP) identified 6 potential textual descriptions / phrase templates of describing time-series data:

1) Referring to the *average*,

2) Referring to the *trend* in a *verbose* way,

3) Referring to the *trend* in a *succinct* way,

4) Referring to the *range* of values observed in a *verbose* way,

5) Referring to the *range* of values observed in a *succinct* way and

6) *Inference* from the data.

Figure 1 shows an example scenario, the graphs that depict the sensor data and 6 corresponding textual summaries for each graph. Note that the order of summaries in Figure 1 corresponds to the template types defined above.

The data collection is conducted online and 69 people are recruited via email (34 males and 35 females). Their background varies, from people with no first aid training to medical doctors. Participants are asked to assign themselves to one of six possible groups regarding their training level in pre-hospital care, as presented in Table 1.

Note that none of the participants classifies themselves as members of Group 4. They are also asked to provide additional information, including their prior experience with sensor data (binary).

Table 1: The different levels of pre-hospital training.
(*n refers to the Number of participants in each group)

Group	Level of training	n
Group 1	None	10
Group 2	First Aid at Work, Emergency First Aid	42
Group 3	Basic First Person on Scene, Intermediate First Person on Scene, Equivalent to BASP Advanced First Aid, Combat Medical Technician 2	5
Group 4	Emergency Medical Technician, Ambulance Technician, Combat Medical Technician 1, Offshore Medic	0
Group 5	Paramedic, Nurse, Physician's Assistant	4
Group 6	Medical Doctor	8

Each participant is shown 4 different emergency scenarios in the same order. Each scenario consists of a textual description of the incident (e.g. see top of Figure 1) and three graphs that correspond to the

physiological measurements of Breathing Rate, Blood Oxygen Saturation (SpO2) and Heart Rate. In the processed time-series data, each time-stamp corresponds to one minute of measurements. For each graph, there are 6 ways of referring to the measured parameter as described previously. Each participant is asked to choose the phrase that s/he would use to describe each condition. A description of each scenario is shown in Table 2.

Table 2: A brief description of each scenario.

Scenario	Brief description	Data trends
Scenario 1: **Smoke inhalation**	The patient inhaled smoke during through a building fire	BR: increasing SpO2: decreasing HR: increasing
Scenario 2: **Drowning**	The patient swallowed water after falling into water	BR: stable SpO2 : stable HR: increasing
Scenario 3: **Falling down stairs**	The patient fell off the stairs	BR: stable SpO2: stable HR: decreasing
Scenario 4: **Bicycle accident**	A cyclist was hit by a car.	BR: increasing SpO2: stable HR: increasing

4. DATA ANALYSIS

The data analysis reveals a number of interesting results regarding the user preferences. Generally the phrase choices are influenced not only by the training level / occupation but also by the incident scenario they describe.

In the following sections, we analyze observed differences in the data with a Pearson's Chi-squared test.

4.1 Scenario / Phrase choice relation & Time-series data / Phrase choice relation

The data reveals that the phrase choice is correlated with two factors: the scenario and the physiological condition (i.e. Breathing Rate, Blood Oxygen Saturation and Heart Rate). For a detailed description of the frequencies please see Table 3.

For example, for the *smoke inhalation scenario*, we find that 66% of the participants choose to describe the Breathing Rate by mentioning the trend (increase) in a verbose way (template 2), whereas only 22% of the participants would describe the Breathing Rate using the succinct phrase of mentioning the trend (template 3).

A similar distribution can be observed for the phrases chosen for the Blood Oxygen Saturation variable: 63% of the participants chose template 2 and 28% template 3. For describing Heart Rate data, 46% of the participants choose the verbose way of

describing the trend (template 2), whereas 36% would choose the succinct way (template 3). In sum, participants mainly varied between template 2 and 3 and this choice was significantly different for different types of physiological data.

Table 3: Phrase choice frequencies (%) of each scenario.

