International Journal of

New Computer Architectures and their Applications

SDIWC

www.sdiwc.net

Overview

The SDIWC International Journal of New Computer Architectures and Their Applications (IJNCAA) is a refereed online journal designed to address the following topics: new computer architectures, digital resources, and mobile devices, including cell phones. In our opinion, cell phones in their current state are really computers, and the gap between these devices and the capabilities of the computers will soon disappear. Original unpublished manuscripts are solicited in the areas such as computer architectures, parallel and distributed systems, microprocessors and microsystems, storage management, communications management, reliability, and VLSI.

One of the most important aims of this journal is to increase the usage and impact of knowledge as well as increasing the visibility and ease of use of scientific materials, IJNCAA does NOT CHARGE authors for any publication fee for online publishing of their materials in the journal and does NOT CHARGE readers or their institutions for accessing the published materials.

IJNCAA

Volume 8, Issue No. 2

2018

CONTENTS

ORIGINAL ARTICLES

International Journal of
NEW COMPUTER ARCHITECTURES AND THEIR APPLICATIONS

The *International Journal of New Computer Architectures and Their Applications* aims to provide a forum for scientists, engineers, and practitioners to present their latest research results, ideas, developments and applications in the field of computer architectures, information technology, and mobile technologies. The IJNCAA is published four times a year and accepts three types of papers as follows:

1. **Research papers**: that are presenting and discussing the latest, and the most profound research results in the scope of IJNCAA. Papers should describe new contributions in the scope of IJNCAA and support claims of novelty with citations to the relevant literature.
2. **Technical papers**: that are establishing meaningful forum between practitioners and researchers with useful solutions in various fields of digital security and forensics. It includes all kinds of practical applications, which covers principles, projects, missions, techniques, tools, methods, processes etc.
3. **Review papers**: that are critically analyzing past and current research trends in the field.

Manuscripts submitted to IJNCAA **should not be previously published or be under review** by any other publication. Plagiarism is a serious academic offense and will not be tolerated in any sort! Any case of plagiarism would lead to life-time abundance of all authors for publishing in any of our journals or conferences.

Original unpublished manuscripts are solicited in the following areas including but not limited to:

- Computer Architectures
- Parallel and Distributed Systems
- Storage Management
- Microprocessors and Microsystems
- Communications Management
- Reliability
- VLSI

A Study for Dynamically Adjustmentation for Exploitation Rate using Evaluation of Task Achievement

Masashi SUGIMOTO

National Institute of Technology, Kagawa College

Email: sugimoto-m@es.kagawa-nct.ac.jp

ABSTRACT

Until now, in reinforcement learning, a ratio of a random action as known as exploration often has not been adjusted dynamically. However, this ratio will be an index of performance in the reinforcement learning. In this study, agents learn using information from the evaluation of achievement for task of another agent, will be suggested. From this proposed method, the exploration ratio will be adjusted from other agents' behavior, dynamically. In Human Life, an "atmosphere" will be existed as a communication method. For example, empirically, people will be influenced by "serious atmosphere," such as in the situation of working, or take an examination. In this study, this atmosphere as motivation for task achievement of agent will be defined. Moreover, in this study, agent's action decision when another agent will be solved the task, will be focused on. In other words, an agent will be trying to find an optimal solution if other agents have been found an optimal solution. In this paper, we propose the action decision based on other agent's behavior. Moreover, in this study, we discuss effectiveness using the maze problem as an example. In particular, "number of task achievement" and "influence for task achievement," and how to achieve the task quantitative will be focused. As a result, we confirmed that the proposed method is well influenced from other agent's behavior.

KEYWORDS

Reinforcement Learning, Exploration ratio, Action Selection Strategy, Multi Agent, Behavior using Communication, Cooperative Work, Interworking Algorithm, Agricultural Weeding Robot

1 INTRODUCTION

Over the years, many studies have been conducted with the objective of facilitating the working of robots in dynamic environments [1, 2, 3]. Various robots have been developed to assist humans in workspaces, such as a house or factory [4]. In general, robots are required to work effectively and safely in a dynamic environment to achieve their tasks. However, it is not easy to make a robot behave like a human in dynamic environments [5, 6]. When they are working in a certain environment, humans select an appropriate course of action through subconsciously predicting all the changes in the environment and their next state. For achievement these problems, in recent years, various machine learning methods have been suggested. In reinforcement learning, it attracts attention as the technique that often use in the actual robot [7, 8, 9, 10, 11]. However, reinforcement learning has some problems. In one of the problems, a robot does not cope with changing purpose in reinforcement learning. Reinforcement learning has been demanded to achieve various purposes, because what request to robot is diversifying and to achieve various purposes in robot have been wanting, as mentioned above. Therefore, it is important to solve this problem.

Until now, in reinforcement learning, a ratio of a random action as known as exploration [11] often hasn't been adjusted dynamically [12, 13]. However, this ratio will be an index of performance in the reinforcement learning. In this study, agents learn using information from the evaluation of achievement for task of another agent, will be suggested. From this proposed method, the exploration ratio will be adjusted from other agents' behavior, dynamically. In Human Life, an "atmosphere" will be existed as a communication method. For example, empirically, people will be influenced by "serious atmosphere," such as

in the situation of working, or take an examination. In this study, this atmosphere as motivation for task achievement of agent will be defined. Moreover, in this study, agent's action decision when another agent will solve the task, will be focused on. In other words, an agent will be trying to find an optimal solution if other agents have been found an optimal solution.

In this paper, we propose the action decision based on other agent's behavior. Moreover, in this study, we discuss effectiveness using a maze problem as an example. In particular, "a number of task achievable" and "influence for task achievement," and how to achieve the task quantitative will be focused. As a result, we confirmed that the proposed method is well influenced from other agent's behavior.

This paper is organized as follows: In section 2, we explain the how to exploration ratio will be adjusted from other agents' behavior, dynamically. In parallel, we provide details about the proposed method. In Section 3, we explain about the setting for the experiment. Finally, in Section 4, we present the conclusions of this study.

2 A CONCEPT OF ACTION-DECISION BASED ON OTHER AGENT'S BEHAVIOUR

2.1 Basic Idea

In Human life, it seems that there is a kind of information transmission methods called "atmosphere." For example, in behavior such as applause and attitude, interaction and cooperation with the surroundings is performed unconsciously. These are accepted as a kind of "atmosphere." Also, at school examination and some work, "serious atmosphere" propagates, moreover, it is empirically occurring that people are influenced, as unconscious and gradually inflation in the same space (perhaps the readers had been might be confirmed empirically). However, since this concept of "atmosphere" is too abstract. Therefore, in this study will be interpreted it in the form of motivation to accomplish the task. Now, let consider in

case of work in human society. When a work is given to someone and work is also given to another person. Think about the pattern that you get. At this time, each one will silently work under "tacit understanding." Of course, at this time, one will quietly and silently attain the work, in the majority of cases, it can be said that it is meaningful to be able to acquire methods to solve work in the shortest time, that is, to finish the work as soon as possible. As mentioned above, behavior selection is done under motivation.

Now, in this paper, the above-mentioned
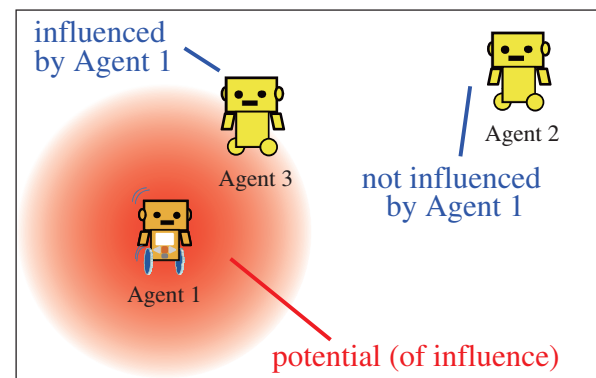
Figure 1. Aim of this study (situation).

"motivation" is greatly thought to propagate like the real environment. This is briefly illustrated in fig. 1. At this time, agent-1 will be contained influence area, and decreasing in a manner of attenuating farther away from himself. When reconsidering mentioned above scenario, if we think that it is working with a maze task, if this influence is within the range, agent-3, which is close to agent-1 in this figure, will be influenced from agent-1 at least. It can be considered harder to receive him influence because agent-2 must be away.

At this time, it can be thought that agent-1 can have some influence in such a way that it decreases as it goes away from itself. If we reconsider previous consideration if operating with a maze task, the number of goals will directly influence the achievement of work, the more seriously it will be accepted as "serious." Moreover, if there is scope for this effect, agent-3 could be thought to be more susceptible than agent-2 and it can be thought that

agent-2 won't be affected by the fact that it is away from it. Hence, the impact of this agent by setting the number of achievements of the task can be expressed as a mathematically model, to show this hypothesis; the task given to N_g times:

$$\exp\left|-\frac{p^2}{\frac{\sqrt{2}}{3}rN_g}\right|. \qquad (1)$$

In this paper, we express this as the potential possessed by the agent, where p is the position of the agent, r is the range of the potential effect of the agent, and N_g is the number of task achievements. Now, in this case, the exponent part may be indefinite in some cases, especially when $N_g = 0$, is naturally to regard the influence as 1 regardless of the state of the agent.

2.2 How to Influence by Other Agent ?
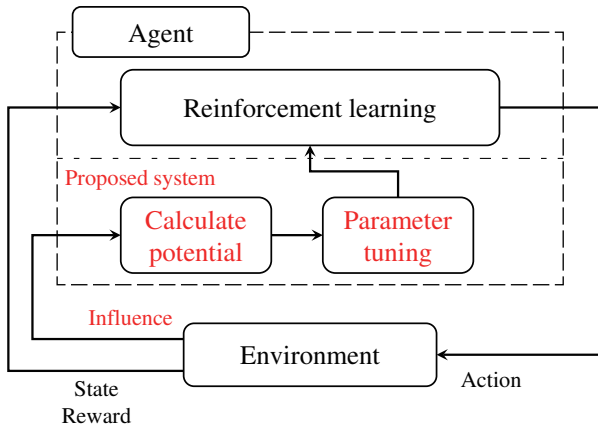
Figure 2. An Outline of the Proposed System.

The figure 2 is the outline of the proposed system. In this method contains two parts; calculate potential part (the part calculating the influence rate of agent) and the parameter tuning part (part calculating the stochastic element from the influence rate). At this time, the agent operating in the proposed system also has the influence of other agents in addition to the state and reward of the environment.

$$\varphi_i = \exp\left|-\frac{(x-p_i)^2}{\frac{\sqrt{2}}{3}r_i N_{g,i}}\right|. \qquad (2)$$

On calculate potential part, the influence degree φ_i received from A_i based on the influence

received from the potential of another agent A_i will be calculated. This influence degree φ_i is calculated by the following expression:

$$\varphi = \sum_{i \in R} \exp\left|-\frac{(x-p_i)^2}{\frac{\sqrt{2}}{3}r_i N_{g,i}}\right|. \qquad (3)$$

Here, x is the position of its own agent, p_i is the position of A_i, r_i is the influence range, and $N_{g,i}$ is the number of achievements of task by A_i. We use the number of achievements as an index. The above expression is based on the expression (2), so the exponent part may be indefinite, so $N_{g,i} = 0, \varphi_i$ is assumed to be 1 irrespective of the state of x, p_i. If R agents affect each other in the same space, as shown in fig. 1.

$$\epsilon = \varphi_i. \qquad (4)$$

Next, we describe parameter tuning part, here we decide the probabilistic element of the influence degree φ_i calculated by calculating potential part. In this study, in particular, the ϵ-greedy strategy is used for the behavior selection method. Considering that decision this ϵ dynamically from φ_i.

That is, when another agent discovers an optimal solution, its own agent also selects a behavior of searching for an optimum solution up to the goal according to it.

If R of agents are affecting each other in the same space, they can be summed from the expression (3) as follows:

$$\epsilon = \frac{\varphi_i}{R}. \qquad (5)$$

Consider using this φ_i as an *indicator of action selected*, that is, apply it to the random action ϵ in the ϵ-greedy strategy, so the agent A_i in the surrounding r_i. The more seriously the goal is, the more individuals are affected by φ_i, which means that the best action is selected at the local point.

3 VERIFICATION EXPERIMENT – COMPUTATIONAL SIMULATION USING THE PROPOSED METHOD (1)

3.1 Outline of the Experiment

We verify the effectiveness of the proposed method up to the previous section by computer simulation. The effectiveness is evaluated by comparing the difference of the convergence speed of the learning of the proposed reinforcement learning with the proposed method. At this time, the ordinary reinforcement learning method is to learn the route that reaches the goal while avoiding walls and pitfalls through trial and error, and the reinforcement learning to which the proposed method is applied. Behavior will be selected according to behavior facing. Also consider the maze environment

Figure 3. An Experimental Environment (Grid Maze) (1).

with walls and pitfalls consisting of a grid of 4×6 shown in fig. 3 as the experimental environment. Moreover, the agent implemented a proposed method will be affected by 2-types agents during task execution.

In figure 3, the water blue-colored mass is the wall and the orange-colored mass is the catchpoint. The two agents are perfect perception and can move up, down, left and right of the grid, among which the pitfall (H). In case of

Figure 4. A Structure of Verification Experiment.

falling to the start point (S) and agent getting a reward -10, or get a reward +1 when agent reaches the goal point (G).

Table 1. Experimental parameters for Agents (1).

Agent-1	Agent-2	Agent-3	Property
0	0	0	Initial value of Q values
0.1	0.1	0.1	Learning rate α
0.95	0.95	0.95	Discount value γ
0.1	Eq.(2)	1.0	Exploration rate ϵ
1.0	0.0	2.0	Influences range area r

3.2 Condition of Simulation

In this experiment, we mainly deal with episodic tasks: Agent-1 is an agent that operates with ordinary reinforcement learning, Agent-2 is an agent that combines the proposed method and reinforcement learning. Moreover, Agent-3 is an agent that operates with a completely random action. Agent-1 and Agent-3 are Agent-2's learning without being affected, Agent-2 will select actions affected by learning and behavior other agents.

When each agent reaches the goal point (G) from the start point (S), the reward is obtained and the process returns to the start (S) Also, as described above, even when falling into the pitfall (H), it returns to the start point (S). Treat this as one episode In this experiment we will

do 400 episodes. Setting of experimental parameters is as shown in the following table 1.
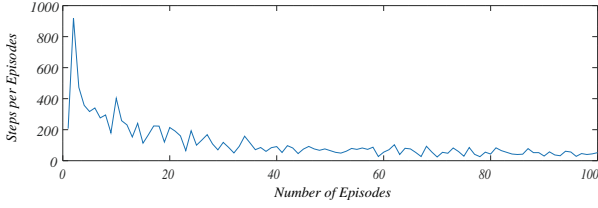
Figure 5. Number of Action per Episodes of Agent-1 (1).
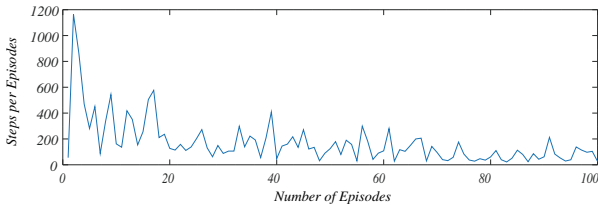
Figure 6. Number of Action per Episodes of Agent-2 (1).
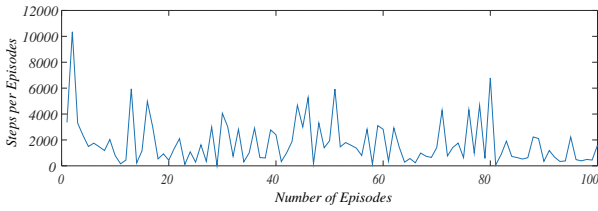
Figure 7. Number of Action per Episodes of Agent-3 (1).

3.3 Discussion on Simulated Results

Figures 5 through 8 shows the results of the experiment. Figures 5 and 7 are the transition of the behavior in each episode by Agent-1 and Agent-3. Figure 6 is the transition of the behavior in each episode of Agent-2 applying the proposed method. The initial value of learning is the number of the behaviors. From these results we can confirm that almost identical to Agent-2, however, as learning progresses, that can be seen that it follows Agent-2 that achieves episode with a fewer number of behaviors than Agent-1.

ϵ is 1 at the beginning of the action and chooses the exploratory behavior, however, as
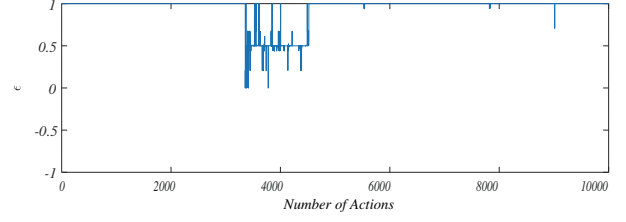
Figure 8. Transition of Exploration Ratio ϵ of Agent-2.

learning progresses, ϵ decreases to 0 as Agent-1's goal number increases, next episode. Moreover, in the next episode, ϵ becomes 1. On the other hands, ϵ had been rising when Agent-1 has faced on Agent-3 was confirmed. We will be considered that the rise resulted from an action strategy of Agent-3. From this result, it could experimentally confirm the fact that actioned will be become while watching the progress of the opponent. Similarly, at the same time. Therefore, the action was realized by searching for an action that finds the optimal solution of the given task (along with it) only when the agent to find the optimal solution of the task that will be confirmed.

4 VERIFICATION EXPERIMENT – COMPUTATIONAL SIMULATION USING THE PROPOSED METHOD (2)

4.1 Outline of the Experiment

In this section, the potential range will be tried to extend such as circle to ellipse. Therefore, we will extend the potential as below:

$$r_i = \frac{L}{1 + \varepsilon_r \cos \arccos \left(\frac{S_i - S_x}{S_i} \right)} \quad (6)$$

$$L = B^2/A \quad (7)$$

$$\varepsilon_r = C/A \quad (8)$$

$$C = \sqrt{A^2 - B^2} \quad (9)$$

On calculate potential part, the influence degree φ_i received from A_i based on the influence received from the potential of another agent A_i will be calculated. This influence degree φ_i is

calculated by the following expression:

$$\varphi = \sum_{i \in R} \exp \left| -\frac{(S_x - p_i)^2}{\frac{\sqrt{2}}{3} r_i N_{g,i}} \right|. \quad (10)$$

Here, S_x is the position of its own agent, p_i is the position of A_i, r_i is the influence range, and $N_{g,i}$ is the number of achievements of task by A_i. We use the number of achievements as an index. Moreover, A, B are the parameters of ellipse. The above expression is based on the expression $(2, 2)$, so the exponent part may be indefinite, so $N_{g,i} = 0, \varphi_i$ is assumed to be 1 irrespective of the state of x, p_i. If R agents affect each other in the same space, as shown in fig. 1.

$$\epsilon = \varphi_i. \quad (11)$$

Next, we describe parameter tuning part, here we decide the probabilistic element of the influence degree φ_i calculated by calculating potential part. In this study, in particular, the ϵ-greedy strategy is used for the behavior selection method. Considering that decision this ϵ dynamically from φ_i.

That is, when another agent discovers an optimal solution, its own agent also selects a behavior of searching for an optimum solution up to the goal according to it.

If R of agents are affecting each other in the same space, they can be summed from the expression (3) as follows:

$$\epsilon = \frac{\varphi_i}{R}. \quad (12)$$

Consider using this φ_i as an *indicator of action selected*, that is, apply it to the random action ϵ in the ϵ-greedy strategy, so the agent A_i in the surrounding r_i. The more seriously the goal is, the more individuals are affected by φ_i, which means that the best action is selected at the local point.

In this experiment, we verify the effectiveness of the proposed method in case of influence potential range will be changed such as a circle to an ellipse. The effectiveness is evaluated by comparing the difference of the convergence speed of the learning of the

proposed reinforcement learning with the proposed method. At this time, the ordinary reinforcement learning method is to learn the route that reaches the goal while avoiding walls and pitfalls through trial and error, and the reinforcement learning to which the proposed method is applied. Behavior will be selected according to behavior facing. Also consider
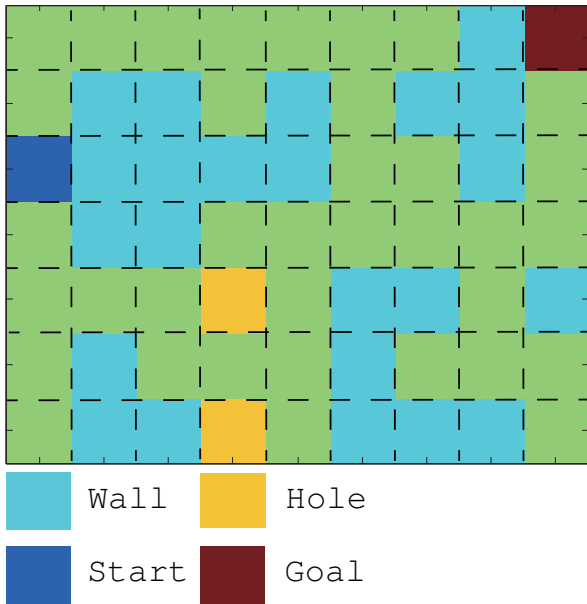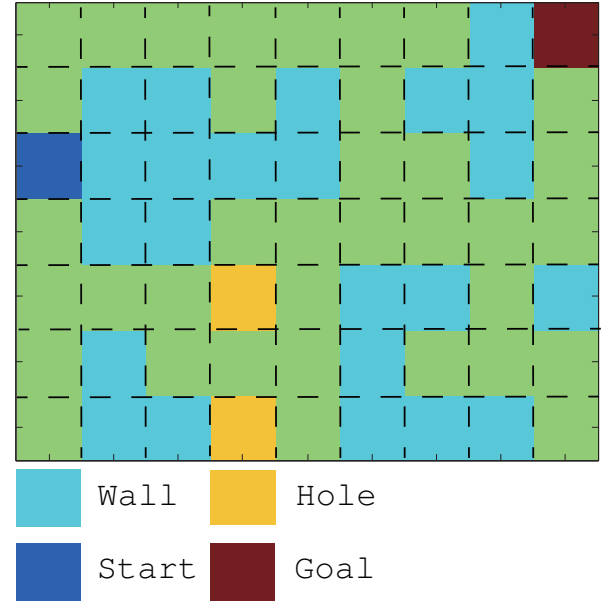

Figure 9. An Experimental Environment (Grid Maze) (2).

the maze environment with walls and pitfalls consisting of a grid of 4×6 shown in fig. 9 as the experimental environment. Moreover, the agent implemented a proposed method will be affected by 2-types agents during task execution.