Scenario	Template	Breathing Rate	SpO$_2$	Heart Rate
Scenario 1	Average	1.3	0	1.3
	Trend verbose	66.2	63.6	46.8
	Trend succinct	22.1	28.6	36.4
	Range verbose	1.3	3.9	11.9
	Range succinct	6.5	0	0
	Inference	2.6	3.9	3.9
Scenario 2	Average	32.3	35.3	0
	Trend verbose	1.5	0	64.7
	Trend succinct	1.5	4.5	26.4
	Range verbose	30.9	44.0	0
	Range succinct	17.6	14.8	1.5
	Inference	16.2	1.4	7.4
Scenario 3	Average	19.0	25.4	0
	Trend verbose	23.8	1.6	60.3
	Trend succinct	15.8	4.8	33.3
	Range verbose	17.4	47.6	3.2
	Range succinct	9.5	19.0	1.5
	Inference	14.0	1.6	1.5
Scenario 4	Average	6.4	30.6	0
	Trend verbose	51.6	1.6	61.3
	Trend succinct	25.8	9.7	32.3
	Range verbose	13.0	41.9	0
	Range succinct	3.2	16.2	1.6
	Inference	0	0	4.8

For the *drowning scenario*, 32% of the users prefer to refer to Breathing Rate by mentioning the average (template 1) and 30% prefer to mention the range in a verbose way (template 2). This is quite different from the observations derived from the smoke inhalation scenario. The results differ for the Blood Oxygen Saturation as well. 35.2% prefer to refer to it by mentioning the average (template 1), whereas 44% prefer to mention the range in a verbose way (template 4). Finally, 93.4% of the users prefer to mention the trend when referring to the Heart Rate (61.2% by mentioning it in a verbose way (template 2) and 32.2% by mentioning it in a succinct way (template 3).

For the *fall down-stairs scenario*, the users' preferences on mentioning the Breathing Rate are spread around the 6 templates. On the other hand, users prefer to mention the range of the Blood Oxygen Saturation in a verbose way (template 4, 47%) or mention the average (template 1, 25.3%). Regarding the Heart Rate, 93.6% of the participants preferred to mention the trend in a verbose (template 2, 60.3) and in a succinct way (template 3, 33.3%).

Finally, for the *bicycle accident scenario*, 51% of the users prefer to talk about the Breathing Rate by mentioning the trend in a verbose way (template 2),

whereas 25% of the users prefer the succinct way of referring to the trend (template 3). For the Blood Oxygen Saturation, the results are similar to the drowning scenario. 30.6% of the users prefer the average template (template 1) and 41.9% of the users prefer to refer to range in a verbose way (template 4). Finally, the preferences for the Heart Rate remain similar, with 61.2% of the users preferring verbose way of referring to trend (template 2) and 32.2% preferring the succinct way (template 3).

From the data available, it is observed that users select summarization templates depending not only on the physiological data but their choices are also informed by the scenarios. For instance, referring to the increasing trend of Breathing Rate in a verbose way in a smoke inhalation event seems highly important, as this factor might be highly correlated with the patient's overall condition.

4.2 Training level / Phrase choice relation

As previously mentioned, the participants were aggregated into 6 groups reflecting their pre-hospital training level. We now examine the data for existing correlations between the phrase choice and the level of training. The data reveals that the participants that belonged to the first three categories have similar preferences in terms of template choice. In contrast, participants belonging to group 5 and 6 have distinctly different preferences. Note that none of our participants belongs to Group 4, as mentioned in Section 3. We can therefore regroup the participants into 3 groups in terms of training level, summarizing groups 1-3 into one consistent preference group, we call the "novice" group, whereas we treat groups 5 and 6 as distinct groups with different levels of expertise.