In figure 9, the water blue-colored mass is the wall and the orange-colored mass is the catchpoint. The two agents are perfect perception and can move up, down, left an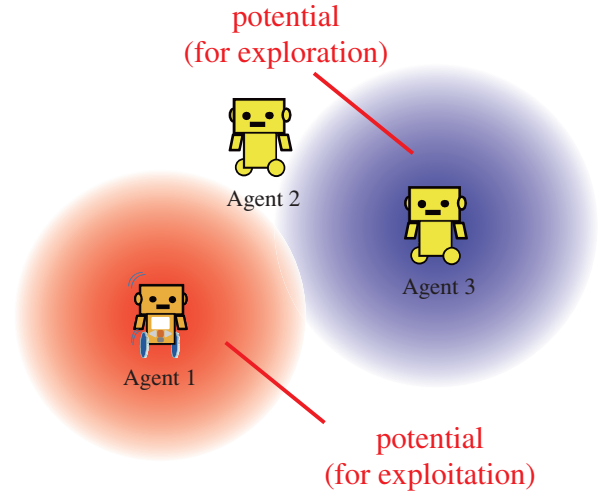d right of the grid, among which the pitfall (H). In case of falling to the start point (S) and agent getting a reward -10, or get a reward +1 when agent reaches the goal point (G).

4.2 Condition of Simulation

In this experiment, we mainly deal with episodic tasks: Agent-1 is an agent that operates with ordinary reinforcement learning, Agent-2 is an agent that combines the proposed

Table 2. Experimental parameters for Agents (2).

Agent-1	Agent-2	Agent-3	Property
0	0	0	Initial value of Q values
0.1	0.1	0.1	Learning rate α
0.95	0.95	0.95	Discount value γ
0.1	Eq.(2)	1.0	Exploration rate ϵ
1.0	0.0	Eq.(6)	Influences range area r

method and reinforcement learning. Moreover, Agent-3 is an agent that operates with a completely random action. Agent-1 and Agent-3 are Agent-2's learning without being affected, Agent-2 will select actions affected by learning and behavior other agents.

When each agent reaches the goal point (G) from the start point (S), the reward is obtained and the process returns to the start (S) Also, as described above, even when falling into the pitfall (H), it returns to the start point (S). Treat this as one episode In this experiment we will do 400 episodes. Setting of experimental parameters is as shown in the following table 2.
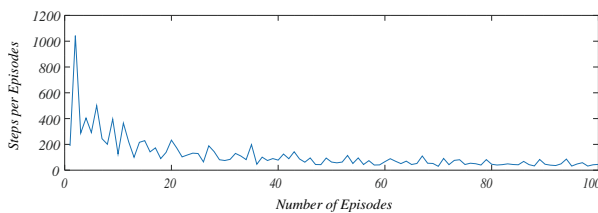
Figure 10. Number of Action per Episodes of Agent-1 (2).
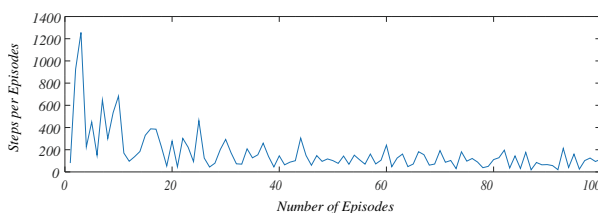
Figure 11. Number of Action per Episodes of Agent-2 (2).

4.3 Discussion on Simulated Results

Figures 5 and 7 are the transition of the behavior in each episode by Agent-1 and Agent-3. Figure 11 is the transition of the behavior
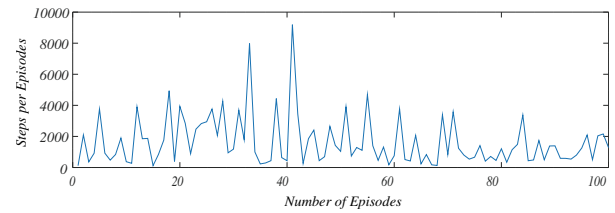
Figure 12. Number of Action per Episodes of Agent-3 (2).

in each episode of Agent-2 applying the proposed method. The initial value of learning is the number of the behaviors. From these results we can confirm that almost identical to Agent-2, however, as learning progresses, that can be seen that it follows Agent-2 that achieves episode with a fewer number of behaviors than Agent-1. However, the behavior of Agent-2 seems randomly than former experiment. This symptom is caused that Agent-2 had fallen within the range of Agent-3. Moreover, the range of Agent-3 is extended than former experiment. From the above, also, the action was realized and affected from other agents, will be confirmed.

5 CONCLUSION

In this paper, a method to dynamically adjust the action-decision strategy based on other agent's behavioral results, has been proposed. In this method, the evaluation of other agent's behavioral results is its number of agent's task achievement. Moreover, in the proposed method, a number of task achievement and a potential that the agent has unique effective range have been defined. Further, these parameters have been affected the exploration ratio as epsilon. From this method, the simulation results showed the proposed method has been acquired actions to reach the goal more efficiently than conventional method. In other words, the number of trials of proposed method's agent is less, while its agent will not be affected by another agent. Inversely, if another agent will be found the route to reach to goal with the shortest action number, the proposed method's agent will be decided the action that reach to the goal with the shortest action number, repeatedly, and affected by his be-

havior.

From these results, the proposed method has been confirmed to efficiently accomplish the task, while adjusting itself looking at the other agent's behavior. Therefore, we conclude that the usefulness of the proposed method has been confirmed.

Now, let's consider the living thing, again. Especially, in Human Life, not only "atmosphere," but also "synchronization" has been existed in social life, moreover, this communication method will be important [14]. Further, in other living things, it is often done to specify actions in the form of cautionary or guidance by pheromones [15]. Especially in humans, interaction and cooperation with the surroundings are doing unconsciously within this atmosphere and synchronization [10, 14]. If a robot working in a dynamic environment, such as daily life, will be implemented, these information methods, its will be perform actions and tasks with higher affinity with humans will be predicted.

REFERENCES

[1] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents series)*. The MIT Press, 2005.

[2] S. Asaka and S. Ishikawa. Behavior Control of an Autonomous Mobile Robot in Dynamically Changing Environment. Journal of the Robotics Society of Japan, 12(4):583-589, 1994.

[3] T. Kanda, H. Ishiguro, T. Ono, M. Imai, T. Maeda, and R. Nakatsu. Development of "Robovie" as Platform of Everyday-Robot Research. IEICE Transactions on Information and Systems, Pt.1 (Japanese Edition), J85-D-1(4):380-389, 2002.

[4] International Federation of Robotics. *All-time-high for industrial robots Substantial increase of industrial robot installations is continuing*, 2011.

[5] T. Sogo, K. Kimoto, H. Ishiguro, and T. Ishida. Mobile Robot Navigation by a Distributed Vision System. Journal of the Robotics Society of Japan, 17(7):1-7, 1999.

[6] J. J. Park, C. Johnson, and B. Kuipers. Robot Navigation with MPEPC in Dynamic and Uncertain Environments: From Theory to Practice. IROS 2012 Workshop on Progress, Challenges and Future Perspectives in Navigation and Manipulation Assistance for Robotic Wheelchairs, 2012.

[7] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. The MIT Press, 1998.

[8] N. Sugimoto, K. Samejima, K. Doya, and M. Kawato. Reinforcement Learning and Goal Estimation by Multiple Forward and Reward Models. IEICE Transactions on Information and Systems, Pt.2 (Japanese Edition), J87-D-2(2):683-694, 2004.

[9] Y. Takahashi and M. Asada. Incremental State Space Segmentation for Behavior Learning by Real Robot. Journal of the Robotics Society of Japan, 17(1):118-124, 1999.

[10] A. Agogino and K. Tumer. Reinforcement Learning in Large Multi-agent Systems. In Proc. of AAMAS-05 Workshop on Coordination of Large Scale Multiagent Systems, 2005.

[11] E. A. Guggenheim. *Boltzmann's Distribution Law*. North-Holland Publishing Company, 1955.

[12] N. Shibuya and K. Kurashige. Control of exploration and exploitation using information content. In Proc. of the Nineteenth International Symposium on Artificial Life and Robotics 2014, pp.48-51, 2014.

[13] T. Masaki and K. Kurashige. Decision Making Under Multi Task Based on Priority for Each Task. International Journal of Artificial Life Research, 6(2):88-97, 2016.

[14] D. Li and Y. Du. *Artificial Intelligence with Uncertainty*, Second Edition. CRC Press, 2017.

[15] C. Blum. Ant Colony Optimization: Introduction and recent trends. Elsevier Physics of Life Reviews, 2:353-373, 2005.

Control of a powerpoint presentation using eye movements

Roy ABI ZEID DAOU[*,**], Elie GEAGEA[*], Miguel SFEIR[*]

[*]Lebanese German University, Faculty of Public Health,
Biomedical Technologies Department
Jounieh, Lebanon
Email: r.abizeiddaou@lgu.edu.lb
[**] MART Learning, Education and Research Center,
Jounieh, Lebanon
Email: roydaou@mart-ler.org

Ali HAYEK, Josef BORCSOK

Kassel University, Department of Computer Architecture
and System Programming
Kassel, Germany
Ali.hayek@uni-kassel.de

Abstract—The purpose of this paper is to present a system capable of helping a person to control the slide shows during his/her presentation without the use of his/her hands. In fact, a particular movement in the eye will allow a change in the slide (towards the next or the previous slide). To do so, some electrodes were used in order to record the eyes movements and some electronic circuits as well as a microcontroller were deployed for recognizing the eye movement signal from the noise in order to limit any faulty signal. A wireless signal is sent to the computer in order to move the presentation up or down. The system was tested on several persons while taking into consideration the age, the skin color, the fatigue, the quantity of light, the use of glasses / medical lenses and the results were very promising especially with the hitting percentage that reached more than 90%, which means, in other words, that the error percentage was less than 10%.

Keywords—Embedded system; EEG system; Electrodes; Microcontroller; powerpoint presentation control; analog filters;

I. INTRODUCTION

The control of a power point presentation is sometimes difficult for people with some movement difficulties or when the computer is placed far from the presenter. Added to that, lots of persons usually use both hands when explaining their ideas which render this tough to handle, in their hand, a remote control during all the presentation.

Lots of researchers were interested in developing tools to help controlling a presentation automatically. From among these researchers, we list the work of *Otsuki* et al. who designed a new method to control presentation in hybridcast [1]. As for *Hsu* et al., they presented a new technic for a multimodal interactive control system in a museum [2]. Concerning *Suta* et al., they implemented a novel way for presentation using tablet PCs in classrooms or meeting rooms [3].

As for the applications deployed using eye movements, they are also numerous and they are mostly developed for paralyzed persons or for those with movement difficulties. In the following, we will list some of these applications: automatic control of a wheelchair by some eye movements [4],

control of computer cursor [5], development of a communication supportive device for patient who have motor paralysis – and therefore can not talk [6], implementation of real time text speller [7] and much more,…

Hence, the aim of this work is to include these two features in order to create a wireless system to control the PowerPoint presentation by some eye movements. In fact, some eye movements may be sufficient to control the slide show in a presentation. Two main signal must be recognized: one to move the slide forward whereas the second is to move it backward.

As for the main constraints of this system, they can be resumed by the following: system must be portable, small in size, light weighted, harmless, easily implemented and used, have affordable price and could be extended to control more advanced systems.

So, based on the above, this paper will be organized as follow: in section 2, some important information about the eye physiology, the electroencephalogram (EEG) and the electrodes will be provided. Section 3 will present the hardware tools needed for the implementation of such system. Section 4 presents the microcontroller as well as the implemented code. In section 5, the whole system will be proposed along with the obtained results. This paper will be summarized in section 6 and some future works will be also projected.

II. EYE PHYSIOLOGY AND APPLIED TOOLS

A. Eye physiology

As proposed by *Litzinger*, the eye is a tiny organ composed of structures that work with respond to others. For example, during its movements, the contraction of the muscles, the dilating, and the narrowing occur [8]. This results in a variation on the electric voltage measured at the eyes boundaries.

Noor et al. presented, in the article entitled "Study the different level of eye movement based on electrooculography (EOG) Technique", the effect of different levels of eye movement strength on the EOG signal. They found out that the

values of the EOG range from 0.05mV to 3.5mV and the frequencies vary from 0 to 100 Hz [9]. The sensors are placed as shown in figure 1. Base on this figure, one can record the five eye movements: up, down, left, right and blink.

In order to measure the left and the write movements, one can measure the voltage difference between HL and HR. as for the up and down movements, the measurement must be made between VU and VL.
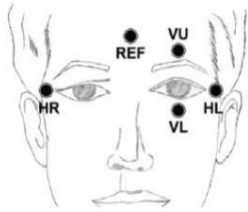
Fig. 1. *Position of the sensors when monitoring eye movements*

B. Biosensors

Biosensors are sensors connected to human body. Special attention should be attributed to these sensors as being non invasive and with no secondary effects.

Two types of biosensors exist: the needle electrodes which require a medical professional to insert them and the surface electrodes which can be mounted by any person but they present lots of noise in the delivered signal.

Once again, as our main target is to develop a low cost system, we have used the surface electrodes that require the insertion of a conductive gel between the skin and the electrode to limit (without eliminating) the noise effects. Figure 2 shows the used electrodes.

Fig. 2. *Position of the sensors when monitoring eye movements*

C. Electrophysiological system

In order to record the eye movements, the EEG (Electroencephalogram) and the EOG (Electrooculogram) can be used. Although the very accurate results both systems deliver, the cost of using such systems is very high. Added to that, the features found in these systems are very developed and will not be used for our system.

Hence, we decided to do our own system that will consist on capturing the values coming from the biosensors and processing the signal as shown in figure 3. The design of the electric circuits needed to do the required tasks will be defined in details in the next section.

III. HARDWARE DESIGN

A. Block diagram

Figure 4 shows the system which is based on three parts:

- the biosensors which read the eye movements and transform it to electrical voltage;

- the processing and transmission circuit which captures, processes, analyses and transmits the commands to the PC via a Bluetooth communication;

- the PC where the powerpoint presentation is running;
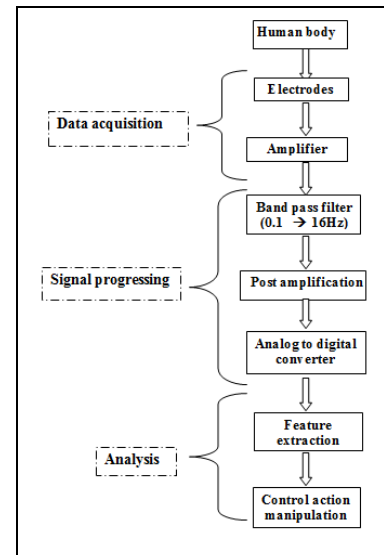
Fig. 3. *Flowchart for the signal processing method*
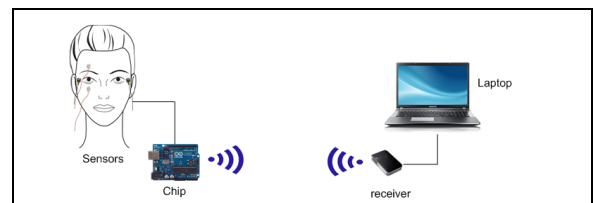
Fig. 4. *Interconnections of the micrcontroller.*

B. Analog filters

Once the signal is captured from the electrodes, its value is in the range of few microvolts. The first step, as indicated in figure 3, is to pre-amplify it. A high amplification would damage the original signal and would not allow recovering the part coming from the eye movement. Hence, amplification by a factor of 50 was used.

The second step of this process required the use of a band pass filter to filter the signals whose frequencies range between 0.1 and 16 Hz. To do so, the cutoff frequencies used were at 0.01Hz (for the low cutoff frequency) and 35Hz (for the high cutoff frequency). A Sallen Key filter was used whereas the filter order is 8 (second order for the low pass and sixth order for the high pass). The gain of this filter is 10.

As the applied gain is still not sufficient, a third step was deployed in order to post amplify the filtered signal. This stage gain is 50. Thus, the system overall gain would be about 25.000 which will allow to have an output signal in the range of hundreds of millivolts.

This signal will be then transmitted to the microcontroller through its AD module in order to convert it to digital signal and to do further processing.

Note here that although a high selective filter was used with very accurate electrical components (for example, the AD620 operational amplifier was used for the pre- and post-filtering as this integrated circuit is known for the low noise ratio it delivers), the output signal is still noisy. A comparator could have been used in order to identify if the user has made a blink or the eye were moving but we have chosen to use the ADC feature of the microcontroller.
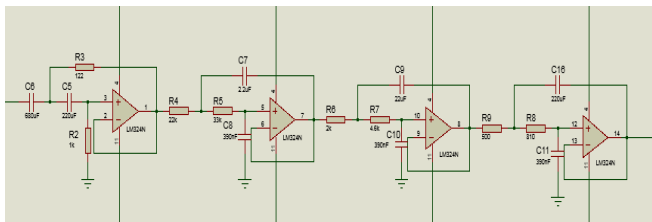
Fig. 5. *Electronic circuits used to filter the signal coming from the electrodes*

IV. MICROCONTROLLER

Several choices could have been deployed for this application. The cost, the simplicity of use and the size were the major aspects taken into consideration. Thus, the Arduino microcontroller was used for this application. In fact, the Analog to Digital module was used as well as the Bluetooth communication module.

As for the interconnections, two analog inputs are used for the electrodes. In fact, the outputs of the electric circuit are connected to A0 and A1. The difference between these two values represents the voltage output. As for the outputs, the TX/RX pins are used to connect the Bluetooth module whereas pins 2 and 4 are connected to the wireless device that controls the powerpoint (as will be shown in figure 7).

A. Flowchart

This section will be dedicated to show the whole system and to present some results after the several tests made.

As we need two different signals to control the powerpoint presentation (one to move to the next slide and the other to move to the previous slide), we have chosen to use the following two eye movements:

- A long blink (that is identified by the output voltage) to move to the next slide;

- Two consecutive blinks (with a delay less than 3 seconds between them) to move to the previous slide;

A normal blink has an average output voltage (delivered at the output of the electric circuit) of 3.2 V whereas a long blink

has an average output voltage of 4V. Thus, we have set 3V to detect a blink movement. If the value is above 3.5V, this blink is considered as a long blink and the slide will move forward. Otherwise, we will be waiting for another blink (without differentiating if it is normal or long) for the next three seconds to move the slide backward.

Note that a special attention was considered especially for the left and right eye movements as the electrode output can't differentiate between these movements. However, the output voltage of these movements ranges between 1.5V and -1.5V as an average, which is far from the values of the blink movements.
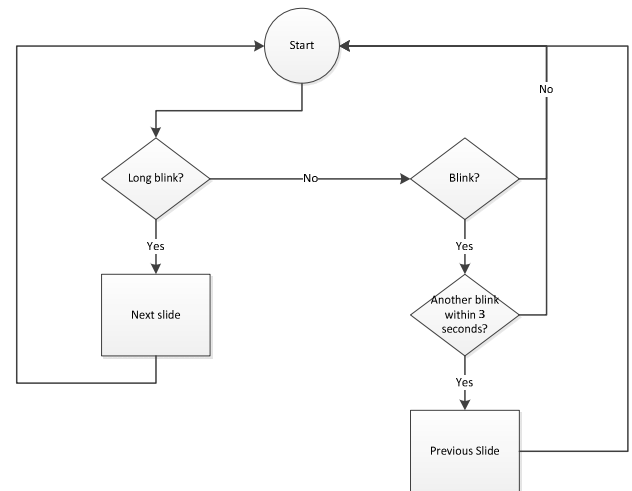
Fig. 6. *Flowchart of the microcontroller program.*

V. SYSTEM ASSEMBLY

This section will be dedicated to show the whole system and to present some results after the several tests made.

Let's start first by presenting some images from the developed system. So, figure 7 shows the whole system whereas figure 8 presents a person wearing it.
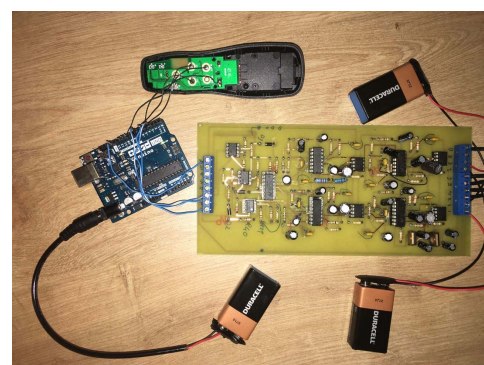
Fig. 7. *Whole assembled system.*

From figure 8, we can recognize that the user is wearing only three electrodes. In fact, as we are only searching for the blink movement, we can place the electrodes at REF, HL and HR or REF, VU and VL. VU or HL are connected to the +Vin

of the op. amp., VL or HR are connected to the –Vin of the op. amp. and REF will be connected to the system ground.