Regarding the Breathing Rate, it was observed that the novice group (Groups 1-3), mostly preferred the verbose descriptions (template 2 and 4, 53%), followed by succinct (template 3 and 5, 20%) and finally the average (template 1, 16%), whereas Group 5 mostly preferred the succinct way (template 3 and 5, 77%), then the inference (template 6, 15%) and finally the verbose way (template 2 and 4, 57%). Medical doctors (Group 6) preferred both the verbose (56%) and succinct (40%) way. The average way was preferred by only 6% of the doctors. Similar observations were made for the Blood Oxygen Saturation parameter. Finally, for the Heart Rate parameter, it was observed that all user groups preferred the phrases that described the trend either in a verbose or a succinct way (90% of Groups 1-3, 92% of Group 5 and 86% of Group 6). If we split these percentages, again the novice group preferred the verbose way over the succinct (60% and 30% respectively). Group 5 preferred the succinct way over the verbose (69% and 23%

respectively) and finally 53% of the doctors preferred the verbose way and 33% the succinct.

These results suggest that the group preferences may vary, but there are common elements within their preferences. For instance, we found a general preference on reporting the trend of the physiological data across all groups (template 2 and 3), but a group-specific preference for either the verbose over the succinct way. Doctors' preferences (Group 6) are not quite clear as to whether they prefer the verbose over the succinct way. We think that this might be due to the fact that doctors usually communicate their findings to groups with different expertise, e.g. explaining results to patients or discuss a condition with other doctors, so they probably customize their descriptions to the interlocutor.

4.3 Gender / Phrase choice relation

There were no observed significant correlations between template choices and gender. When we examine the Breathing Rate parameter, 54% of male and 47% of female participants choose to refer to it in a verbose way. On the other hand, 33% of women choose the succinct way, but only 17% of men choose the succinct way (template 2 and 4). Whereas, for the Heart Rate parameter, we observe a different pattern: 86% of male participants choose the verbose way, but only half of the female participants (56%) choose the verbose way; 40% of women choose the succinct way, whereas 26% of men would choose the succinct way.

We conclude that for generating data descriptions, we do not need to take user gender into account, as gender preferences vary for different parameters (confirming our findings in Section 4.1).

4.4 Previous experience with sensor data / Phrase choice relation

We finally investigate the influence that previous experience with sensor data has on template choice preference. We find that prior experience with sensor data yield mixed results.

For the Breathing Rate parameter, there is no significant difference between the preferences of people with prior experience and people without prior experience. For the Blood Oxygen Saturation parameter, we observe that, on the one hand, the majority (58%) of the users without prior experience in sensor data prefer the verbose way of referring to the data trend, whereas only 30% of the users with prior experience prefer the verbose way.

On the other hand, 42% of users with prior experience prefer the succinct way of referring to the data. Notably, 22% of the same group prefers referring to the average. And an almost equal percentage (21%) of the group without previous

experience prefers the average reference. Finally, regarding the Heart Rate Parameter, users with prior experience almost equally prefer the verbose (both by referring to the trend and describing the value range) and the succinct way.

5. CLUSTERING REGARDING USER PREFERENCES

In sum, the above results show that, individual user characteristics, such as medical training level, gender, or experience with medical sensor data, do not have a significant effect on the template choice. The only significant factors are scenario and physiological parameters (Breathing Rate, Blood Oxygen Saturation and Heart Rate).

We conclude that, categorizing users depending on their pre-hospital training level, or their gender, or prior experience with sensor data, does not necessarily yield distinctive user groups. For instance, users that have received training at work can have similar preferences to medical doctors. We therefore consider automatic clustering to define user groups in terms of phrase choice, regardless of their training background, gender or experience with sensors, following a similar approach to the approach presented by [1]. However, Dethlefs et al. approach [1] addresses known users, in a sense that the user preferences are defined via previous ratings on generated text. In comparison, here, we deal with unknown users and therefore, placing a user into a group is not possible.

Cluster analysis allows to group a set of objects in such a way that objects in the same group (called a cluster) are more similar (here in terms of their phrase choice) to each other than to those in other groups/ clusters. For instance, people that prefer referring to the average value of time-series are more similar and thus they belong to the sane cluster, whereas people that prefer to refer to the trend in a verbose way belong to a different cluster etc. In this way, users are grouped according to their preferences and regardless of their profession, gender, or level of training.