Fig. 8.   *Photo of a person using this system*

As for the results, the system was tested over 80 persons while taking into consideration the following aspects:

- Color skin (white, corny and black);

- Age (ranging between 20 and 55 years);

- Fatigue (tests made at the morning, at noon and at the night);

- Light mode (dark, medium light, high quantity of light);

- Eyes (wearing medical lenses, glasses, nothing);

- Normal and paralyzed people;

The differences between the obtained values were very small. Hence, table 1 summarized the recorded values of 7 participants.

TABLE I.   MEASURED VALUES WHEN TESTING THE SYSTEM ON SEVERAL SUBJECTS

Movements (lightened room/dark room) Subject at rest	S1	S2	S3	S4	S5	S6	S7
Upwards	3.2 V	2 V	3V	1.5V	2V	2.1V	2.1V
Downwards	-1.6 V	-1 V	-2.1V	-1.8V	-1V	-1.5V	-1.8V
Right	2 V	1.2 V	1.8V	1V	1V	1V	1.5V
Left	-2 V	-1 V	-1.6V	-1V	-1.8V	-1V	-1.9V
Long Blink	6 V	4.5 V	4.1V	2V	4V	6V	3.8V
Normal Blink	3.8V	2.9V	3.4	1.7V	2.7V	3.7V	2.9V

When comparing the obtained values to the thresholds set in the microcontroller, one can see that the considered values work well for all subjects except subject 4 where all output voltages are very low compared to the average voltage.

Added to that, one can see clearly that the difference between the upward (moving the eye up) and the blink is big and both signals can be differentiated.

VI. CONCLUION AND FUTURE WORKS

As a summary, we were able to develop a system, based on some eye movements, to control a powerpoint presentation. This system relays on the use of biosensors (electrodes), an electrical circuit to process the signal coming from the electrodes and a microcontroller to deliver the desired output in order to control the powerpoint. The accuracy of this system was about 90% after testing it on several persons taking into consideration their age, skin color, fatigue, the lenses or glasses they are using and the quantity of light inside the room where the tests have been made.

As a future work, lots of ideas can enrich this work. We will list the following:

- Design of a mask that embeds these sensors in order to facilitate the use of the system;

- Develop a genetic algorithm that could control the powerpoint presentation and which will lead to a lower error percentage;

- Apply the eye movement system on other manual system in order to facilitate the life of humans, especially the ones with movements disabilities;

- Work on developing this system using a smaller size to be handled easier;

REFERENCES

[1] K. Otsuki, H. Ohmata, A. Fujii, K. Majima and T. Inoue, "A method of controlling presentation for applications in hybridcast," in *IEEE International Conference on Consumer Electronics (ICCE)*, Berlin, Germany, 2012.

[2] F. Hsu, W. Lin, G. You, T. Tsai and H. Huang, "U-Garden: An interactive control system for multimodal presentation in museum," in *IEEE International Conference on Multimedia and Expo (ICME)*, Singapore, 2010.

[3] P. Suta, P. Mongkolnam and W. Eamsinvattana, "Controlled and collaborative presentation via tablet PCs for classroom and meeting room uses," in *12th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, Songkhla, Thailand, 2015.

[4] M. Challagundla, k. Reddy and N. Harsha Vardhan, "Automatic motion control of powered wheel chair by the movements of eye blink," in *International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)*, Ramanathapuram, India, 2014.

[5] H. Septanto, A. Prihatmanto and A. Indrayanto, "A computer cursor controlled by eye movements and voluntary eye winks using a single channel EOG," in *International Conference on Electrical Engineering and Informatics (ICEEI)*, Malaysia, 2009.

[6] J. Hori, K. Sakano and Y. Saitoh, "Development of communication supporting device controlled by eye movements and voluntary eye blink," in *26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEMBS)*, Hangzhou, China, 2004.

[7] A. Raheel, S. Anwar, M. Majid, B. Khan and E. Haq, "Real time text speller based on eye movement classification using wearable EEG sensors," in *SAI Computing Conference (SAI)*, London, 2016.

[8] T. Litzinger and K. Del Rio-Tsonis, "Eye Anatomy," Miami, 2002.

[9] N. M. Mohd Noor and M. Qawiem bin Kamarudin, "Study the Different Level of Eye Movement Based on Electrooculography (EOG) Technique," *Advanced Science Letters,* 2010.

Diversifying Search Result Leveraging Aspect-based Query Expansion

Md Shajalal[a*]   Masaki Aono[b*]   Muhammad Anwarul Azim[c*]

shajalal@baust.edu.bd   aono@tut.jp   azim@cu.ac.bd

[*]Department of Computer Science and Engineering

[a]Bangladesh Army University of Science and Technology, Nilphamari, Bangladesh

[b]Toyohashi University of Technology, Toyohashi, Aichi, Japan

[c]University of Chittagong, Chittagong, Bangladesh

ABSTRACT

Web search queries are short, ambiguous and tend to have multiple underlying interpretations. To reformulate such queries, query expansion is a prominent method that leads to retrieve a set of relevant documents. In this paper, we propose an aspect-based query expansion technique for diversified document retrieval. At first, query suggestions and completions are retrieved from major commercial search engines. A frequent phrase-based soft clustering algorithm is then applied to group similar retrieved candidates into clusters. Each cluster represents different query aspect. The expansion terms are selected from the generated cluster labels for each cluster. To estimate the relevancy between the expanded query and the documents, multiple new lexical and semantic features are introduced using the content information, and word-embedding model, respectively. Finally, a linear ranking approach is employed to re-rank the documents retrieved for the original query using the extracted features. We conduct experiments on Clueweb09 document collection using TREC 2012 Web Track queries. The experimental results clearly demonstrate that our proposed aspect-based query expansion method is effective to diversify the retrieved documents and outperformed baseline and some known related methods in terms of diversity metrics ERR-IA, α-nDCG and NRBP at the cut of 20.

KEYWORDS

Query Ambiguity, Query Expansion, Diversified Search, Query Aspect, and Word Embedding

1 INTRODUCTION

Web search has become the predominant method for users to fulfill their information needs. In this aspect, users describe their information needs by providing a set of keywords. These keywords are collectively called a search query for each user. Since expressing information need through keywords is difficult, some users fail to choose the precise terms while others tend to omit important terms needed to clarify search intentions [1, 2]. Therefore, a large number of the web search queries are usually short, ambiguous, and prone to have multiple interpretation [3, 4, 5]. Generally, the short queries mean a lot of ambiguity as to what information needs the users express. Consider a short and ambiguous query "*Java*", which could be interpreted as a programming language, island, coffee, etc.

For such type of queries, the search engine may generate a ranking of documents with maximum redundancy covering a very few user information needs. To mitigate these issues, search result diversification (SRD) can be used to generate the effective ranking of documents. Some clustering algorithms applied in different perspectives [6, 7] can be used for SRD. Diversification approaches re-rank the retrieved documents considering intents or aspects for the user query. Therefore, the retrieved documents contain less redundant documents. In turns, the retrieved documents also cover user query aspects as much as possible. The common principle used in the existing SRD approaches is to select as diverse results as possible from a given set of retrieved documents. The final ranking list is much dependent on the initial retrieval results, which may not have a good coverage of the different aspects of the query.

To overcome this drawback, some existing studies on SRD attempted to expand the original query before diversifying the results [8, 9].

Query expansion is a classic technique to reformulate the query, which generates diversified expansion terms to enhance the original query. There are lot of approaches to expand the original query using different resources and techniques including pseudo-relevance feedback [10], word-embedding [11, 12, 13], ConceptNet and WordNet [14], and Freebase [15] etc. Query expansion techniques are widely applied for improving the efficiency of the textual information retrieval systems. These techniques help to overcome vocabulary mismatch issues by expanding the original query with additional relevant terms and reweighing the terms in the expanded query.

In this research, we propose an aspect-based query expansion technique to diversify the retrieved documents for the original query. Query suggestions and completions from search engines are good resources to reformulate the original query. Therefore, our proposed approach is to retrieve query suggestions and completions for each query from three commercial web search engines namely, Google, Yahoo, and Bing. The aggregated list of suggestions and completions are used as a resource to expand the original query. A frequent phrase based soft clustering algorithm is then applied to group similar candidates into clusters. Here every cluster represents different query aspect. The generated cluster labels are then used to expand the query. We employ all the terms from the cluster labels except query terms and stop words to expand the query. Finally, the retrieved documents are re-ranked based on the relevancy with the expanded query. To estimate the relevancy between web document and the query, we propose multiple semantic and lexical features using word-embedding and their content information, respectively. We conducted experiments using the Clueweb09 document collection with TREC 2012 Web Track query set. The experimental results clearly illustrate that our proposed aspect-based query expansion method is effective to diversify web documents. There are two distinct contributions in proposed method:

1. A novel query expansion technique based on users' aspect and

2. Multiple new semantic and lexical features to estimate the relevance between expanded query and documents

The rest of the paper is structured as follows: In section **2**, we summarize related work on query expansion and document retrieval. In section **3**, we briefly explain two classical retrieval model. We present our proposed method in section **4**. The experiments and evaluation to show the effectiveness of our proposed method is presented in section **5**. Some concluded remarks and future directions are described in section **6**.

2 RELATED WORK

Usually, queries to web search engines are short and not written carefully, which makes it more difficult to understand the intent behind a query and retrieve relevant documents. A common solution is query expansion, which uses a larger set of related terms to represent the user's intent and improves the documents' ranking.

Pseudo Relevance Feedback (PRF) algorithms are widely used in query expansion. These algorithm assume that top ranked documents for the original query are relevant that contain good expansion terms. The researchers proposed a model that selected expansion terms based on their term frequency in top retrieved documents, and weights them by documents' ranking scores [16]:

$$s(q) = \sum_{d \in D} p(t|d) f(q, d)$$

where D is the set of top retrieved documents, $p(t|d)$ denotes the probability that term t that generated by document d's language

model, and $f(q, d)$ denotes the ranking score of the document provided by the retrieval model. Later, another study [17] added inverse document frequency (IDF) to demote very frequent terms:

$$s(q) = \sum_{d \in D} p(t|d) f(q, d) \log \frac{1}{p(t|C)}$$

where $p(t|C)$ denotes the probability of term t in the corpus language model C.

Another PRF approach has also been proposed using a Mixture Model [18]. In that study, the researchers assumed that the terms in top retrieved documents are drawn from a mixture of two language models: query model θ_q and a background model θ_B. The likelihood of a top retrieved document d is defined as follows:

$$\log p(d|\theta_q, \alpha_d, \theta_B) = \sum_{t \in D} \log(\alpha_d \, p(t|\theta_q)$$
$$+ (1 - \alpha_d) \, p(t|\theta_B))$$

α_d denotes a document-specific mixture parameter. For this equation, the query model θ_q can be learned by maximizing the top retrieved documents' likelihood. The terms that have non-zero probability in θ_q are used for query expansion.

Knowledge base such as Freebase can also be applied by query expansion methods [15]. Those methods identified the entities associated with the query, and used the entities to perform query expansion. A supervised model combined information derived from Freebase descriptions and categories to select terms that are effective for query expansion. The researchers also proposed a method to expand the original query with the help of WordNet and ConceptNet [14]. Their approach extended the query with the synonyms generated from WordNet.

Recently, word-embedding techniques are used for query expansion [11, 12, 13]. An Automatic Query Expansion (AQE) framework has been proposed by using distributed neural language model, word2vec [13]. They

trained a word2vec model that learned a low dimensional embedding for each vocabulary entry using the semantic and contextual relation in a distributed and unsupervised approach. They selected the related terms to the query by applying a K-nearest neighbor technique and those terms were used for expansion. A query expansion technique is introduced for *adhoc* retrieval using a locally trained word embedding model [11]. They presented local embedding which capture the nuances of topic-specific language better than global embeddings. Another study also proposed word-embedding based method for query expansion [12]. They applied continuous-bag-of-words implementation of word2vec over the entire corpus on which search is performed and selected terms that are semantically related to the query. Their method either used the terms to expand the original query or integrate them with the effective pseudo-feedback-based relevance model.

3 CLASSICAL RETRIEVAL MODEL

In this section, we discuss two classical retrieval models, Okapi BM25 [19] and Language model [16]. In this research, these two classical retrieval models are utilized as baseline document retrieval.

3.1 Okapi BM25 Model

Let d be an unstructured document in the collection C. We may consider this as a vector $\vec{d} = (tf_1, \cdots, tf_V)$, where tf_i denotes the term frequency of the i-th term t_i in the document d and V is the total number of terms in the vocabulary. In order to score such a document against a query, most ranking functions define a term weighting function $w_j(\vec{d}, C)$. BM25 is an example of such functions. For ad-hoc retrieval, and ignoring any repetition of terms in the query, BM25 [19] can be simplified as follows:

$$w_j(\vec{d}, C) = \frac{(k_1 + 1) \cdot tf_j}{k_1(1 - b) + b\frac{dl}{avdl}} \log \frac{N - df_j + 0.5}{df_j + 0.5} \tag{1}$$
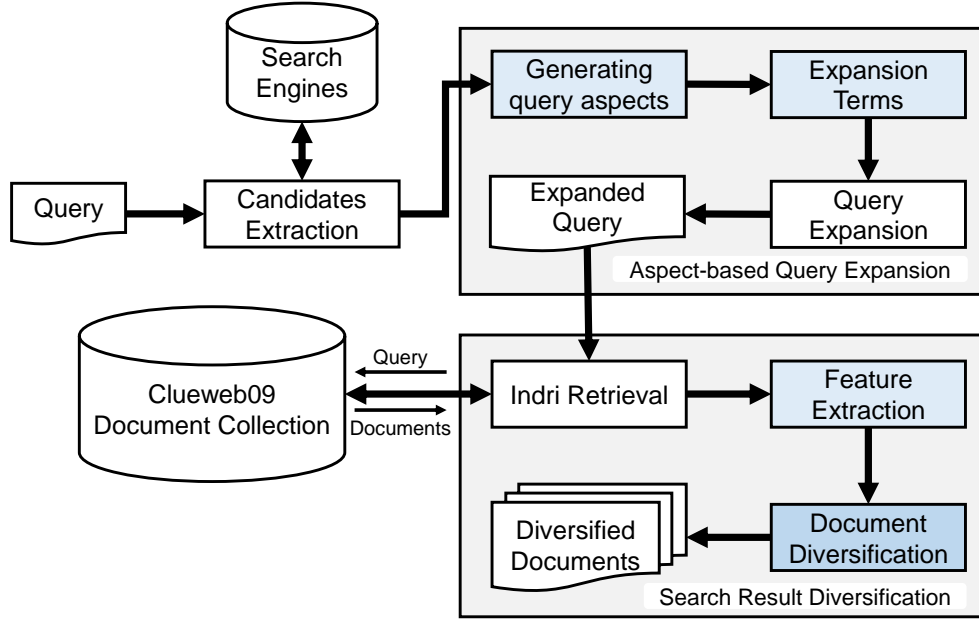
Figure 1. An Aspect-based Query Expansion Framework for SRD

where df_j is the document frequency of *j*-th term, dl is the document length, $avdl$ is the average document length in the collection, and k_1 and b are tuning parameters.

The document score is then obtained by adding the document term weights of term matching the query q:

$$W(\vec{d}, q, C) = \sum_j w_j(\vec{d}, C) \cdot tf_{q_j} \qquad (2)$$

3.2 Language Model

Language model is a quite general formal approach in information retrieval. Query likelihood model is the most basic method for using language models in information retrieval. Let us assume a simple unigram for each document, where each document is represented as the standard bag-of-words and their language model is distributed over a vocabulary of a single word. The maximum likelihood estimate of term w occurring in document d for a multinomial distribution is given below [16]:

$$P_{ML}(q|d) = \frac{tf_{w,d}}{|d|} \qquad (3)$$

where $tf_{w,d}$ is term frequency (number of times term w appears in document d) of the term w in document d and $|d|$ denotes the total number of terms in d.

Given a query $q = \{q_1, q_2, q_3...q_k\}$, the likelihood can be computed for the document d as follows:

$$P(q|d) = \prod_{i=1}^{k} P(q_i|d) \qquad (4)$$

This likelihood is computed for each document and used for ranking. Ranking documents in this procedure is known as query likelihood language model.

The smoothed $p(w|d)$ with Jelinek-Mercer smoothing is estimated as follows:

$$p(w|d) = \lambda \frac{tf_{w,d}}{|d|} + (1 - \lambda)\frac{cf_w}{|C|} \qquad (5)$$

where cf_w is the term frequency of term w in the collection C and λ is a smoothing parameter.

4 OUR PROPOSED METHOD

This section presents our proposed aspect-based query expansion method for search result diversification. For a given query, our method produces a diversified list of documents using aspect-based query expansion. The high-level building blocks of our proposed method are illustrated in Fig. 1. There are two major parts in our method, *query expansion*

Table 1. Multiple candidates reflect one aspect.

Query	Candidate	Aspect
grilling	memorial day grilling recipes easy grilling recipes grilling recipes	Grilling Recipes
	grilling chicken grilling chicken breasts grilling chicken leg	Grilling Chicken
	grilling corn on the cob grilling corn	Grilling Corn
	grilling lobster tails grilling lobster	Grilling Lobster
	grilling tips outdoor grilling tips perfect grilling tips	Grilling Tips

and *search result diversification* using the expanded query.

Given a query, the *query expansion* technique expands the original query covering user query aspects as much as possible. In this regards, we propose a new aspect-based approach to select expansion terms. The *search result diversification* technique returns a list of diversified documents with respect to the expanded query. In this part, multiple new semantic and lexical features are introduced to estimate the relevancy between the query and document. The remainder of this section presents the complete explanation of the *query expansion* and the *search result diversification* techniques.

4.1 Aspect-based Query Expansion

Generally, search queries are very short in length. The existing study on query structure suggested that, the average length of search queries is around 2.3 terms per query [20]. Usually, the short queries mean a lot of ambiguity as to what information needs the users express. Therefore, we reformulate the original query by appending more related terms that reflect different user aspects. This process is called query expansion. Here our expansion method tries to select terms with various aspects underlying a query as much as possible. Our hypothesis is that, if the query covers more

user aspects, the task of estimating the relevancy between the expanded query and documents will be more easier. The original query is expanded using the following three steps.

4.1.1 Candidate Extraction

Query reformulation with related query suggestions and completions are more effective for searching the most relevant documents that maximize the coverage [8, 21]. The query suggestions and completions from commercial search engines are employed as a resource to find the expansion terms. We retrieve query suggestions and query completions from major search engines (Google, Yahoo, and Bing) for a given query. Then the duplicates are filtered out after aggregating all suggestions and completions.

4.1.2 Generating Query Aspects

Multiple query suggestions and completions may contain candidates which reflect the same query aspect. Our observation on this aspect is that a group of candidates covers similar query aspect rather covering unique aspects. Table 1 depicts an example where we can see that multiple candidates represent the same aspect. In this table, we can see five aspect of query "*grilling*" including "*grilling recipes*", "*grilling chicken*", "*grilling corn*", "*grilling*

Table 2. lexical features

Type	Features	Description
Lexical Features	01. $f_{LS}(q_{exp}, d)$	Lexical similarity based on edit distance
	02. $f_{TO}(q_{exp}, d)$	% of overlapping query terms
	03. $f_{SynO}(q_{exp}, d)$	% of overlapping synonym of query terms.
	04. $f_{BR}(q, d)$	Baseline rank of each individual document
	05. $f_{VisTerm(d)}$	Number of visible terms on the page
	06. $f_{TTerm}(title(d))$	Number of terms in the page <title>field
	07. $f_{avgTL}(d)$	Avg. length of visible terms on the document
	08. $f_{fracAT}(d)$	Fraction of anchor text on the document
	09. $f_{fracVT}(d)$	Fraction of visible text on the document
	10. $f_{fracS}(d)$	Stopword and Non-stopword ratio

lobster", and "*grilling tips*".

A soft clustering technique is then applied to the candidates based on frequent phrases to identify the query aspects. We make use of Lingo Clustering algorithm [22] to group the candidates into clusters. Some candidates may belong to more than one cluster. Then we used the cluster labels generated by the clustering algorithm as query aspects.

4.1.3 Selecting Expansion Terms

For a given query q, let assume that $L_q = \{l_1, l_2, l_3, ..., l_K\}$ be the set of cluster labels generated by the previous section. We generate and select the expansion terms from these labels to expand the query. To select the expansion terms from the generated cluster labels, we introduce a expansion term selection algorithm. The pseudo-code of our expansion term selection (ETS) algorithm is as follows:

Algorithm 1: Expansion Term Selection: ETS(q, L_q)

Input: Set of generated cluster labels for query q, $L_q = \{l_1, l_2, l_3, ..., l_K\}$
Output: Set of expansion terms, E_t
$E_t = [\emptyset]$;
for *each term $t \in L_q$* **do**
    if $t \in l$ && $t \notin q$ **then**
        | $E_t = E_t \cup t$
    end
end

---

where E_t denotes the set of expansion terms, l is the cluster label, and t is the term in label l. $t \in l$ and $t \notin q$ state that term t exists in l and t does not exist in q, respectively.

4.1.4 Query Expansion

Let $E_t = \{t_1, t_2, t_3, ..., t_n\}$ be the set of selected terms for query q. We make use of these terms to expand the query. We append the selected terms with the query q and the expanded query is as like as follows:

$$q_{exp} = q$$
$$for\ each\ term\ t;\ t \in E_t \qquad (6)$$
$$q_{exp} = q_{exp} \cup t$$

where q_{exp} denotes the expanded query.

4.2 Search Result Diversification

This section presents the diversification approach to re-rank the retrieval result for original query. The documents are re-ranked with respect to their relevancy between the expanded query q_{exp} and the documents. We first extract multiple semantic and lexical features, then we apply a linear ranking function to rank the documents. Since the expanded query covers multiple query aspects, the re-ranked documents satisfy the diversity.

4.2.1 Feature Extraction

The lexical and semantic features are estimated using WordNet [23] and the content information of document and expanded query and the

Table 3. Semantic features

Type	Features	Description
Semantic Features	01. $f_{MPP}(d)$	$\frac{\sum_{t \in d} I(POS(t) \in M)}{|d|}$
	02. $f_{ACS}(q_{exp}, d)$	$\frac{\sum_{t_i \in q_{exp}} \sum_{t_j \in d} ConSim(t_i, t_j)}{|q_{exp}| * |d|}$
	03. $f_{SIM_{w2v}}(q_{exp}, d)$	$\frac{\sum_{t_i \in q_{exp}} \sum_{t_j \in d} cosine(\vec{t_i}, \vec{t_j})}{|q_{exp}| * |d|}$

pre-trained word2vec[1] model on *Google News Corpus*. The lexical and semantic features are summarized in Table 2 and 3, respectively.

The notations in the Table 3 are defined as follows:

- In Meaningful POS (part-of-speech) percentage (MPP) feature $f_{MPP}(d)$, $I(POS(t) \in M)$ returns 1 if the POS of term t belongs to the set $M = \{Noun, Verb, Adjective, Adverd\}$ [24].

- In average concept similarity (ACS) feature $f_{ACS}(q_{exp}, d)$, $ConSim(t_i, t_j)$ returns the conceptual similarity between term t_i and t_j [24].

- In similarity (SIM) feature $f_{SIM_{w2v}}(q_{exp}, d)$ based on *Word2Vec*, $\vec{t_i}$ and $\vec{t_j}$ denote the 300 dimensional vector representation of term t_i and t_j from pre-trained word2vec model, respectively.

We make use of *MinMax* normalization to normalize the features value into the range [0,1] as follows:

$$\bar{x} = \frac{x - min(x)}{max(x) - min(x)}$$

where x is the feature value and $\bar{x}$ is the normalized feature value. $min(x)$ and $max(x)$ denote the minimum and maximum feature values of a specific feature, respectively.

4.2.2 Document Ranking

To re-rank the retrieved documents for original query, we estimate the document relevancy using a linear ranking approach considering all extracted features as follows:

$$
\begin{aligned}
Rel(q_{exp}, d) &= \frac{\sum_{i=1}^{N} wt_i \cdot f_i(q_{exp}, d)}{\sum_{i=1}^{N} wt_i} \\
&= \frac{wt_1 \cdot f_1(q_{exp}, d)}{\sum_{i=1}^{N} wt_i} + \frac{wt_2 \cdot f_2(q_{exp}, d)}{\sum_{i=1}^{N} wt_i} \\
&+ \cdots + \frac{wt_N \cdot f_N(q_{exp}, d)}{\sum_{i=1}^{N} wt_i}
\end{aligned}
$$

$$(7)$$

where $f_i(q_{exp}, d)$ denotes the *i*-th feature and wt_i denotes the feature importance. The higher the value of $Rel(q_{exp}, d)$, the higher the relevancy of the document d with the expanded query q_{exp} is.

5 EXPERIMENTS AND EVALUATION

This section presents the details of the experiments and evaluation results of our proposed method on a standard dataset and compare with some known related methods.

5.1 Data Collection

We use the Web Track dataset from TREC 2012 [25]. There are 50 queries, each of which includes 3 to 8 subtopics identified by TREC assessors. All experiments are conducted on ClueWeb09 [26] collection. We used query suggestions and completions from Googe, Yahoo, and Bing search engine provided by NTCIR-10 English subtopic mining dataset [27]. The pre-trained Word2Vec model using Google News Corpus are employed to extract semantic features. To find the POS

of each term, we make use of Stanford NLP Parser. We make use of Indri search engine [28] to retrieve top 500 documents from the clueweb09 collection.

5.2 Evaluation Metrics

Several metrics have been used in order to evaluate the diversification effectiveness of search engines. A good diversification system is the one that satisfies multiple information needs (or user intents) underlying a query that is submitted to that system by different users, or by the same user in different contexts. In the context of search result diversification, a query is represented by a set of subtopics or aspects (which generally correspond to user intents). The relevance of a document with respect to a query is judged separately for each subtopic, and is estimated by the ability of that document to cover different subtopics of the same query. In this research, we utilized three diversity metrics which are official in the diversity task of TREC Web track.

5.2.1 α-nDCG (α-normalized Discriminative Cumulative Gain)

α-nDCG@k [29] is computed as follows:

$$\alpha - nDCG@k = \frac{\alpha - DCG@k}{\alpha - DCG'@k}$$

where $\alpha - DCG'@k$ is a normalization factor corresponding to the maximal value of $\alpha - DCG@k$ that gives the ideal document ranking. $\alpha - DCG@k$ is computed as follows:

$$\alpha - DCG@k = \sum_{j=1}^{k} \frac{\sum_{s \in S(q)} rel(d_j, s)(1-\alpha)^{\sum_{i=1}^{j-1} rel(d_i,s)}}{log_2(1+j)}$$

where the parameter α ($\alpha \in [0,1]$) represents the user satisfaction factor for the set of documents that have been already browsed by the user. This parameter (α) is generally fixed to 0.5. q is a query, $S(q)$ is the set of subtopics underlying q, and d_i (resp. d_j) is the document ranked at the ith (resp. jth) position. $rel(d, s)$ is a function that evaluates the relevance of a document d with respect to a given subtopic s. Note also that $\alpha - nDCG$ considers the set of already $(k-1)$ selected documents when evaluating a document at position k. This means that

the metric takes into account the dependency between the returned documents. Finally, note that $(1-\alpha)^{\sum_{i=1}^{j-1} rel(d_i,s)}$ penalizes the coverage of already covered aspects of the query and α controls the amount of penalization.

5.2.2 ERR-IA (Expected Reciprocal Rank - Intent Aware)

ERR-IA(q, D) [30] for a given query q and over a set of returned documents D with respect to q is defined as follows:

$$ERR - IA@k = \sum_{s \in S(q)}^{k} p(s|q) \cdot ERR(s, D)$$

where $ERR(s, D)$ is the expected reciprocal rank and $p(s|q)$ denotes the importance of subtopic s regarding to the query q (the more popular the subtopic s for q, the higher is $p(s|q)$.

5.2.3 NRBP (Novelty and Rank-Biased Precision)

NRBP [31] is an extension of the RBP (Rank-Biased Precision) metric [32]. The basic intuition that NRBP uses is that, the user has some specific intent and is generally interested in one particular aspect of the query, at least at that time. NRBP is defined as follows:

$$NRBP = \frac{1 - (1-\alpha)^\beta}{N}$$
$$\cdot \sum_{k=1}^{\infty} \beta^{k-1} \cdot \sum_{i=1}^{N} J(d_k, i)(1-\alpha)^{C(k,i)}$$

where d_k denotes the kth document, N is the (possible) number of aspects of a given query, $J(d, i) = 1$ if document d is relevant to the ith aspect of the query, and $J(d, i) = 0$ otherwise, $C(k, i)$ is the number of documents at cut-off k that have been judged to be relevant to the ith aspect of the query, parameter $\beta \in [0,1]$ is used to model the patience level of the user, and parameter $\alpha \in [0,1]$ refers to the user declining interest.

In short, we use the above three official diversity evaluation metrics used in TREC Web Track.

Table 4. Summary of all experimental settings.

Run	Description
LM	Documents retrieved using language model
BM25	Documents retrieved with BM25 retrieval model
QFLR	Original query and linear ranking with all features
$Q_{n\_exp}FLR$	Query expansion with suggestions's terms and linear ranking with all features
$Q_{a\_exp}LFLR$	Aspect-based query expansion and linear ranking with lexical features
$Q_{a\_exp}FLR$	Aspect-based query expansion and linear ranking with lexical and semantic features

Table 5. Experimental results of our method, baseline and some known related methods on TREC Web Track 2012 in terms of ERR-IA, α-nDCG, and NRBP at cut of 20. **Boldface** indicates the **best** performance ammong all.

Type	Method	ERR-IA@20	α-nDCG@20	NRBP
Baseline Retrieval	BM25 [19]	0.2253	0.3105	0.1738
	LM [16]	0.157889	0.2143	0.1237
Our Method	**$Q_{a\_exp}FLR$**	**0.3447**	**0.4438**	**0.3033**
	$Q_{a\_exp}LFLR$	0.3015	0.3925	0.2685
	$Q_{n\_exp}FLR$	0.2475	0.3475	0.2348
	QFLR	0.2354	0.3257	0.1925
Related Methods	ICTNET [25]	0.326	0.422	0.280
	udel [25]	0.325	0.419	0.282
	LIA [25]	0.318	0.424	0.268
	udel fang [25]	0.300	0.420	0.241

5.3 Experimental Results

To measure the effectiveness of our aspect based query expansion method for SRD, we carried out experiments using different experimental settings. We retrieved top 500 documents for each query by using two baseline retrieval model, Language model with Jelinek-Mercer smoothing and Okapi BM25. We already described these to classical model in section 3. We used the parameter *mu*=2500, *lambda*=0.4 for Jelinek-Marcer smoothing and *k1*=1.2, *b*=0.75 and *k3*=7 for BM25 based method. We denote these two experimental settings as LM and BM25, respectively.

We carried out experiments using our re-ranking method presented on section 4.2 with documents retrieved by BM25. In setting QFLR, we re-ranked the retrieved documents by using a linear ranking method (i.e Eq. 7)with extracted features with respect to the original query. Then we expanded the

query with the query suggestions' terms except stopwords and query terms. This setting is denoted by $Q_{n\_exp}FLR$. Then we applied our aspect-based query expansion method to expand the query and linear ranking with only lexical features to re-rank the documents in setting $Q_{a\_exp}LFLR$. Finally, in setting $Q_{a\_exp}FLR$, we used all features instead for lexical features to re-rank the documents with respect to the expanded query q_{exp}. Table 4 summaries the description of all experimental settings.

The performance of our proposed methods, baseline, and some related [25] methods on TREC Web track 2012 dataset in terms of three diversity metrics including ERR-IA, α-nDCG, and NRBP at the cut of 20 are reported in Table 5. Boldface indicates the best performance among all methods. We can see that, our two aspect-based query expansion methods $Q_{a\_exp}LFLR$ and $Q_{a\_exp}FLR$ performed better
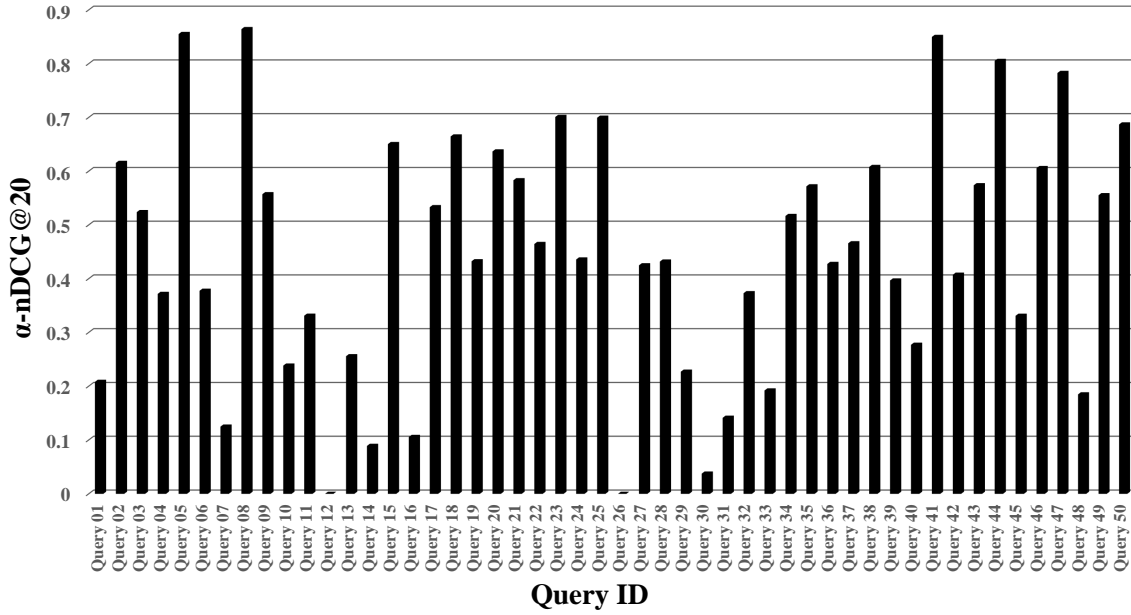
Figure 2. The query-wise performance of our proposed method method in terms of α-nDCG@20. The X-axis represents individual Query-ID and Y-axis represents the value of α-nDCG@20 for each individual query.

than all other methods including the normal query expansion based method Q_{n_exp}FLR. Therefore, we can conclude that aspect-based query expansion can capture different query aspects which helps to increase the diversity of the ranking. Our proposed semantic features are not applied in setting Q_{a_exp}LFLR whereas those are applied in setting Q_{a_exp}FLR. The experimental results clearly demonstrate that our semantic features are effective to capture better relevancy.

The query-wise performance of our method in terms of the diversity metric α-nDCG@20 on TREC Web Track 2012 dataset is depicted in Fig. 2. The figure illustrates that the performance for each individual query are varied widely. We can see that, our method achieved more than 80% accuracy (i.e Query 05, Query 09, Query 41, etc.) for several queries . For an example query "porterville" (Query 09), our method achieved 85% accuracy. The aspect-based query expansion technique selected the expansion terms for this queries are: "recorder," "college," "ca," "school," "district," "fair," "weather," "police," "department," and "courthouse". These expansion terms are distinct from each other and all are related to

the query. We can also see that each term is representing different user's query aspect for the query "porterville". The figure also conclude that our method failed for only two queries (Query 12 and Query 26). Considering the query, "dnr" (Query 12) which is an abbreviation query. The observation for this query is that our frequent phrase-based clustering algorithm was not good enough to generate meaningful cluster labels. In turns the expansion terms were not related to the abbreviation queries. We think this might be one of the plausible reasons for the failure. However, we can conclude from the experimental results on a benchmark dataset that our proposed aspect-based query expansion method and semantic features contributed to effectively diversify the documents. Our method also outperformed the baselines and known related methods in terms of all three standard diversity evaluation metrics.

6 CONCLUSION AND FUTURE DIRECTIONS

This paper proposed an aspect-based query expansion method and multiple semantic features for search results diversification. To

identify the query aspects, we applied a frequent phrase-based soft clustering technique to the query suggestions. Then we select the expansion terms from the cluster labels. We also proposed multiple semantic and lexical features to estimate the relevancy between the expanded query and the document. The experimental results on a benchmark TREC dataset clearly conclude that our proposed method is effective for search result diversification.

For future work, we have a plan to extract expansion terms from top retrieved web document for query expansion. Furthermore, it will be interesting to apply an aspect-based document diversification approach for results diversification.

REFERENCES

[1] Se-Jong Kim, Jaehun Shin, and Jong-Hyeok Lee. Subtopic mining based on three-level hierarchical search intentions. In *European Conference on Information Retrieval*, pages 741–747. Springer, 2016.

[2] Se-Jong Kim and Jong-Hyeok Lee. Subtopic mining using simple patterns and hierarchical structure of subtopic candidates from web documents. *Information Processing & Management*, 51 (6):773–785, 2015.

[3] Rodrygo LT Santos, Craig Macdonald, Iadh Ounis, et al. Search result diversification. *Foundations and Trends® in Information Retrieval*, 9(1):1–90, 2015.

[4] Kevyn Collins-Thompson, Craig Macdonald, Paul Bennett, Fernando Diaz, and Ellen M Voorhees. Trec 2014 web track overview. Technical report, MICHIGAN UNIV ANN ARBOR, 2015.

[5] Behrooz Mansouri, Mohammad Sadegh Zahedi, Maseud Rahgozar, Farhad Oroumchian, and Ricardo Campos. Learning temporal ambiguity in web search queries. pages 2191–2194, 2017.

[6] Rahmat Widia Sembiring, Jasni Mohamad Zain, and Abdullah Embong. Dimension reduction of health data clustering. *International Journal of New Computer Architectures and their Applications (IJNCAA)*, pages 1018–1026, 2011.

[7] Yongxin Zhang Hong Shen Shitian Xu. Cluster sampling for the demand side management of power big data. *International Journal of New Computer Architectures and their Applications (IJNCAA*, pages 114–121, 2016.

[8] Rodrygo LT Santos, Craig Macdonald, and Iadh Ounis. Exploiting query reformulations for web search result diversification. In *Proceedings of the 19th International Conference on World Wide Web*, pages 881–890. ACM, 2010.

[9] Rodrygo LT Santos, Jie Peng, Craig Macdonald, and Iadh Ounis. Explicit search result diversification through sub-queries. In *European Conference on Information Retrieval*, pages 87–99. Springer, 2010.

[10] Mohannad Almasri, Catherine Berrut, and Jean-Pierre Chevallet. A comparison of deep learning based query expansion with pseudo-relevance feedback and mutual information. In *European conference on information retrieval*, pages 709–715. Springer, 2016.

[11] Fernando Diaz, Bhaskar Mitra, and Nick Craswell. Query expansion with locally-trained word embeddings. *arXiv preprint arXiv:1605.07891*, 2016.

[12] Saar Kuzi, Anna Shtok, and Oren Kurland. Query expansion using word embeddings. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pages 1929–1932. ACM, 2016.

[13] Dwaipayan Roy, Debjyoti Paul, Mandar Mitra, and Utpal Garain. Using word embeddings for automatic query expansion. *arXiv preprint arXiv:1606.07608*, 2016.

[14] Arbi Bouchoucha, Xiaohua Liu, and Jian-Yun Nie. Towards query level resource weighting for diversified query expansion. In *European Conference on Information Retrieval*, pages 1–12. Springer, 2015.

[15] Chenyan Xiong and Jamie Callan. Query expansion with freebase. In *Proceedings of the 2015 International Conference on The Theory of Information Retrieval*, pages 111–120. ACM, 2015.

[16] Victor Lavrenko and W Bruce Croft. Relevance-based language models. In *ACM SIGIR Forum*, volume 51, pages 260–267. ACM, 2017.

[17] Chengxiang Zhai and John Lafferty. A study of smoothing methods for language models applied to ad hoc information retrieval. In *ACM SIGIR Forum*, volume 51, pages 268–276. ACM, 2017.

[18] Hamed Zamani, Javid Dadashkarimi, Azadeh Shakery, and W Bruce Croft. Pseudo-relevance feedback based on matrix factorization. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pages 1483–1492. ACM, 2016.

[19] Andrew Trotman, Antti Puurula, and Blake Burgess. Improvements to bm25 and language models examined. In *Proceedings of the 2014 Australasian Document Computing Symposium*, page 58. ACM, 2014.

[20] Claudio Carpineto and Giovanni Romano. A survey of automatic query expansion in information retrieval. *ACM Computing Surveys (CSUR)*, 44(1):1, 2012.

[21] Youngho Kim and W Bruce Croft. Diversifying query suggestions based on query documents. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 891–894. ACM, 2014.

[22] Claudio Carpineto, Stanislaw Osiński, Giovanni Romano, and Dawid Weiss. A survey of web clustering engines. *ACM Computing Surveys (CSUR)*, 41(3):17, 2009.

[23] Christiane Fellbaum. *WordNet*. Wiley Online Library, 1998.

[24] Md Shajalal, Md Zia Ullah, Abu Nowshed Chy, and Masaki Aono. Query subtopic diversification based on cluster ranking and semantic features. In *Advanced Informatics: Concepts, Theory And Application (ICAICTA), 2016 International Conference On*, pages 1–6. IEEE, 2016.

[25] Ian Soboroff, Iadh Ounis, Craig Macdonald, and Jimmy J Lin. Overview of the trec-2012 microblog track. In *TREC*, volume 2012, page 20, 2012.

[26] Jamie Callan, Mark Hoy, Changkuk Yoo, and Le Zhao. Clueweb09 data set, 2009.

[27] Tetsuya Sakai, Zhicheng Dou, Takehiro Yamamoto, Yiqun Liu, Min Zhang, Ruihua Song, MP Kato, and M Iwata. Overview of the ntcir-10 intent-2 task. In *NTCIR*, 2013.

[28] Bhaskar Mitra, Fernando Diaz, and Nick Craswell. Luandri: a clean lua interface to the indri search engine. *arXiv preprint arXiv:1702.05042*, 2017.

[29] Charles LA Clarke, Maheedhar Kolla, Gordon V Cormack, Olga Vechtomova, Azin Ashkan, Stefan Büttcher, and Ian MacKinnon. Novelty and diversity in information retrieval evaluation. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 659–666. ACM, 2008.

[30] Olivier Chapelle, Shihao Ji, Ciya Liao, Emre Velipasaoglu, Larry Lai, and Su-Lin Wu. Intent-based diversification of

web search results: metrics and algorithms. *Information Retrieval*, 14(6): 572–592, 2011.

[31] Charles LA Clarke, Maheedhar Kolla, and Olga Vechtomova. An effectiveness measure for ambiguous and underspecified queries. In *Conference on the Theory of Information Retrieval*, pages 188–199. Springer, 2009.