We apply the Expectation-Maximization (EM) clustering algorithm provided by the WEKA toolkit [5]. EM is useful when the number of the clusters is not obvious (unknown), as in our dataset. Taking into account the analysis described in the previous section, it is obvious that we cannot cluster the users in terms of their occupation, gender or expertise. Therefore, we need a clustering algorithm that is able to determine the number of clusters automatically. EM initially assigns a probability distribution to each instance which indicates the probability of it belonging to each of the clusters. It uses cross validation to determine the number of clusters following these steps: (1) sets the number of clusters to 1; (2) splits the training set into 10 folds randomly; (3) EM algorithm is applied 10 times as normally in cross validation; (4) it averages the log likelihood over all 10 results; and (5) if log likelihood has increased, the number of clusters is also increased by 1 and it repeats until convergence is achieved.

Our clustering task was formed as follows: given the choices that a participant makes over all scenario, assign the user into a group. Accordingly, the features used for clustering are all the template choices a user makes for all 4 scenarios. EM clusters the data in two consistent user groups, where the first cluster consists of 27 participants and the second consists of 43 participants. Figure 2 shows the distribution of members of a group (as per level of training, see Table 1) for each cluster. We again used Chi-squared test to evaluate the consistency of the clusters in terms of the scenarios and the template choice. We also notice that the scenarios and the time-series data are multi-collinear, i.e. the scenarios and the time-series data are highly correlated. Therefore, the analysis using the scenarios produces the exactly same results as if we use the trends of time-series data instead. In the following two sections, we discuss each cluster in detail.
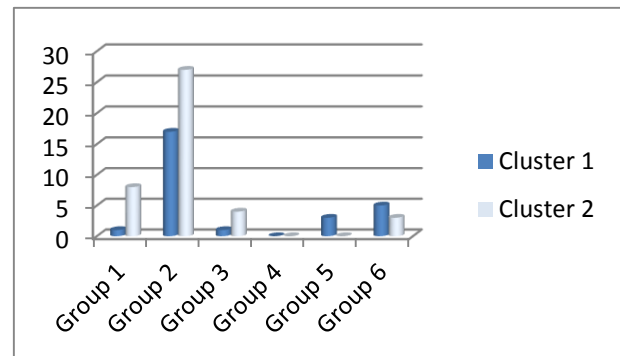

Figure 2: *Number of users per group belonging to each cluster, where group membership reflects level of expertise.*

5.1 Analysis of Cluster 1

The first cluster consists of 10 male and 17 female participants (Figure 3). 1 participant belongs to Group 1, 17 to Group 2, 1 to Group 3, 0 to Group 4, 3 to Group 5 and 5 to Group 6. 9 participants have previous experience with sensor data and 18 do not. Table 4 shows the phrase choice frequencies for each scenario for cluster 1. We easily observe that participants in Cluster 1 prefer the succinct ways of referring to time-series data generally.

Table 4: Phrase choice frequencies (%) of each scenario for cluster 1.