[32] Alistair Moffat and Justin Zobel. Rank-biased precision for measurement of retrieval effectiveness. *ACM Transactions on Information Systems (TOIS)*, 27(1):2, 2008.

Interactive Face Robot

Michel J. Owayjan, Fady C. Elias-Sleiman, Hasan A. Ezzidine, and Nourhane H. Haidar
Department of Computer and Communications Engineering
American University of Science & Technology
Beirut, Lebanon
mowayjan@aust.edu.lb, {fady.e.sleiman, hasanezzidine, nourhanehhaidar}@gmail.com

Abstract- **The interaction between humans and robots has increased and become more personalized in this century. Robots are being used in various fields such as medicine, manufacturing, automation, etc.; however, they lack the interaction capability that is found in a human mainly because more than 60% of human communication is non-verbal, and thus, is hard to replace him. This project presents an interactive face robot which is able to mimic the six universal human facial expressions: happiness, surprise, fear, sadness, anger, and disgust. It can interact with the user using speech recognition in order to mimic a career/major adviser. The Interactive Face Robot is made out of a rubber mask to simulate the skin. Facial expressions are created by moving muscle-like structures based on a facial mass-spring model. Nine servo motors move the different action units in the face to create the different expressions. The Interactive Face Robot mimics a career/major adviser using speech recognition software, the BitVoicer, which recognizes the user's speech and communicates with a LabVIEW™ program that evaluates the answers and calculates the best career/major choice. The LabVIEW™ program, which runs on a laptop computer, also converts the pre-defined questions that the robot asks into voice, and communicates with an Arduino Mega the expression that the face is to generate. Using pulse-width modulation, the Arduino Mega controls the needed servo motors to create the necessary facial expression. The Interactive Face Robot has proven to be user-friendly, and interacted with humans as pre-programmed. Facial expressions provide robots with more capabilities for human-like interactions, making them more adequate for applications requiring interactions with humans. This Interactive Face Robot can replace a career/major adviser, and can be programmed as a receptionist, front-desk operator, or can be used in therapeutic applications.**

Keywords—HCI; Facial Expressions; Interactive Robot

I. Introduction

As the use of service robots continues to grow rapidly, an increase in the human-robot interaction (HRI) is required [1]. More than 60% of the human interaction is inducted nonverbally [2]. Thus a human-friendly interface should be developed to communicate information the same way we do with humans. For a robot to be effective in the human world, it must respond to the human emotional state by generating back humanlike expressions.

This Interactive Face Robot is able to mimic five universal expressions that benefit users in traditional roles such as servants, assistants, or companions. In a typical work day, an employee might face certain personal issues or might be exhausted; a state that is reflected on the individual's facial expressions. Since facial expressions play a major role in communication, this might send the wrong signal to the other person. Interactive robots have been proved to play a major role in improving the users' relaxation and motivation levels by reducing stress [2]. Moreover, it was found very efficient in helping children with autism by engaging them in social

interaction interface [1]. The proposed robot face is a user friendly system. However, its true worth is in the therapeutic characteristics.

The significance of this project is to infer a broad scope of uses in the field of HRI. The system is a fully automated real-time one. The objective is to deliver a complete system that generates facial expressions, recognizes speech, and generates speech in real time. The proposed technique is made up of three primary key points: artificial expressions generation, speech generation, and speech recognition. The proposed scheme can be applied in different industries and circumstances and for different purposes. The goal proposed in this paper is to make the face robot an administrator at a university. Students usually have a hard time deciding what major they should go into. Therefore, the face robot will undertake a career test with the student to find out what the student likes and what not, and accordingly tells the candidate the major that most fit him/her.

II. State-of-the-Art

A Japanese teacher called Hiroshi Ishiguro developed his own robotic doppelgänger. The doppelgänger can act as the teacher by explaining lessons and answering questions. [3] Also, Hisashi Ishihara, Yuichiro Yoshikawa, and Prof. Minoru Asada of Osaka University in Japan developed a child robot that can generate realistic facial expressions. [4]

Within the recent decades, many researches have been trying to compose strategies to automatically classify facial expressions. Various methods have been used. The representative research-work based on this approach includes the work done by Hanson Robotics [5]. The newest addition to their family was Han, a face robot that can engage in conversation and generate facial expressions and make eye contact [5]. On the other hand, the Einstein robot head used 31 servo motors, 27 of them are used to generate facial expressions [5]. What remarks this project from others is that the facial expressions are detected by the movement of the muscles under the skin; these are learned by the robot and then generated when necessary. The project describes how each servo motor is used with every particular expression [5].

Moreover, Karsten and Jochens constructed a robotic head ROMAN. The paper delineates the three major steps in this project: first how the robot is able to realize facial expressions, then the mechatronic system of this robot including the neck and eye construction, and finally the software architecture of ROMAN [6].

III. MATERIALS AND METHODS

The project's design can best be described in three parts: hardware, model construction, and software. Figure 1 shows the block diagram of the project, detailing the interaction between the hardware and the software parts which will be described in more details hereafter.

A. Hardware

Every step that went into the construction of the hardware is presented in Figure 2. To begin with, the face was divided into three parts: the upper, middle, and lower. The upper part contains the eyebrows and their corresponding mechanism. The middle part contains the eyes, the upper lip, and the fixed upper jaw. Finally, the lower part contains the lower lip and moving lower jaw, and rotates to model the mouth as it opens and closes.

The eyes are ping pong balls that have been reshaped and colored to look like the human eyes. The position of the eyes required several calculations. The human eyes are positioned at the same distance from the right and the left nostrils, and to the line perpendicular to the mouth. Moreover, when looking from the top plane of the face, the distance in which the eyes are pushing forward or backwards should be taken into consideration. After fixing the eyes, the back of every eye is attached with an L shaped plastic, as shown in Figure 3, that was fabricated using a 3D printer. The end of the L shape is connected to a joint, as shown in Figure 4, attached to the servo motors arm. This allows the eyes to move right and left simultaneously. To apply a real lips movement, a guitar string is fixed around the lower and upper teeth. Other guitar strings are fixed at two points on the lower jaw and three points on the upper jaw with a round string connecting each part to model each lip. Every fixed point is tightened by a servo motor.

Figure 2. Making of the hardware.
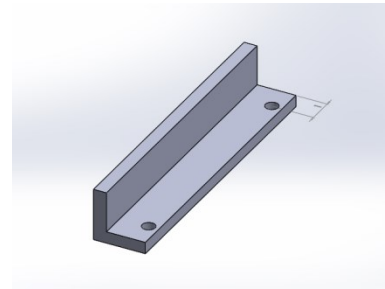
Figure 3. L-shape 3-D CAD design.
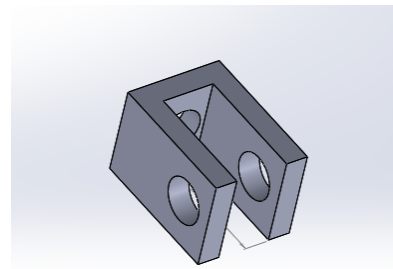
Figure 4. Joint 3-D CAD design.
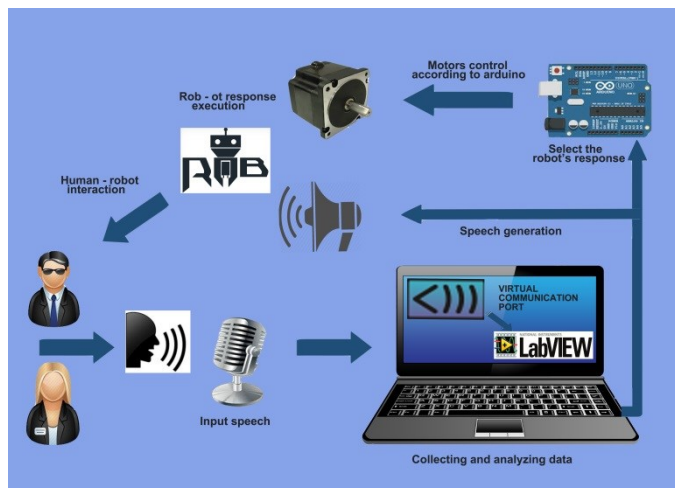
Figure 1. System's block diagram.

Influenced by the idea behind the bicycles' brakes, the guitar strings are navigated through plastic pipes that are fixed on the mannequins head. The end of the pipe is immobile, allowing that end of the steel string to be tightened using servo motors.

The lower jaw needed two servo motors to allow easy movement. Moving to the upper jaw, knowing that there are three fixed points on the round guitar string bridge that modeled the upper lip, three servo motors are used. Due to lack of space, a polyamide body was designed to hold the three motors as seen in Figure 5. The motor wings hold a plastic rod that fixes the guitar strings on it in order to move the lips. In addition to that, a cable conjunction is fixed in front of the polyamide body in order to fix the plastic pipes with a screw. The conjunction itself is fixed by a screw so that the pipe, the rods on the motor end, and the openings above the lips are on the same line of sight.
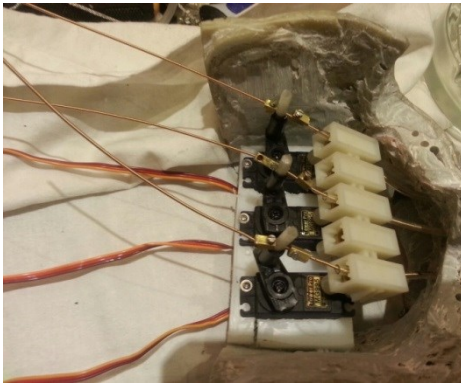
Figure 5. Upper jaw.

The lower jaw and the upper jaw are connected in such a way that allows the mouth to open and close. Therefore, the two parts are fixed using an iron joint as seen in Figure 6.

To allow rotation, first the lower jaw is connected with a loosened screw, and smoothened from its back-right and beck-left to make an angle of approximately 30 degrees with the upper jaw. The movement mechanism is done using two steps. The first is to fix three springs from the front end of the lower jaw to the upper jaw. While the second is to connect the back end of the lower jaw to a servo motor fixed on the back of the head, using a plastic thread. This will pull the jaw and make the required rotation. A spring is attached to pull the jaw back to its initial position.

The next step covers the eyebrows. The eyebrows are fixed 2 cm above the eyes. They are made up of fiber masked by a nylon sheet. The edges are smoothed. They are later drilled and fixed by their middle to be able to rotate easily to perform happy, angry, and neutral emotions. This mechanism is made up of plastic threads that are connected from the center of the eyebrows, up to the upper part of the head, then into the head where they are connected to the servo motors. This will allow the eyebrows to move up to 60 degrees anti-clock wise.

As for the neck, an iron rod is attached on a base platform that is connected to a rotating disk around a fixed axis. The base can be adjusted to the left and right using a servo motor. Such a mechanism will not be affected by the robot's weight.

B. Model Construction

After assembling all the parts together, there comes the binding stage between the rubber face and the skull. This was too risky because of the incompatibility between them. Knowing that facial expressions are produced due to fixed AUs and flexible ones, the mask is secured in some places, and left loose in others. Also, the cheeks are filled with sponge to achieve the smooth touch of the face and give it a realistic form. The nose is also thickened by sponge and compressed to have the desired shape. After fixing all the face, the motion of every motor was tested. You can see the before and after of the Interactive Face Robot in Figure 7.

C. Software

The project's system is based on three software: LabVIEW™ (National Instruments, Austin, TX, USA), Arduino™ (Arduino, Northern Italy), and BitVoicer™ (BitSophia, So Paulo, Brazil) which are connected via a serial communication (Figure 8). The LabVIEW™ program

is called the decision maker since it connects the BitVoicer™ to the Arduino™. The LabVIEW™ program gets its input from the BitVoicer™, and then according to a predefined algorithm, it will send a command to the Arduino™ microcontroller which will drive the corresponding servo motors.
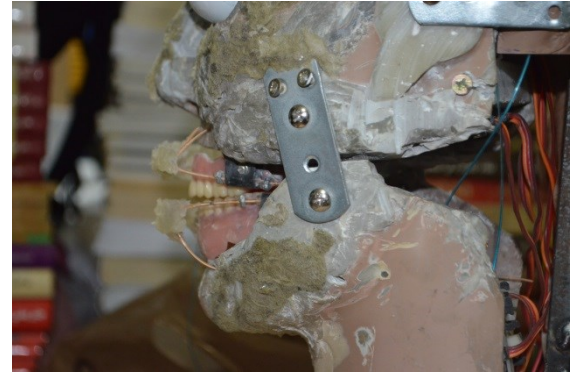

Figure 6. The connection of the jaws.

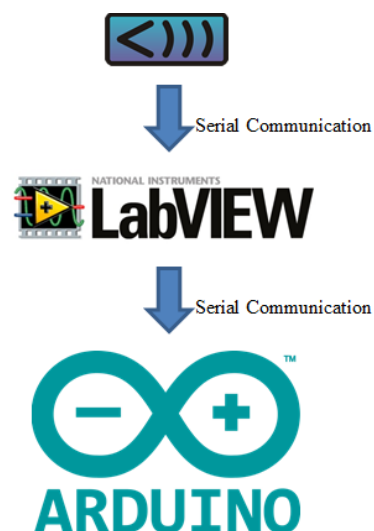Figure 7. Before and after of the interactive face robot.

Figure 8. Software block diagram.

The face robot can be used in a variety of fields. One of them is to act as a university administrator. One of the hardest things that a student encounters when moving from the school phase to the university phase is to choose the right major. That is why, the previously mentioned program was edited to be able to do a career test with new students. A sample career test was performed. Eleven majors were taken into considerations, which are: Radio TV, Interior Design, Journalism, Public Relations, Computer Science and Informatics, Finance, Management, Marketing and Advertising, Computer and Communications Engineering, Biomedical Engineering, and Clinical Laboratory Sciences. The sample career test displayed in Table. I consists of eighteen questions. The users' answers are a simple yes/no answers. Whenever the user answers yes on a certain question, the majors related to the question should increment once. The increments are distributed between the majors equally with no favor of one over the other. Finally, when the test is over, the program should be able to know the major with the highest score and relay the information as speech to the user. The Arduino code controls the motors in the face to produce an emotion. Table II displays the angles that the ten motors rotate to create every emotion. There are one motor for the eyes, one for the neck, one for every eyebrow, one for the jaw, three for the upper lip, and two for the lower lip.

TABLE I
SAMPLE CAREER TEST

	Question	Incremented Majors
1	Do you like mathematics?	Computer and Communications Engineering Computer Sciences and Informatics Biomedical Engineering Finance Interior Design
2	Are your communication skills good and do you like to meet new people?	Journalism Marketing and Advertising Public Relations
3	Do you find out-of-the-box solutions for your problem?	Computer and Communications Engineering Computer Sciences and Informatics Marketing and Advertising
4	Are you confident and do you like to show yourself to the public?	Journalism Radio TV Public Relations
5	Do you like to learn about living things?	Biomedical Engineering Clinical Laboratory Scientist
6	Are you good in analyzing and interpreting data?	Finance Management
7	Do you have an analytical mind and do you organize your responsibilities?	Finance Management
8	Are you interested in acting, film making, or media production?	Journalism Radio TV
9	Did you spend most of your high school time in laboratories?	Biomedical Engineering Clinical Laboratory Scientist
10	Do you have the ability to recognize general principles in certain situations?	Clinical Laboratory Scientist
11	Do you need the freedom to create and be creative?	Interior Design
12	Do you like designing and artwork?	Interior Design
13	Are you appealing?	Public Relations
14	Do you like tinkering with electronics and robots?	Computer and Communications Engineering
15	Do you like to sit in front of your laptop?	Computer Sciences and Informatics
16	Do you interact with people in patience and congeniality?	Public Relations
17	Do you have the ability to link things together?	Management
18	Do you like to do heavy researches?	Marketing and Advertising

TABLE II
EMOTIONS VS ACTION UNITS

Emotion	FACS Code	
	Muscle Description	Associated AUs
Normal	Poker face	Neck : 50° Left eyebrow: 120° Right eyebrow: 30° Eyes: 180° Jaw: 130° Left lower lip: 150° Right lower lip: 90° Left upper lip: 170° Middle upper lip: 170° Right upper lip: 170°
Surprised	Eyebrows raised Mouth open Eyes open Lips protruded	Neck : 50° Left eyebrow: 90° Right eyebrow: 0° Eyes: 180° Jaw: 50° Left lower lip: 120° Right lower lip: 130° Left upper lip: 170° Middle upper lip: 170° Right upper lip: 170°
Angry	Nostrils raised Mouth compressed Furrowed brows Eyes wide open	Neck : 50° Left eyebrow: 140° Right eyebrow: 40° Eyes: 180° Jaw: 130° Left lower lip: 150° Right lower lip: 90° Left upper lip: 170° Middle upper lip: 170° Right upper lip: 170°
Frustrated/Contempt	Lip protrusion Upper lip raised Eyebrows raised	Neck : 50° Left eyebrow: 90° Right eyebrow: 0° Eyes: 180° Jaw: 130° Left lower lip: 150° Right lower lip: 90° Left upper lip: 170° Middle upper lip: 170° Right upper lip: 170°
Sad	Mouth depressed Eyebrows partially raised	Neck : 50° Left eyebrow: 115° Right eyebrow: 10° Eyes: 180° Jaw: 130° Left lower lip: 120° Right lower lip: 90° Left upper lip: 170° Middle upper lip: 170° Right upper lip: 170°
Happy	Upper lip raised at corners Eyebrows partially raised Jaw partially opens	Neck : 50° Left eyebrow: 100° Right eyebrow: 0° Eyes: 180° Jaw: 100° Left lower lip: 150° Right lower lip: 90° Left upper lip: 100° Middle upper lip: 170° Right upper lip: 100°
Move left	Turn neck to the left	Neck: 130°
Move right	Turn neck to the right	Neck: 30°
Move eyes	Move eyes left then right	Eyes: from 0° to 180° with delay 50ms for every 5°
Talk	Open and close mouth	Neck : 50° Left eyebrow: 120° Right eyebrow: 30° Eyes: 180° Jaw: 130° Left lower lip: 150° Right lower lip: 90° Left upper lip: 170° Middle upper lip: 170° Right upper lip: 170° Delay (200ms) Neck : 50° Left eyebrow: 120° Right eyebrow: 30° Eyes: 180° Jaw: 90° Left lower lip: 120° Right lower lip: 120° Left upper lip: 150° Middle upper lip: 150° Right upper lip: 150°

IV. TESTING AND RESULTS

This project presented results for a fully automated real-time expression generation system that has been proven to be able to generate the five universal expressions: normal, happiness, sadness, surprise, and anger (Figure 9).

The primary goal of this work is to solve an engineering problem: how to make the robot's interaction be as humanly as possible. Facial expressions play a major role in this interaction. Therefore, the appearance of human facial muscles was approximated with available motors. The career test was experimented on several graduate friends, and almost all of them got the result of their university major or something that they have always wanted to do, but never got the chance to. Moreover, they enjoyed taking the test since the robot was interacting with them non-verbally. The robot was smiling and making funny expressions so that the user felt comfortable and happy. It was observed that users interacted more with the robot than with either a laptop or with a human.

V. CONCLUSIONS AND FUTURE WORK

Because of the demand to study human robot interaction, a robot platform was produced to meet this requirement. The exact task of this project is to have a robot face that is human-like and produces facial expressions similar to those in humans. The methodology followed is simple. The robot takes as an input the users' speech. Then, the robot interacts while generating facial expressions by feeding the Arduino with specific command that will thus move the servo motors to generate the desired expression by stretching and releasing the rubber face on the AUs that were generated in the testing phase. Moreover, in order to make this interaction more realistic, speech was added for the robot to talk back with the user. The main task of the robot is as an administrator at a university. It asks new students questions to conduct a career test. According to the students' responses, the robot will tell him/her what major he/she should get into.

The scope of this project is promising. It is accurate to say that such a scheme can be used for different purposes easily. For example, the project can be used in the therapeutic domain to help people with autism; also, it can be useful in the education application or even in the care of the aged people. For this project to reach its foremost goal, we should allow it to interpret the facial expressions of the human facing it and act accordingly. Also, artificial intelligence should play a role so that the robot will know how to act without predefined scenarios. Moreover, the speech domain should become broad enough to make sure that the robot can reply to most of what he is receiving.
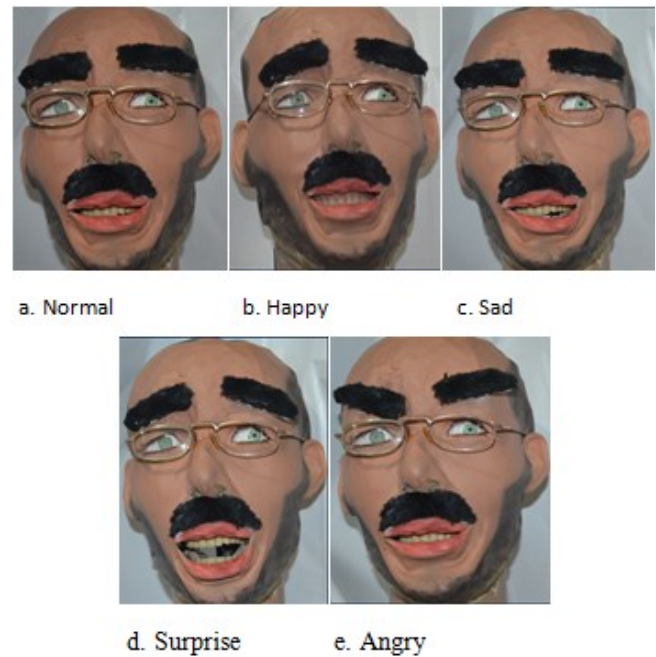
Figure 9. Results.

REFERENCES

[1] M. A. Goodrich1 and A. C. Schultz, "Human–Robot Interaction: A Survey," *Foundations and TrendsR in Human–Computer Interaction,* vol. 1, no. 3, pp. 1-20, 2007.

[2] D. Feil-Seifer and M. J. Matarić, "Human–Robot Interaction," *Encyclopedia of Complexity and Systems Science,* R. A. Meyers (eds.), Springer New York, 4643-4659, 2009.

[3] E. Guizzo, "Hiroshi Ishiguro: The Man Who Made a Copy of Himself," April 2010. [Online]. Available: http://spectrum.ieee.org/robotics/humanoids/hiroshi-ishiguro-the-man-who-made-a-copy-of-himself.

[4] N. Kageki, "Meet Affetto, a Child Robot With Realistic Facial Expressions," Feb 2011. [Online]. Available: http://spectrum.ieee.org/automaton/robotics/humanoids/meet-affetto.

[5] Hanson Robotics. "Face Robot," 2015. [Online]. Available: http://www.hansonrobotics.com/robot/han/

[6] K. Berns and J. Herth, "Control of facial expressions of the humanoid robot head ROMAN," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3119-3124, 2006, October 9-15, Beijing, China.

Power Efficient Gurumukhi Unicode Reader Design and Implementation on FPGA

Amanpreet Kaur, Sunny Singh
Chitkara Institute of Engineering &
Technology, Chitkara University,
Chandigarh,Punjab
amanpreet.kaur@chitkara.edu.in
sunny.singh@chitkara.edu.in

Bishwajeet Pandey, Nisha Pandey
Gyancity Research Lab, India
Gurgaon, India
gyancity@gyancity.com,
nisha@gyancity.com

D M Akbar Hussain
Department of Energy Technology
Aalborg Univeristy
Esbjerg, Denmark
akh@et.aau.dk

Abstract—**Gurumukhi is found to be the most widely used language of Pakistan, and it is ranked 3rd in Canada, 7th in India and almost 4th most spoken language in U.K. This Unicode Reader is cost effective solution for learning as well as understanding the Punjabi language by the people across the globe .This reader helps the user to understand, whether written text is consonants, vowels or digits of Gurumukhi scripts. This paper can also be the solution to the various problems occurred in research of Punjabi natural language processing. Hardware is designed for Gurumukhi Unicode Reader (GUR) and is implemented on Virtex-6 FPGA on Xilinx software. This GUR design is tested on different frequencies by applying frequency scaling techniques .The reader is also observed on different IO Standards of two logic families i.e. on SSTL (Stub-Series Terminated Logic) and LVDCI (Low Voltage Digitally Controlled Impedance) logic families to make this design more energy efficient. It is concluded that using LVDCI_DV2_15 rather than SSTL18_II_DCI, the total power can be saved up to 51.22% with the device operating at a frequency of 1MHz.**

Keywords— Energy Efficient, Gurmukhi Scripts, Unicode Reader, Punjabi Language Reader, Natural Language Processing

I. INTRODUCTION

Even in the era of computer and telecommunication, constant development efforts are still being made in order to make the communication faster, easy, economical and less burdensome. But we are still away from the dream in which the ideas and the information can be freely exchanged despite of the language and script in which the data is written/spoken. Unicode is a one step ahead in making this dream true. Unicode is a universal character encoding standard. Unicode is designed in order to support characters from all languages around the world, unlike ASCII codes which are designed only to describe few basic English characters. So in order to make communication in Gurumukhi more easy and compatible, Power Efficient Unicode Reader has been designed in the following paper that consumes less power and less energy in order the meet the energy crisis being faced by the world .For the implementation of Gurmukhi Unicode reader, many other languages Unicoder Readers were also being [1-5]. In order to design a power efficient Gurumukhi, the power utilities of SSTL (Stub-Series Terminated Logic)

[6] & LVDCI (Low Voltage Digitally Controlled Impedance) [7] logic families are compared at different IO (Input-Output) standards [8-10]. The power utility of the device can also check at other logic families such as HSTL, LVCMOS, Mobile DDR [11-13]. In conjunction with the change in the IO standards, the frequency of the FPGA device has also been scaled from lowest frequency of 1MHz to the highest frequency of 1 THz in order to see the effect of the frequency on the total power consumption of the device, this process of scaling is called as frequency scaling [14]. The frequency can be varied to more than 1THz and to less than 1MHz.In this way by applying the frequency scaling technique and changing the IO standards the least power consumed IO standard and frequency has been find out , thus making the device power efficient by operating it on such a particular frequency and IO standard. Voltage scaling [15], Capacitance scaling [16] and many other powers efficient techniques can also be applied to the devices in order to make devices more power efficient.
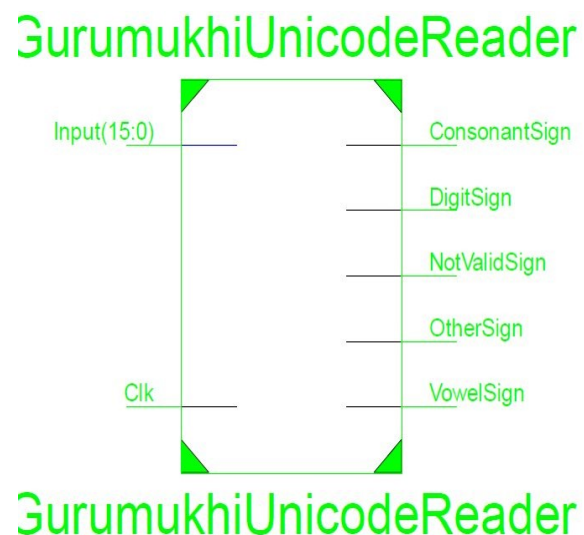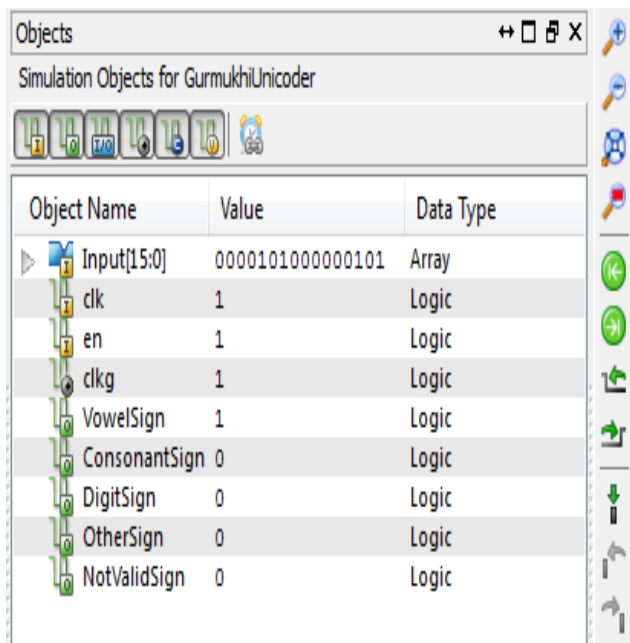
II. SCHEMATIC OF GURMUKHI UNICODE READER

Fig. 1: RTL Schematic of Gurmukhi Unicode Reader

RTL (Register Transfer Level) schematic of the Gurumukhi Unicode reader is shown in the fig.1. Two inputs have been applied to the reader one is the 16 bit hexadecimal code which is unique for each and every alphabet or character of Gurumukhi. Another is the clock input in order to make it a synchronous circuit. Five corresponding outputs are coming out of the circuit. The hexadecimal range lies from 0x0A01-0x0A70 is solely dedicated to the characters of the Gurumukhi. If any range comes out to be less or more than this dedicated range, then not valid sign will be high, else if the hexadecimal input range confines within 0x0A05-0x0A14, then the output of the vowel sign will be high as shown in the figure 2. The input applied to the Gurmukhi Unicode Reader is 0x0A10 that lies within the ranges of Vowels of Gurmukhi language; hence the output is 1 corresponding to the vowel sign. Similarly if the input range lies within 0x0A15-0x0A39, then the output at the consonant sign will be high, if the input range lies within 0x0A66-0x0A6F, then the output at the digit sign will be high. If any other hexadecimal number is applied as an input within the range of 0x0A01-0x0A70 except the vowel, consonant and digit, then the output at other sign will be high.

Fig. 2: Waveform of Gurmukhi Unicode Reader corresponding to the Vowel Sign

III RESULT ANALYSIS

The total amount of power dissipated by the device when operated at 6 different IO standards of SSTL and LVDCI logic family and 6 differe frequency ranges as depicted in the result analysis.

(i)Total power dissipated when the device operates at LVDCI_15, SSTL2_II IO standards and when its frequency is also scaled from 1 MHz to 1THz as shown in table 1.

Table 1: Power Analysis

Frequencies	Total Power Dissipated (LVDCI_15)	Total Power Dissipated (SSTL2_II)
I THz	32.325	39.921
100GHz	3.927	4.980
10GHz	1.040	1.417
1GHz	0.750	1.061
100MHz	0.721	1.024
10MHz	0.718	1.021
1MHz	0.718	1.020

It has been observed that out of all the frequencies the frequency at which the amount of dissipation of power minimum is a frequency of 1MHz whereas the maximum power is dissipated at highest 1THz frequency. It has also been observed that out of LVDCI_15 and SSTL2_II IO standards , SSTL2_II is the one at which the dissipation of power is maximum as compared to LVDCI_15.Hence it is most feasible to work with LVDCI logic family in comparison to SSTL. Mathematically there is 97.7% of power is reduced on operating at 1 MHz frequency rather than 1THz in case of LVDCI_15 and there is 97.4% reduction in power of SSTL2_II. Hence if we compare both LVDCI_15 and SSTL2_II families, then it can be concluded that on operating the device at a frequency of 1MHz in case of LVDCI_15, we can save total 29.6% of dissipated power rather than in case of SSTL2_II as shown in figure3.
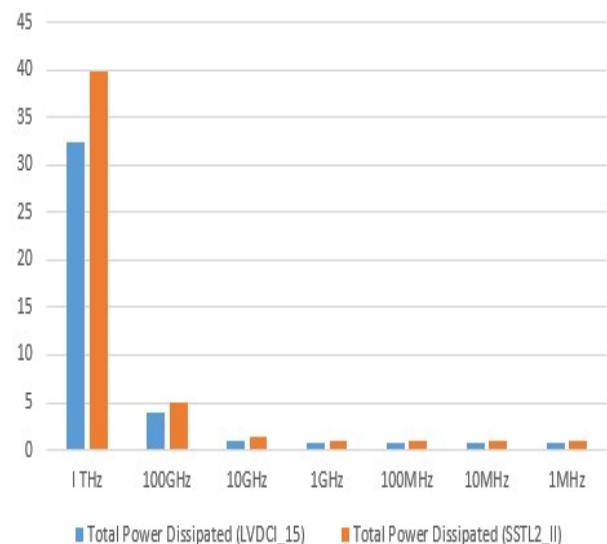

Fig. 3: Comparison of Power Dissipation with LVDCI_15 and SSTL2_II

(ii)Total power dissipated when the device operates at LVDCI_18, SSTL18_II IO standards and when its frequency is also scaled from 1 MHz to 1THz as shown in table 2.

Table 2: Power Analysis

Frequencies	Total Power Dissipated (LVDCI_18)	Total Power Dissipated (SSTL18_II)
I THz	36.678	34.621
100GHz	4.378	4.426
10GHz	1.089	1.351
1GHz	0.760	1.044
100MHz	0.726	1.012
10MHz	0.723	1.009
1MHz	0.723	1.009

Table 3: Power Analysis

Frequencies	Total Power Dissipated (LVDCI_25)	Total Power Dissipated (SSTL15)
I THz	46.967	33.384
100GHz	5.450	4.827
10GHz	1.210	1.326
1GHz	0.786	1.030
100MHz	0.743	0.999
10MHz	0.739	0.996
1MHz	0.738	0.996

It has been observed that out of all the frequencies the frequency at which the amount of dissipation of power minimum is a frequency of 1MHz whereas the maximum power is dissipated at highest 1THz frequency. . It has also been observed that out of LVDCI_18 and SSTL18_II IO standards , SSTL18_II is the one at which the dissipation of power is maximum as compared to LVDCI_18.Hence it is most feasible to work with LVDCI logic family in comparison to SSTL. Mathematically there is 98.0% of power is reduced on operating at 1 MHz frequency rather than 1THz in case of LVDCI_18 and there is 97.08% reduction in power of SSTL18_II. Hence if we compare both LVDCI_18 and SSTL18_II families, then it can be concluded that on operating the device at a frequency of 1MHz in case of LVDCI_18, we can save total 28.34% of dissipated power rather than in case of SSTL18_II as shown in figure 4.

It has been observed that out of all the frequencies the frequency at which the amount of dissipation of power minimum is a frequency of 1MHz whereas the maximum power is dissipated at highest 1THz frequency. . It has also been observed that out of LVDCI_25 and SSTL15 IO standards , SSTL15 is the one at which the dissipation of power is maximum as compared to LVDCI_25.Hence it is most feasible to work with LVDCI logic family in comparison to SSTL. Mathematically there is 98.42% of power is reduced on operating at 1 MHz frequency rather than 1THz in case of LVDCI_25 and there is 97.01% reduction in power of SSTL15. Hence if we compare both LVDCI_25 and SSTL15 families, then it can be concluded that on operating the device at a frequency of 1MHz in case of LVDCI_25, we can save total 25.90% of dissipated power rather than in case of SSTL15 as shown in figure 5.

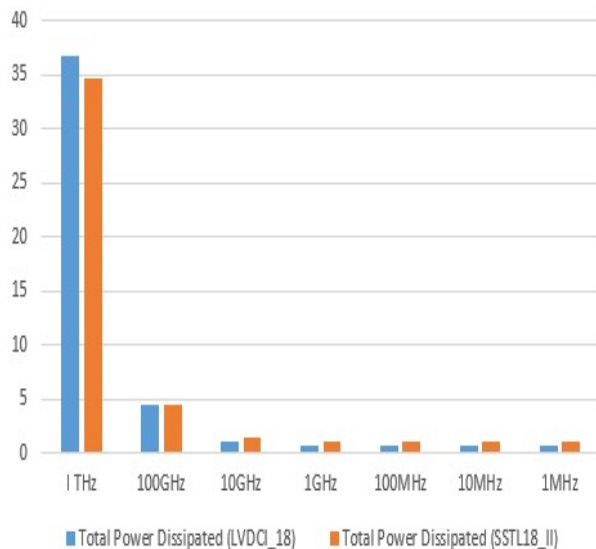Fig. 4: Comparison of Power Dissipation between LVDCI_18 and SSTL18_II

Fig. 5: Comparison of Power Dissipation between LVDCI_25 and SSTL15

(iii)Total power dissipated when the device operates at LVDCI_25, SSTL15 IO standards and when its frequency is also scaled from 1 MHz to 1THz as shown in table 3.

(iv)Total power dissipated when the device operates at LVDCI_DV2_15, SSTL2_II_DCI IO standards and when its

frequency is also scaled from 1 MHz to 1THz as shown in table 4.

Table 4: Power Analysis

Frequencies	Total Power Dissipated (LVDCI_DV2_15)	Total Power Dissipated (SSTL2_II_DCI)
I THz	32.321	32.419
100GHz	3.923	5.075
10GHz	1.036	2.289
1GHz	0.747	2.011
100MHz	0.717	1.982
10MHz	0.714	1.979
1MHz	0.714	1.979

It has been observed that out of all the frequencies the frequency at which the amount of dissipation of power minimum is a frequency of 1MHz whereas the maximum power is dissipated at highest 1THz frequency. . It has also been observed that out of LVDCI_DV2_15 and SSTL2_II_DCI IO standards , SSTL2_II_DCI is the one at which the dissipation of power is maximum as compared to LVDCI_DV2_15.Hence it is most feasible to work with LVDCI logic family in comparison to SSTL. Mathematically there is 98.0% of power is reduced on operating at 1 MHz frequency rather than 1THz in case of LVDCI_DV2_15 and there is 97.08% reduction in power of SSTL2_II_DCI. Hence if we compare both LVDCI_DV2_15 and SSTL2_II_DCI families, then it can be concluded that on operating the device at a frequency of 1MHz in case of LVDCI_DV2_15, we can save total 28.34% of dissipated power rather than in case of SSTL2_II_DCI as shown in figure 6.

Fig. 6: Comparison of Power Dissipation between LVDCI_DV2_15 and SSTL2_II_DCI

(v)Total power dissipated when the device operates at LVDCI_DV2_18, SSTL18_II_DCI IO standards and when its

frequency is also scaled from 1 MHz to 1THz as shown in table 5.

Table 5: Power Analysis

Frequencies	Total Power Dissipated (LVDCI_DV2_18)	Total Power Dissipated (SSTL18_II_DCI)
I THz	36.673	32.709
100GHz	4.373	4.645
10GHz	1.083	1.788
1GHz	0.754	1.502
100MHz	0.720	1.473
10MHz	0.717	1.470
1MHz	0.717	1.470

It has been observed that out of all the frequencies the frequency at which the amount of dissipation of power minimum is a frequency of 1MHz whereas the maximum power is dissipated at highest 1THz frequency. . It has also been observed that out of LVDCI_DV2_18 and SSTL18_II_DCI IO standards , SSTL18_II_DCI is the one at which the dissipation of power is maximum as compared to LVDCI_DV2_18.Hence it is most feasible to work with LVDCI logic family in comparison to SSTL. Mathematically there is 98.04% of power is reduced on operating at 1 MHz frequency rather than 1THz in case of LVDCI_DV2_18 and there is 95.50% reduction in power of SSTL18_II_DCI. Hence if we compare both LVDCI_DV2_18 and SSTL18_II_DCI families, then it can be concluded that on operating the device at a frequency of 1MHz in case of LVDCI_DV2_18, we can save total 51.22% of dissipated power rather than in case of SSTL18_II_DCI as shown in figure 7.
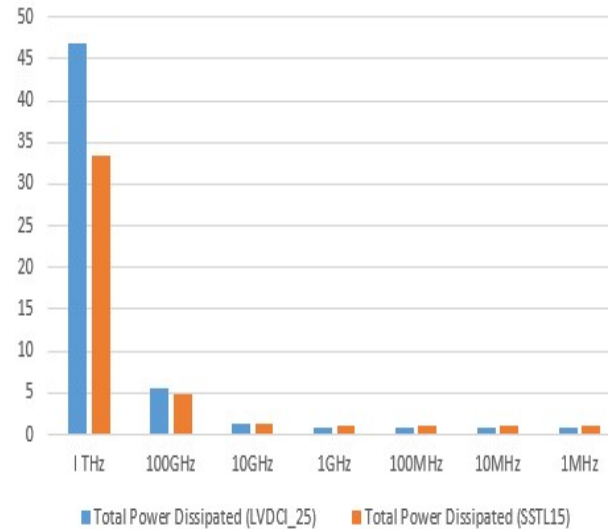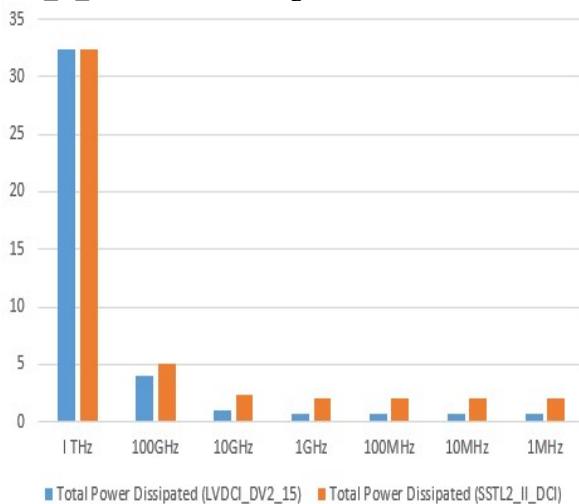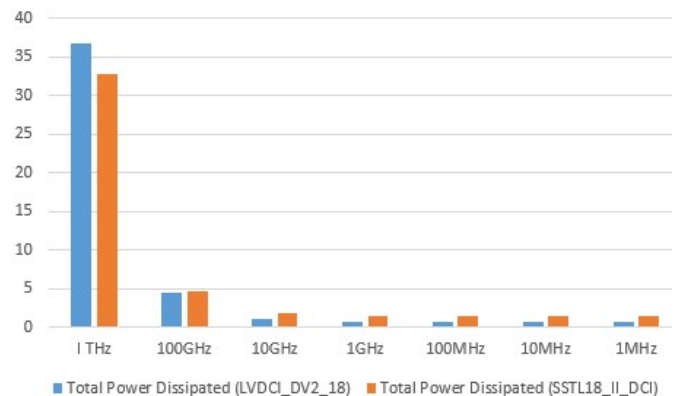

Fig. 7: Comparison of Power Dissipation between LVDCI_DV2_18 and SSTL18_II_DCI

(v)Total power dissipated when the device operates at LVDCI_DV2_25, SSTL15_II_DCI IO standards and when its frequency is also scaled from 1 MHz to 1THz as shown in table 6.