| Scenario	Template	Breathing Rate	SpO$_2$	Heart Rate
Scenario 1	Average	0	0	0
	Trend verbose	33.2	25.9	3.75
	Trend succinct	**51.8**	**66.6**	**92.5**
	Range verbose	0	3.75	3.75
	Range succinct	15	0	0
	Inference	0	3.75	0
Scenario 2	Average	24.0	20.0	0
	Trend verbose	0	4.0	16.0
	Trend succinct	8.0	12.0	**72.0**
	Range verbose	16.0	24.0	0
	Range succinct	**40.0**	**36.0**	4.0
	Inference	12.0	4.0	8.0
Scenario 3	Average	16.6	12.5	0
	Trend verbose	4.2	0	8.3
	Trend succinct	**37.5**	12.5	**83.3**
	Range verbose	8.35	25.0	0
	Range succinct	25.0	**45.8**	4.2
	Inference	8.35	4.2	4.2
Scenario 4	Average	4.2	16.6	0
	Trend verbose	12.5	0	4.2
	Trend succinct	**66.6**	25.0	**83.3**
	Range verbose	8.35	20.8	0
	Range succinct	8.35	**37.6**	4.2
	Inference	0	0	8.3
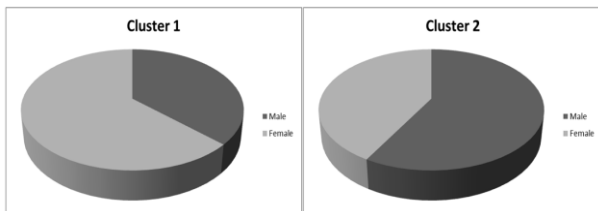
Figure 3: *Males/Females in Cluster 1 and Cluster 2.*
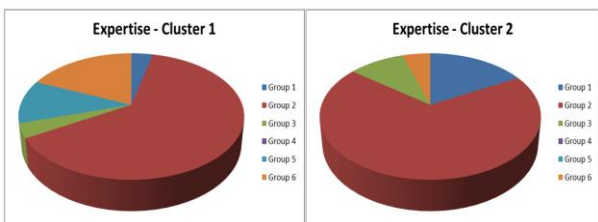

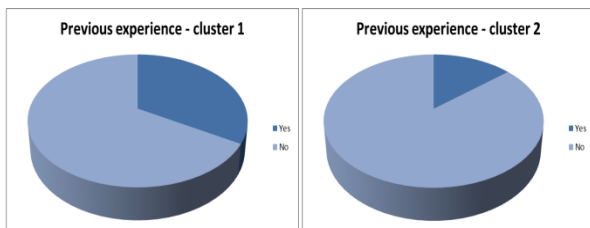**Figure 4:** *Different level of expertise in the two clusters.*

Figure 5: *Previous experience with sensor data in the two clusters.*

5.1 Analysis of Cluster 2

The first cluster consists of 25 male and 18 female participants. 8 participants belong to Group 1, 27 to Group 2, 4 to Group 3, 0 to Group 4, 0 to Group 5 and 3 to Group 6. 6 participants have previous experience with sensor data and 37 do not. Table 5 shows the phrase choice frequencies for each scenario for cluster 2. In contrast with cluster 1, we observe that the users in this cluster prefer the verbose ways of referring to time-series data.

Table 5: Phrase choice frequencies (%) of each scenario for cluster 2.

Scenario	Template	Breathing Rate	SpO$_2$	Heart Rate
Scenario 1	Average	2.3	0	2.4
	Trend verbose	**93.1**	**93.1**	**79.0**
	Trend succinct	0	0	0
	Range verbose	2.3	4.6	16.2
	Range succinct	0	0	0
	Inference	2.3	2.3	2.4
Scenario 2	Average	34.1	39	0
	Trend verbose	4.8	0	**97.5**
	Trend succinct	0	0	0
	Range verbose	**41.5**	**58.5**	0
	Range succinct	2.5	2.5	0
	Inference	17.1	0	2.5
Scenario 3	Average	23.0	38.4	0
	Trend verbose	**35.9**	0	**87.2**
	Trend succinct	0	0	2.6
	Range verbose	20.6	**59**	7.6
	Range succinct	2.6	2.6	0
	Inference	17.9	0	2.6
Scenario 4	Average	10.6	39.4	0
	Trend verbose	**73.7**	2.7	**97.3**
	Trend succinct	0	0	0
	Range verbose	15.7	**55.2**	0
	Range succinct	0	2.7	0
	Inference	0	0	2.7

In particular, as seen in Table 5, users in Cluster 2 prefer mentioning the trend for all variables in a verbose way in the smoke inhalation scenario. In the drowning scenario, the users prefer the sentences that describe the range of values for the Breathing Rate and Blood Oxygen Saturation, where these two variables remain stable throughout the monitoring. Users in Cluster 1 also preferred to mention the range of values, but in a succinct way. However, the Heart Rate variable increases and the users prefer to refer to the trend. Similar preferences are observed for the other two scenarios.