Table 6: Power Analysis

Frequencies	Total Power Dissipated (LVDCI_DV2_25)	Total Power Dissipated (SSTL15_DCI)
I THz	57.265	33.323
100GHz	6.497	4.499
10GHz	1.302	1.565
1GHz	0.783	1.271
100MHz	0.731	1.241
10MHz	0.725	1.238
1MHz	0.717	1.238

It has been observed that out of all the frequencies the frequency at which the amount of dissipation of power minimum is a frequency of 1MHz whereas the maximum power is dissipated at highest 1THz frequency. . It has also been observed that out of LVDCI_DV2_25 and SSTL15_DCI IO standards , SSTL15_DCI is the one at which the dissipation of power is maximum as compared to LVDCI_DV2_25.Hence it is most feasible to work with LVDCI logic family in comparison to SSTL. Mathematically there is 98.74% of power is reduced on operating at 1 MHz frequency rather than 1THz in case of LVDCI_DV2_25 and there is 96.28% reduction in power of SSTL15_DCI. Hence if we compare both LVDCI_DV2_25 and SSTL15_DCI families, then it can be concluded that on operating the device at a frequency of 1MHz in case of LVDCI_DV2_25, we can save total 42%of dissipated power rather than in case of SSTL15_DCI as shown in figure 8.
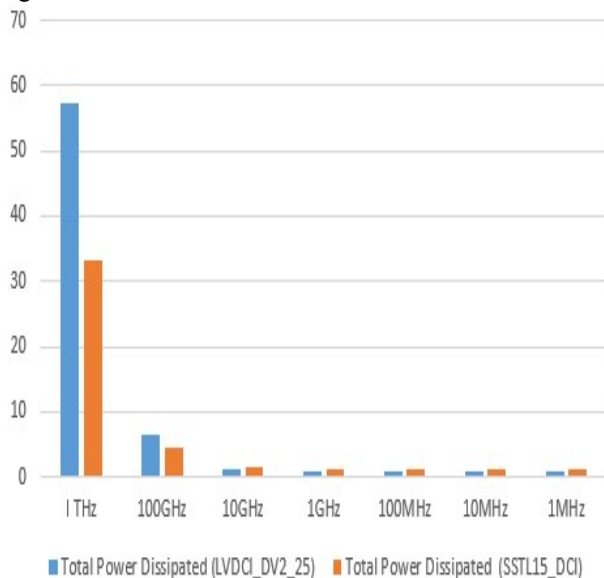

Fig. 8: Comparison of Power Dissipation between LVDCI_DV2_25 and SSTL15_DCIs

IV. CONCLUSION

It has been concluded that out of all the IO standards of SSTL and LVDCI logic families, the maximum amount of power has been saved in case of SSTL18_II_DCI if we want to operate the device on SSTL logic family and maximum power savage is in case of LVDCI_DV2_15 IO standard if we want to operate the device on LVDCI logic family providing that both the logic families are operating on a frequency of 1 MHz. Also if we compare amongst the two logic families, then it has been clearly seen from the table 7 that there is always a more power savage in case of LVDCI logic family in comparison to SSTL logic family.

Table 7: Comparison of Power Savage amongst LVDCI and SSTL logic family

S.NO.	If we move to	Power Saving
1.	LVDCI_DV2_15 rather than SSTL18_II_DCI	51.22%
2.	LVDCI_DV2_25 rather than SSTL15_DCI	42.08%
3.	LVDCI_15 rather than SSTL2_II	29.6%
4.	LVDCI_18 rather than SSTL18_II	28.34%
5.	LVDCI_DV2_15 rather than SSTL2_II_DCI.	28.34%
6.	LVDCI_25 rather than SSTL15	25.90%

It is concluded that the LVDCI_DV2_15 rather than SSTL18_II_DCI, the total power can be saved up to 51.22% with the device operating at a frequency of 1MHz.

V. FUTURE SCOPE

We can also make the Unicoder on other languages such as Shahmukhi, Urdu, Sindhi and many more. Till now we have only worked upon the two logic families. We can also implement the same design on other logic families such as LCMOS, HSTL, LVPECL, RSDS and many more. In the implementation of Gurumukhi Unicode reader Virtex-6 FPGA has been used. There is further scope to implement the same design on some other advanced FPGA technologies such as Kintex-7 and Artix-7 in which the channel length of the gate is 28nm which is less than 40nm of Virtex-6.

REFERENCES

[1] A.Kaur,A.Singh,S.Singh,F.Fazili,M.D.Minve,V.Sharma,"Capacitance Scaling Based Energy Efficient and Tera Hertz Design of Malayalam Unicode Reader on FPGA" , International Journal of u- and e Service, Science and Technology Vol.8, No. 8 (2015), pp.151-158.

[2] A.Kaur, B.Pandey,A.Sharma, K.Sharma,S.Singh, "SSTL IO Standard Based Tera Hertz & Energy Efficient MALAYALAM Unicode Reader Design and Implementation on FPGA", Bilingual International Conference on Information Technology: Yesterday, Today, and Tomorrow at Defence Scientific Information & Documentation Centre (DESIDOC), DRDO, Ministry of Defence, Government of India during 19-21 February, 2015, In Special Edition of "Managing Information Technology", Vol.6, No.48, February, 2015.

[3] P.Sharma, A. Kaur, S.Gupta, "Energy Efficient Meetei Mayek Unicode Reader Design ", IEEE International Conference on "Computing for Sustainable Global Development (INDIACOM), Bharati Vidyapeeth, Delhi, 1-3 March 2017.

[4] N.Sharma,A.Kaur,B.Verma, "Stub-Series Terminated Logic Based Energy Efficient Devnagri Unicode Reader Design On 40nm And 28nm FPGA" ,IEEE International Conference on Next Generation Computing Technologies (NGCT), 4-5 September 2015, UPES, Dehradun, India.

[5] B.Pandey, A.Uddin, S.Banshal, M.A.Rahman, T.Das, T.Kumar., "Thermal Aware Energy Efficient Bengali Unicode Reader in Text

Analysis", IEEE International Conference on Reliability Optimization & Information Technology (ICROIT), Faridabad, 6-8 Feb, 2014.

[6] B.Pandey, V.Thind, S. Kaur Sandhu, T. Walia, and SSharma. "SSTL Based Power Efficient Implementation of DES Security Algorithm on 28nm FPGA." International Journal of Security and Its Application 9, no. 7, July 2015, Page: 267-274.

[7] A. Kaur, D. Bhat, S. Singh, "Low Voltage Digitally Control Impedance Based Solar Charge Controller Design on FPGA", IEEE International Conference on "Computing for Sustainable Global Development (INDIACOM), Bharati Vidyapeeth, Delhi, 11-13 March 2015.

[8] S.Madhok,A. Kaur, B. Pandey, "Different IO Standard Based Energy Efficient Decoder Design For 64-bit Processor Architecture", IEEE International Conference on "Computing for Sustainable Global Development (INDIACOM), Bharati Vidyapeeth, Delhi, 11-13 March 2015.

[9] S. Mandhok, B. Pandey, A. Kaur, M. H. Minver, D. M. Akbar Hussain "HSTL IO Standard Based Energy Efficient Multiplier Design using Nikhilam Navatashcaramam Dashatah on 28nm FPGA" International Journal of Control and Automation Vol.8, No.8 (2015), pp.35-44.

[10] H. Kaur, A. Kaur, H. Sohal,I. Gupta, S. Singh and S. Nagpal , "Design and performance analysis of RAM_WR_Control module using Xilinx ISE 14.2",International Conference on Recent Trends in Computer Science and Electronics Engineering(RTCSE),Kuala Lumpur,Malaysia,02-03 Jan.

[11] A.Singla, A. Kaur, and B. Pandey, "LVCMOS Based Energy Efficient Solar Charge Sensor Design on FPGA", IEEE Sixth India International Conference on Power Electronics (IICPE), NIT Kurukshetra, 8-10 December, 2014.

[12] T. Kumar, B. Pandey, T. Das, and B. S. Chowdhry, "Mobile DDR IO Standard Based High Performance Energy Efficient Portable ALU Design on FPGA", Wireless Personal Communications, An International Journal, Volume 76, Issue 3 (2014), pp. 569-578.

[13] S.Madhok , B. Pandey and A. Kaur , Mohamed Hashim Minver and D M Akbar Hussain, "HSTL IO Standard Based Energy Efficient Multiplier Design using Nikhilam Navatashcaramam Dashatah on 28nm FPGA", International Journal of Control and Automation Vol.8, No.8 (2015), pp.35-44.

[14] S. Bhatt, A. Kaur and S. Gupta, "Frequency Scaling Based Power Efficient Kannada Unicode Reader Design on 40nm and 28nm FPGA", IEEE International Conference on "Computing for Sustainable Global Development (INDIACOM), Bharati Vidyapeeth, Delhi, 1-3 March 2017.

[15] T. Gupta, G.Verma, A. Kaur, B. Pandey, A. Singh and T. Kaur, "Energy Efficient Counter Design Using Voltage Scaling On FPGA", IEEE International Conference on Communication Systems and Network Technologies (*CSNT*), Gwalior, India, April 4-6, 2015.

[16] A.Kaur, G.Singh, B.Pandey and F. Fazili, "Capacitance Scaling Based Green Gurumukhi Unicode Reader Design for Natural Language Processing", IEEE International Conference on "Computing for Sustainable Global Development (INDIACOM), Bharati Vidyapeeth, Delhi, 11-13 March 2015.

An Experimental Platform for QoE Studies of WebRTC-based Multi-Party Video Communication

Doreid Ammar*, Katrien De Moor†, Poul Heegaard†
*Emlyon Business School, Ecully, France
ammar@em-lyon.com
†NTNU, Norwegian University of Science and Technology, Trondheim, Norway
{katrien.demoor|poul.heegaard}@ntnu.no

Abstract—**Web Real-Time Communication (WebRTC) enables easy to use, no installation, in browser applications such as Google hangout and appear.in. To increase the understanding of the Quality of Experience with the use of such applications for multi-party video communication requires an experimental platform where controlled studies can be conducted. This paper presents such a platform and a use case with a simple setup and test to demonstrate the features of this platform.**

Keywords—WebRTC, quality of experience (QoE), performance statistics, testbed

I. Introduction

The open source Web Real-Time Communication (WebRTC) project [1] has, since it was established by Google nearly 6 years ago, been used in numerous applications supporting e.g., videochat and file transfer. WebRTC is a royalty-free and open framework that makes it possible to have high-quality real-time communications between different parties, by using a web browser supporting WebRTC (e.g., Google Chrome, Opera, Mozilla Firefox). The strength and appeal of many WebRTC-based applications from a user perspective lies in their simplicity and the fact that they can be accessed from any device connected to the internet (regardless the type of connection and device used), as long as the browser used supports WebRTC. Moreover, some of these browsers even enable to easily collect application level performance statistics, which provide insight into potential performance issues at different levels. Given the many affordances and possible experiences (e.g., different devices, different types of network access, many parties in one call) and the availability of the WebRTC performance statistics, WebRTC is a very interesting use case from a Quality of Experience (QoE)-point of view. For example, in a multi-party scenario where the conditions are asymmetric, evaluating and predicting QoE is not that straightforward. Even if only one party in a call suffers from reduced performance (has a bad network connection or calls from a device with constrained CPU), the other parties in the call may also suffer a lot in terms of quality impairments and interruptions.

It is currently not fully understood how such technical and non-technical factors play a role in this respect, and what this implies for users and their Quality of Experience (QoE), i.e., their degree of delight or annoyance when using the application [1]. This may largely be due to the fact that evaluating QoE in such complex and diverse settings is not straightforward. Yet, such increased insights are highly valuable as they can guide the optimisation of QoE for WebRTC applications in different scenarios and circumstances. More concretely, more profound insights into the (ii) origins and occurrence of performance issues, (ii) impairments which may hinder smooth communication and a pleasureable experience during a call from the user point of view, and (iii) the ways in which users (may) react to this, are needed.

In this paper, we introduce an experimental platform that can help to gain deeper insights into the above issues. The platform can be used for conducting both controlled laboratory and longitudinal "living lab", empirical studies on QoE of WebRTC-based video communication and for collecting different types of complementary data. It consists of a testbed that (i) enables real-time audiovisual conversations for up to eight parties, (ii) provides real-time logging of performance- and session-related statistics, and (iii) enables the collection of both explicit and implicit user feedback. It also contains a dashboard that enables easy data visualisation and that can be used to perform a number of in-depth data analytical approaches. Next to the introduction of the experimental platform, we present results from a small-scale subjective study in which the testbed was used to conduct a controlled laboratory study on QoE in the context of WebRTC based audiovisual communication. More specifically, we investigated whether and how packet loss influences experienced quality in a two-party call setting.

The paper is organized as follows: in Section II, we briefly point to relevant related work on WebRTC QoE and experimental testbeds for investigating QoE in the context of multi-party video communication. Next, we introduce the experimental platform, its features and different components in more detail. In Section IV, we present a use case (subjective user study, its set-up and obtained results) and demonstrate the affordances and applicability of the testbed. Finally, in Section V we summarize the paper and point to the limitations of the testbed as well as the opportunities for future research.

II. Related Work

An increasing number of studies has looked at QoE and factors bearing an influence on it in the context of multi-party video-conferencing (also called telemeeting QoE). Broadly speaking, three classes of QoE influence factors have been defined in the literature [2]: namely human, system and context influence factors. It goes beyond the scope of this paper to comprehensively classify existing studies along these three

---

[1]https://webrtc.org

categories, therefore we point to [3] for a useful overview and instead, limit the overview here to a number of examples. When it comes to human factors, for instance the work presented in [4] indicated that the extent to which a participant is involved in a conversation matters. Similarly, in [5], participants were found to have a different sensibility to delay depending on how active they took part in a conversation. When it comes to contextual factors, [3] argue that *e.g.*, the task and setting of a conversation (*e.g.*, business vs. leisure) are likely to correspond with different quality thresholds and expectations. A useful example here is [6], which shares results from a series of studies. These results point amongst others to differences in QoE depending on the task and social context in which the conversational tests took place.

The primary focus in most studies published so far however has been on system-level influencing factors (*e.g.*, different types of devices, video quality, network limitations and their implications) or on how to enable investigating to which extent they matter. A testbed for investigating QoE of video-mediated communication in controllable and reproducible lab settings was reported in [7]. This testbed allows modification and monitoring of network and media conditions in real-time. Numerous studies has been conducted using the later testbed. *Schmitt et al.* investigated in [5] the impact of asymmetric delay on QoE in video-mediated group discussions. The results indicated a negative impact of asymmetric delay on QoE. In [8], the relationship between QoE and different layout, video quality (low vs. high quality) and network limitations (packet loss) was studied. The results showed a slightly larger impact of packet loss distortions on users' QoE than a reduction of the video quality. In [9], the impact of different encoding bitrate was investigated. The authors claim that 1 Mb/s is sufficient as encoding bitrate for HD streams in video conferencing.

Some previous studies have also focused more specifically on QoE in the context of WebRTC-based video communication. In [10], one of the few studies comparing different types of devices in the context of mobile multi-party telemeetings via WebRTC, the authors pointed to the importance of device capabilities and processing power as a key factors, especially in the context of WebRTC. In [11], *Ammar et al.* investigated the impact of performance-related potential Video QoE killers for two-party video conferencing. The results showed that the performance indicators that could be used to identify severe video freezes are significant drops in data rate, non-zero packet loss ratios, non-zero Picture Loss Indication, and non-zero bucket delay.

III. EXPERIMENTAL PLATFORM

In this section, we describe the different main components of our experimental platform in more detail. More specifically, it consists of a testbed that (i) enables real-time audiovisual conversations in constrained conditions (e.g., by adding network impairments) (ii) provides real-time logging of objective and subjective data and (iii) enables the easy capturing of explicit and implicit user feedback. Additionally, it encompasses a dashboard that allows to visualise the different data sources in a comprehensive way and to perform a number of data analytical procedures.

A. WebRTC Testbed

1) Testbed Topology: Our WebRTC testbed consists of a WebRTC-based application called appear.in[2], a *network emulator* and a *controller*. The testbed enables video conferencing communication for up to eight clients and is designed to run in a controlled environment. Fig. 1 illustrates a simplified version of our testbed with a 3-party setup.
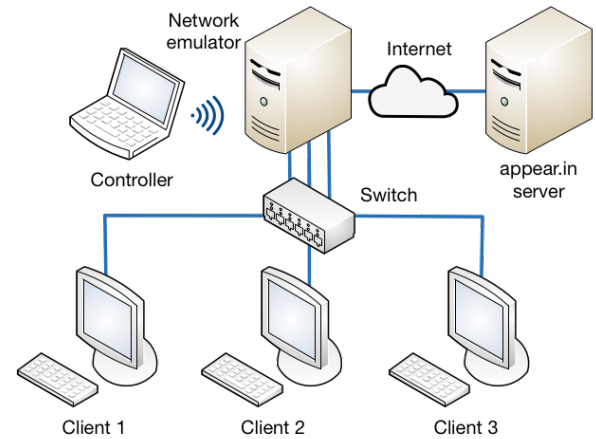

Figure 1: Testbed topology of three-party video conferencing

In the testbed, we use a research version of appear.in, which is nearly identical to the public version of appear.in. The main differences are situated at the level of (i) the feedback form that is displayed when a caller leaves a session (a more fine-grained feedback form is used and this feedback form can be easily modified in the research version), and (ii) the link used to access appear.in (which is slightly different). Apart from these minor differences, the research version of the application can be used by users in the same way and with the same look and feel as the public version of the application.

The *network emulator* controls the network conditions and provides network emulation functionality by emulating network impairments. It allows: (i) bandwidth throttling; (ii) adding delay and its variation (jitter); (iii) adding packet loss and emulate packet burst losses, hereafter referred to as Mean Lost Burst Size ($mlbs$). In our testbed, the network conditions are controlled by *NetEm*. Additionally, our testbed can also allow CPU throttling, controlled on the client side.

The *controller* remotely controls the network conditions in real-time at a running system, which is required, since modern networks have typically fluctuating performances during one session. It is also worth noting that the clocks at the clients' devices need to be properly synchronized via the Network Time Protocol (NTP) in order to take the temporal aspects of the collected data into account.

2) Data Acquisition: The developed testbed enables the collection of different types of data, including explicit, subjective and implicit, behavioral user feedback, and performance-related statistics. We will now discuss the latter in more detail.

---

[2]Appear.in: https://appear.in

2.1) WebRTC Statistics in Google Chrome: The *webrtc-internals*[3] functionality enables observation of the performance of the WebRTC connections locally in the browser. The stats gathered here are based on the statistics specified in the WebRTC Application Programming Interfaces (APIs). The W3C WebRTC statistics define objects to observed Real-Time Transport Protocol (RTP) statistics for the different data channels and tracks. For more background on these statistics and how they can be classified we refer to [11]. In webrtc-internals, a JSON file contains all PeerConnection objects defined in W3C APIs as well as some Google-specific stats. The reports include several media and network statistics, such as (video)frame rate, packet loss, and bitrate. End users can view these statistics in real-time or download them in a single file any time during a session, or download the statistics immediately after a session before the browser window is closed.The statistics are collected per browser, which means that in order to assess the performance of a multi-party session, the statistics from all browsers used in the session need to be recorded, downloaded, and (manually) combined and synchronized. Despite their inherent limitations, these statistics can be used to study QoE aspects of WebRTC services [12].

2.2) WebRTC Statistics in getstats.io: Appear.in recently launched a customizable WebRTC Analytics interface called *getstats.io*. It provides real-time visualization of WebRTC service performance stats. This interface integrates the performance stats of all parties involved in a call (gathered with a sampling time of every 10 seconds). The stats files collected from Google Chrome and getstats.io are both stored in a JSON format, but with a different structure and to some extent a different content.

2.3) Network performance: Network-oriented performance measurement tools, such as Wireshark, are complementary to WebRTC API stats and can be used for QoE troubleshooting and network diagnosis, etc. Compared to the data gathered from these network performance measurement tools, WebRTC API stats are application-oriented and more directly linked to how the user actually experiences the application and what is perceivable at the user side.

2.4) Explicit user feedback and questionnaires: Our testbed enables gathering of explicit QoE feedback immediately after a session ends (post-experience). The provided feedback can have different forms, e.g., a star rating, or answering a short questionnaire. Questions can relate to different quality features and how they were perceived by the user, to the context of the call, e.g., professional vs. leisure context, to the affective state of the user, the experienced impact on the interaction between the different conversation partners, etc. As mentioned, the requested feedback can be modified depending on the focus of the study. Both closed and open-ended questions are supported, as the questionnaires can be easily defined in an electronic format and displayed to the participants in an experiment. The feedback form at the end of a session can also be made an intrinsic part of an experiment, e.g., in the sense that the completion of a post-experience questionnaire can trigger the next step of the experiment. Additionally, further insights into the user's personal traits can be gained by integrating corresponding questions into the questionnaires before or after the experiments. Biases and attitudes, e.g, towards the application under investigation, as well as familiarity and criticality with online multi-party video communication in general and in particular appear.in, should be considered.

2.5) Behavioral User Feedback: When users experience different types and gradations of quality impairments (e.g., video freezes, bad or no audio), they may react in different ways. Users reactions may, for example, include: muting audio, disabling video, changing video resolution, aborting the call, retrying to connect, etc. Our testbed offers the possibility of measuring and monitoring what is going on at the performance level, what is perceivable by users and allows to gain an insight into how how users (actively or passively) react on a range of audiovisual impairments and their related tolerance levels. Evaluating and understanding such active users' reactions may help to improve WebRTC-based applications and optimization of QoE in constrained scenarios.

2.6) Audio and Video Recordings: The transmitted audio and video contents of each individual participant are recorded on the sending and on the receiving side. This allows us to analyze the conversation after the experiment from the perspective of each participant. We created a web interface tool for visualizing an experiment session and analyzing audio and video contents along with objective and subjective data. This tool is described in Section III-B. Furthermore, quality measurement tools for voice and audiovisual analysis based on human perception (e.g., PEXQ[4]) can also be used in evaluating the recorded audio and video contents. These tools provide a perceptual evaluation of audio / video quality for Mean Opinion Score (MOS) scoring of such telemeeting systems and applications.