In conclusion, it is obvious that the two clusters are very similar in that for the same scenarios, all users seem to agree on the way that they would refer to time-series data (e.g. referring to trend over the range).

6. SUMMARY AND FUTURE WORK

In this paper, we presented results from a user study on individual preferences regarding textual descriptions of sensor data. We find that individual user characteristics, such as medical training level, gender, or experience with medical sensor data (although we did not ask about their experience with graphs), do not have a significant effect on the summarization template choice.

The only significant factors regarding this choice are the scenarios and the physiological parameters (Breathing Rate, Blood Oxygen saturation and Heart Rate). We therefore used cluster analysis to discover consistent user groups in terms of their template preferences with regard to scenarios and physiological parameters. Using EM, we discovered two consistent groups, which include participants across all different levels of training.

In future, we will use these findings to design a Natural Language Generation algorithm for summarizing time-series data from physiological sensors, which can equally adapt to both groups, following [3].

These automatically generated summaries can, for example, be used as part of first aid apps, such as the Red Cross app[3] or St John's app[4] , in order to address the problem of dealing with unknown users.

7. ACKNOWLEDGEMENTS

8. REFERENCES

[1] Dethlefs, N., Cuayáhuitl, H., Hastie, H., Rieser V. and Lemon, O. 2014. Cluster-based Prediction of User Ratings for Stylistic Surface Realisation. 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL).

[2] Gatt, A., Portet, F., Reiter, E., Hunter, J., Mahamood, S., Moncur, W. and Sripada, S. 2009. From data to text in the Neonatal Intensive Care Unit: Using NLG technology for decision support and information management. AI Communications, 22(3): 153 - 186.

[3] Gkatzia, D, Hastie, H and Lemon, O. 2014. Finding middle ground? Multi-objective Natural Language Generation from time-series data. Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL).

[4] Gkatzia, D, Hastie, H and Lemon, O. 2014. Multi-adaptive Natural Language Generation from time-series data. Proceedings of the 8th International Natural Language Generation Conference (INLG).

[5] Hall, M, Eibe, F., Holmes, G., Pfahringer, B., Reutemann, P., Witten. I. H. 2009. The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.

[6] Hunter, J., Freer, Y., Gatt, A., Sripada, Y., Sykes, C. and Westwater, D. 2011. BT-Nurse: Computer Generation of Natural Language Shift Summaries from Heterogeneous Medical Data. American Medical Informatics Association, 18(5): 621 - 624.

[7] Mahamood, S. and Reiter, E. 2011. Generating Affective Natural Language for Parents of Neonatal Infants. Proceedings of the 13th European Workshop on Natural Language Generation (ENLG).

[8] van den Meulen, M., Logie, R.H., Freer, Y., Sykes, C., McIntosh, N. and Hunter, J. 2010. When a graph is poorer than 100 words: A comparison of computerised natural language generation, human generated descriptions and graphical displays in neonatal intensive care. Applied Cognitive Psychology, 24(1): 77 - 89.

[9] Petre, M. 1995. Why looking isn't always seeing: Readership skills and Graphical Programming. Transactions of the ACM. 38(6): 33-44.

[10] Reiter, E. 2007. An architecture for data-to-text systems. Proceedings of the 11th European Workshop on Natural Language Generation (ENLG).

[11] Schneider, A., Vaudry, P., Mort, A., Mellish, C., Reiter, E., and Wilson, P. 2013. MIME - NLG in Pre-hospital Care. In Proceedings of 14th European Natural Language Generation Workshop (ENLG).

---

[3] http://www.redcross.org.uk/What-we-do/First-aid/Mobile-app

[4] http://www.sja.org.uk/sja/support-us/the-difference/helpless/mobile-phone-app.aspx