2.7) Integration of additional methods and tools to capture user feedback: As these different data types are gathered in real-time, and can easily be visualised using the dashboard, the experimenter can easily have a look at the gathered data and ask for clarifications or additional information after the experiment, for instance in an individual debriefing interview for each participant or focus group discussion. Moreover, the testbed can easily be augmented with tools allowing e.g., the collection of physiological data, which can be gathered continuously and unobtrusively and which may provide more detailed insights into the user's affective state.

B. WebRTC Dashboard

The WebRTC dashboard consists of a web interface tool that makes it possible to visualize the most relevant factors that may influence user's QoE in an easy and straightforward way. The procedure is as follows: A real-time audiovisual conversation using the research version of appear.in takes place and Google Chrome is used as browser to access appear.in. Then, Real-time logging of performance and session-related statistics of the appear.in session is performed. For this purpose, we use both Google Chrome's WebRTC internals functionality and the WebRTC Analytics interface getstats.io. Additionally, subjective user feedback (collected at the end of a session), implicit and behavioral user feedback (collected during a session), as well as audio and video recordings, are collected.

---

[3]Chrome: chrome://webrtc-internals, and in Opera: opera://webrtc-internals

[4]PEXQ: http://www.opticom.de/products/pexq.html

Finally, the developed dashboard replays the collected data and exhibits them in a graphical and interactive way. See Fig. 2 for an illustration.
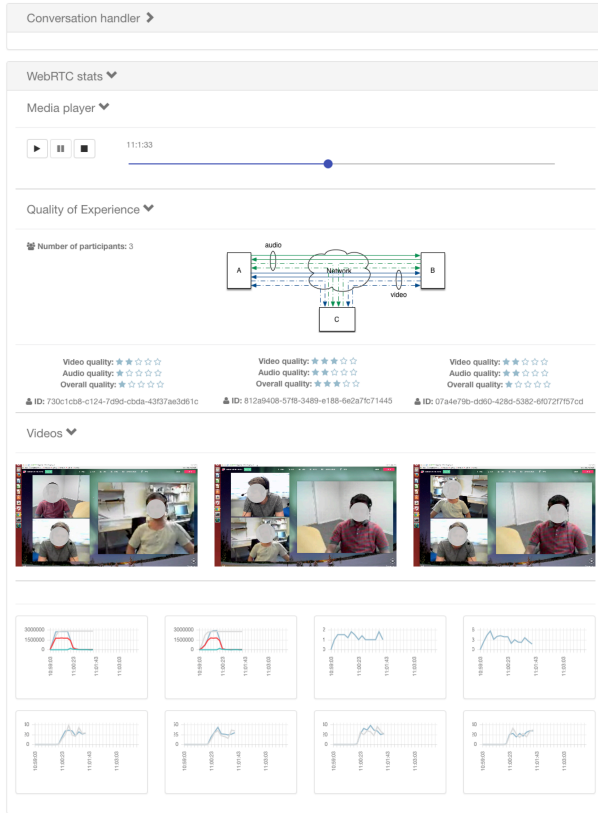

Figure 2: Snapshot of WebRTC dashboard that replays collected data of a three-party video communication

The WebRTC dashboard is composed of two main panels: *Conversation Handler* and *WebRTC stats*. The *Conversation Handler* is responsible for handling WebRTC-based conversations. It allows users to choose and manage the settings of a conversation, and customize the presentation of the collected data in the *WebRTC stats* panel. The *WebRTC stats* provides performance results and offers an evaluation, both objective and subjective, of audio and video contents[5].

IV. USE CASE: SUBJECTIVE ASSESSMENT OF A TWO-PARTY VIDEO CONFERENCING USING APPEAR.IN

In this section, we report on the set-up and results from a small-scale subjective study on the impact of packet loss on perceived quality in the context of a two-party audiovisual conversation using appear.in. The study took place in a controlled lab setting and under different network conditions, enabled by the WebRTC testbed. The features and the practical applicability of our experimental platform are further illustrated here.

A. General set-up and methodology

The experiment consisted of 2 parts. First the participants were welcomed by the experimenter, who briefed them about

[5]This dashboard available at: http://appear01.item.ntnu.no:3000/

Table I: Technical specifications

Hardware	Description
Desktops	HP Compaq Elite 8100 SFF – Intel® Core™ i7
Displays	HP LP2465 24" 1920 x 1200 (60Hz)
Headsets	Koss SB45
Webcams	Microsoft LifeCam Studio

the scope of the study and who explained the procedure of the experiment in detail. The participants received instructions about their task and about the post-session questionnaire to fill in. Next, they were asked to sign a written consent form and were invited to fill in a pre-questionnaire with some general information about themselves and their familiarity with multi-party video communication. Before moving to the test rooms, they were given to possibility to ask any questions they still had and the experimenter provided clarifications if needed. For the second part (and start of the actual experiment), they were escorted to different test rooms. They were seated within a viewing distance (around 70 cm) from the monitor, as recommended by ITU-T P.910 [13]. Participants were then asked to begin the experimental task (see further), and they were exposed to the first of five experimental conditions (see Section IV-D for details).

The whole procedure had a total duration of one hour, with an active QoE testing part of about 30 min. The QoE testing part contains 5 conditions and includes a 3 to 4 min break in-between the conditions. Each testing condition consisted of two-party video conference session using appear.in with a duration of 4 min. After each session participants were asked to rate the perceived quality using a 5-point absolute category rating (ACR) scale with the following values: (1) bad; (2) poor; (3) fair; (4) good; (5) excellent. To be more precise, we asked the following questions, which were answered within 2 min after each session:

Q1: *How would you rate the overall audiovisual quality of the session (the overall combined audio and video quality)?*
ACR scale: 5-Excellent; 4-Good; 3-Fair; 2-Poor; 1-Bad.

Q2: *How would you rate the video quality of the session?*
ACR scale: 5-Excellent; 4-Good; 3-Fair; 2-Poor; 1-Bad.

Q3: *How would you rate the audio quality of the session?*
ACR scale: 5-Excellent; 4-Good; 3-Fair; 2-Poor; 1-Bad.

Q4: *Which quality-related issues have you experienced during the session? Several answers are possible.*
Answer: Audio problems: bad audio or no audio at all; Video problems: bad video or no video at all; Bad synchronization between audio and video; Not applicable (never experienced any problems); Other, please specify.

Q5: *Have you considered quitting the session because of quality-related issues?*
Answer: Yes; No.

Q6: *Did you perceive any reduction in your ability to interact with the other party during the session?*
Answer: Yes (If Yes, specify the problem if you could); No.

In total 30 test conversations took place (6 groups and 5 conditions per group, see further). The subjective assessment followed the test procedure described in ITU-T Recommendation P.1301 [14] as much as possible.

B. Test Environment

The experiment was conducted in an office environment. The physical parameters during the testing sessions were slightly different for different participants, as the conversation partners were located in separate rooms. However the two different rooms that were used were selected so that they had very similar dimensions, lighting and background conditions. Both rooms were equipped with identical desktop computers, displays, webcams and headsets (see Table I for details).

C. Test participants

In total, we collect data from 12 participants, including 2 females and 10 males aged between 22 and 41 years (mean 27.67, median 25.5), recruited among the employees and students at the Norwegian University of Science and Technology (NTNU), Norway. All participants reported to frequently use online video communication: for 4 out of 12 participants this means 2-3 times per month and the remaining 8 use online video communication services at a weekly basis (3 of them use such services even daily). Generally speaking, all participant indicated in the pre-questionnaire that good audio quality is important to very important. For video quality, the expectations are still high, yet slightly lower than for audio. 5 out of 12 participants indicate to be rather indifferent to the video quality. Participants were divided into six groups with 2 members each to take part in the test sessions. They were placed together with another participant that they knew, so that the being unfamiliar with the conversation partner would not be an additional influence factor to consider.

D. Experimental Conditions

To study the impact of packet loss on QoE ratings, we consider the following five conditions:

$C1$: No packet loss ($plr = 0\%$ and $mlbs = 0$ packet)
$C2$: $plr = 10\%$ and $mlbs = 1.5$ packets
$C3$: $plr = 10\%$ and $mlbs = 3$ packets
$C4$: $plr = 20\%$ and $mlbs = 1.5$ packets
$C5$: $plr = 20\%$ and $mlbs = 3$ packets

where plr=packet loss ratio and $mlbs$=mean loss burst size.

The controlled variables (also called *independent variables*) are listed in Table II. Packet losses were remotely introduced (see Fig. 1) 30 seconds after the beginning of each session and lasted for three minutes, followed by 30 seconds of good quality (no packet losses), which was needed to be be able to collect the performance statistics via the WebRTC internals functionality.

In this experiment, a symmetric call condition was considered, so every participant would see similar distortions. Note that the five experimental conditions were randomly ordered

for each of the six groups, so that potential confounding effects due to the following of a specific order of the experimental conditions could be avoided.

E. Conversation task

To assess the conversational quality of multi-party video conferencing, it is necessary to involve conversation partners in an appropriate conversation task. Within our lab experiment, we employed the survival task from ITU-T P.1301 [14], and we adapted it for a two-party video conferencing. The goal of the survival task is to explore the performance characteristics of a decision-making group. Participants were invited to imagine themselves in a survival situation based on an accident (space rocket, plane, etc.). They have a list of twelve items (listed with photographs to avoid that participants take too much time to look at their sheet), which were divided into two different 6-item lists, one for each participant. They have to share their objects, then discuss them together and come to an agreement on six objects to be selected (see ITU-T P.1301 [14] for details).

F. Experimental results

Fig. 3 depicts the average quality ratings with 95% confidence intervals for the three dependent variables (audiovisual, audio and video quality) in the five experimental conditions. It shows that both plr and $mlbs$ seem to impact the overall audiovisual quality, as evaluated by the participants. Lower audiovisual quality ratings were found when plr is set to large value, typically $plr = 20\%$. This pattern was also reflected in the answers to questions 5 and 6 (see section IV.A): In the experimental conditions 4 and 5 (for both $plr = 20\%$ applies), the percentages of participants that considered quitting the session are respectively 17% and 33% (compared to 0% in conditions 1 and 3 and 8% in condition 2). Similarly, one out of four participants felt that it was more difficult to interact with the other party in condition 4. For condition 5, this percentage increases even further (42%). On the other hand, the perceived audio quality rating decrease slowly when either the plr or $mlbs$ increases. The only case with poorer audio quality ratings (average MOS equal to 2.7) is for $plr = 20\%$ and $mlbs = 3$. Finally, the perceived video quality ratings are highly impacted with the increasing values of plr and $mlbs$. It is also worth noting that smaller values of the $mlbs$ – high frequency of packet loss bursts with a small number of consecutive packet losses per burst – lead to lower video quality ratings. More specifically, for the same plr values, for instance when $plr = 10\%$, the frequency of the packet loss bursts has a larger influence on the video quality than larger burst sizes – large number of consecutive packet losses per burst – with low frequency of loss bursts.

V. Conclusion

In this paper, we introduced an experimental platform for conducting research on Quality of Experience (QoE) and experienced quality in the context of WebRTC-based video communication. The platform consists of a testbed which enables the set up and running of QoE studies in a controlled environment with up to 8 parties in one call, and a Dashboard that enables easy data visualization and in-depth data analytics. The use of our experimental platform was demonstrated though

Table II: Controlled variables

Controlled Variable	Description	Levels
plr	Packet loss ratio	0 (none), 10, 20
$mlbs$	Mean Lost Burst Size	0 (none), 1.5, 3

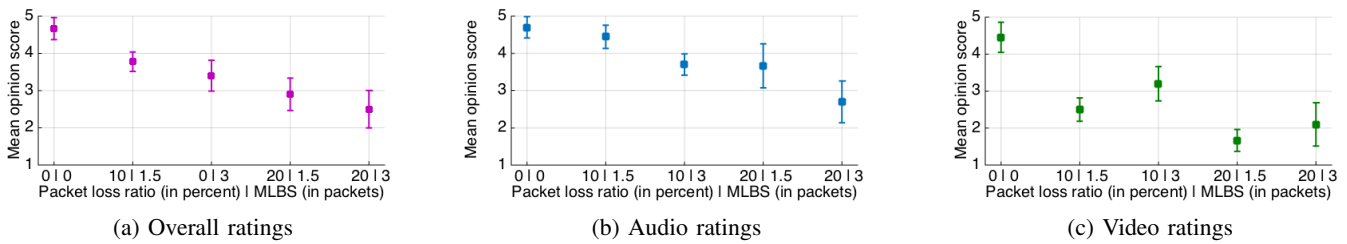(a) Overall ratings   (b) Audio ratings   (c) Video ratings

Figure 3: Influence of packet losses on quality ratings of a two-party video conferencing using appear.in

a small-scale pilot test in a controlled lab setting. The focus was on study of the impact of packet loss on quality ratings in a two-party WebRTC-based video communication and indicated that the experienced audio quality was worst in the test scenario which had a packet loss ratio of 20 % and a mean loss burst size of 3. The experienced video quality was however worst when participants were exposed to a packet loss ratio of 20 % and a mean loss burst size of 1,5. We showed that the platform allows us to control and directly manipulate several system- and in particular network-related parameters. As the experimental platform also provides real-time logging of performance and session-related statistics and as it enables the easy collection of explicit and implicit user feedback, it is also a very promising tool for data gathering outside of the lab, in test users' natural environment. More specifically, it enables to run longitudinal user studies, which could focus on how performance issues and short and longer-term QoE ratings are related, and how they may be interlinked with a set of other technical and non-technical influence factors which are typically hard (or impossible) to investigate in a controlled lab environment only approach.

REFERENCES

[1] K. Brunnström, S. A. Beker, K. De Moor, A. Dooms, S. Egger, M.-N. Garcia, T. Hossfeld, S. Jumisko-Pyykkö, C. Keimel, M.-C. Larabi, B. Lawlor, P. Le Callet, S. Möller, F. Pereira, M. Pereira, A. Perkis, J. Pibernik, A. Pinheiro, A. Raake, P. Reichl, U. Reiter, R. Schatz, P. Schelkens, L. Skorin-Kapov, D. Strohmeier, C. Timmerer, M. Varela, I. Wechsung, J. You, and A. Zgank, "Qualinet White Paper on Definitions of Quality of Experience," Mar. 2013, qualinet White Paper on Definitions of Quality of Experience Output from the fifth Qualinet meeting, Novi Sad, March 12, 2013. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00977812

[2] U. Reiter, K. Brunnström, K. De Moor, M.-C. Larabi, M. Pereira, A. Pinheiro, J. You, and A. Zgank, *Factors Influencing Quality of Experience*. Cham: Springer International Publishing, 2014, pp. 55–72. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-02681-7_4

[3] J. Skowronek, K. Schoenenberg, and G. Berndtsson, "Multimedia Conferencing and Telemeetings," in *Quality of Experience: Advanced Concepts, Applications, and Methods*. Springer, 2014.

[4] J. Skowronek, F. Schiffner, and A. Raake, "On the influence of involvement on the quality of multiparty conferencing," in *4th International Workshop on Perceptual Quality of Systems, Vienna*, 2013, pp. 25–30.

[5] M. Schmitt, S. Gunkel, P. Cesar, and D. Bulterman, "Asymmetric Delay in Video-Mediated Group Discussions," in *International Workshop on Quality of Multimedia Experience (QoMEX)*, Sep 2014, pp. 19–24.

[6] G. Berndtsson, M. Folkesson, and V. Kulyk, "Subjective quality assessment of video conferences and telemeetings," in *Packet Video Workshop (PV), 2012 19th International*, May 2012, pp. 25–30.

[7] M. Schmitt, S. Gunkel, P. Cesar, and P. Hughes, "A QoE Testbed for Socially-aware Video-mediated Group Communication," in *Proceedings of the 2Nd International Workshop on Socially-aware Multimedia*, 2013, pp. 37–42.

[8] S. Gunkel, M. Schmitt, and P. Cesar, "A QoE study of different stream and layout configurations in video conferencing under limited network conditions," in *Quality of Multimedia Experience (QoMEX), 2015 Seventh International Workshop on*, May 2015, pp. 1–6.

[9] M. Schmitt, J. Redi, P. Cesar, and D. Bulterman, "1mbps is enough: Video quality and individual idiosyncrasies in multiparty hd video-conferencing," in *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*, June 2016, pp. 1–6.

[10] D. Vucic and L. Skorin-Kapov, "The impact of mobile device factors on QoE for multi-party video conferencing via WebRTC," in *Telecommunications (ConTEL), 2015 13th International Conference on*, July 2015, pp. 1–8.

[11] D. Ammar, K. D. Moor, M. Xie, M. Fiedler, and P. Heegaard, "Video qoe killer and performance statistics in webrtc-based video communication," in *2016 IEEE Sixth International Conference on Communications and Electronics (ICCE)*, July 2016, pp. 429–436.

[12] D. Ammar, P. Heegaard, M. Xie, K. De Moor, and M. Fiedler, "Revealing the dark side of webrtc statistics collected by google chrome," in *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*, June 2016.

[13] "ITU-T P.910: Subjective video quality assessment methods for multimedia applications ," April 2008.

[14] "ITU-T P.1301: Subjective Quality Evaluation of Audio and Audiovisual Multiparty Telemeetings ," July 2012.

Evaluation of Mobile Application Prototype in Context of Design Against Human and Computer Interaction

Mat Redhuan Samsudin [1], Tan Tse Guan [2], Anuar Mohd Yusof [3]
1,2,3, Faculty of Creative Technology and Heritage,
Universiti Malaysia Kelantan, Locked Bag 01,
16300 Bachok, Kelantan, Malaysia
redhuansamsudin@gmail.com [1], tan.tg@umk.edu.my [2], anuarmy@umk.edu.my [3]

ABSTRACT

Application design is part of a process to produce a prototype or multimedia product by integrating several multimedia elements, including text, graphics, audio, video and animation. In the process, it involves some of the important features of Human Computer Interaction (HCI). This study aimed to identify appropriate application designs and evaluate the quality of the prototype which is a tool for teaching special education students. In this study was conducted a summative test of prototypes to see the effectiveness of application design taking into the characteristics of human and computer interaction. Results showed the application excellent in the quality of the application and management. In terms of the pedagogy strategies and integration of multimedia elements showed is good and accepted to proceed as a tool in teaching and learning. In the conclusion, the design of teaching and learning tool should follow that criteria for giving the impact to the students.

Keyword : Human computer interaction, prototype design, mobile application, special education, teaching and learning tool.

1 INTRODUCTION

The prototype is the process of developing a model that has the same characteristics or the same as the actual model. Prototype is a system that is almost identical to the real product.

The application, even it is not fully completed yet, but the basic function is like the real system. The prototype is also must have in multimedia product development, where it is used to develop into one end product that everyone can use. Additionally, it is also an efficient and effective way to review and optimize application designs or content, assignments through repeat discussion, browsing, testing and review processes.

The development of prototypes is aimed at obtaining feedback from consumers' targets in an educational context known as teaching design or teaching technology. It aims to ensure more effective teaching and learning process and meet the needs of students [1]. In the context of teaching and learning, some elements need to be emphasized as teachers, teaching materials, media, and learning environments. Researchers in this study design mobile applications based on the ADDIE model. Based on that model, its design is systematic that involves several phases including analysis, design, development, implementation and evaluation. The development of this application uses adobe flash 6 software with 3.0 action script, and adobe premiere.

2 PROBLEM STATEMENT

Recent studies showed that there are many advantages of mobile applications practiced in teaching and learning but in the context of human computer interaction design (HCI) it is still under-explained especially in the context of special education. As practiced at the university level, mobile applications has been widely used for communication, establishing collaboration and building clearer knowledge and understanding.

This learning pattern is more effective. Therefore, researchers need to look at this important aspect to improve the quality of disability students especially at the school level [2]. Special studies need to be done to identify the designs that suit the needs of the students.

Besides that the study performed by showing mobile applications has access problems to browsers which takes a long time to loading the files [3]. It may be affected by slow internet access as well as large file sizes. Therefore, a summative evaluation is done on mobile application design to improve the quality of an application that is appropriate to the student's level and needs. This assessment is also expected to be a guide to designers to develop a mobile application.

3 RESEARCH OBJECTIVES

i. To identify appropriate application designs for students with hearing impairments and conversations

ii. To evaluate the quality of the prototype which is a tool in teaching and learning for special education students

4 HUMAN COMPUTER INTERACTION

Human computer interaction (HCI) is the concept of interaction or relationship between humans and computers as well as applications [4]. The association for computing machinery states that (HCI) is a discipline that includes design, evaluation and implementation of interactive computer systems for humans to make it easy to use. The main thing in the interaction between humans and computers is the satisfaction of using the system in performing a task or learning.

In our daily lives, we will always interact with others to do all the important things in life. Computer is a tool used in our daily life that gives importance to people in education, employment, communication and it becomes a necessity in life.

Computer also has an influence in daily life in improving good achievement in most aspects of life. While computers come with a great variety of technologies, this human computer interaction concept needs a variety of support tools such as keyboards, mouse, and others as a support tool for interacting.

The concept of human computer interaction (HCI) can provide interaction processes in a better learning environment. Although the computer was originally only used by users to deliver messages that were understandable, it had gone through the process of providing more effective interaction. Hand movement is one of the elements in human computer interaction [5] and it is an element that has a great influence on sign language learning in education [6]. In the context of sign language learning, hand movement elements are the main means of translating feelings, actions, and instructions because sign language is the language used by deaf and mute persons.

5 EVALUATION CRITERIA

In this study the prototype has been developed through the testing process to see the quality or usage of the application. Usability testing is a common thing that comes after a given application is designed to see how far an application works according to the needs of the user. Based on feedback from developer users to analytics to improve the quality of the app.
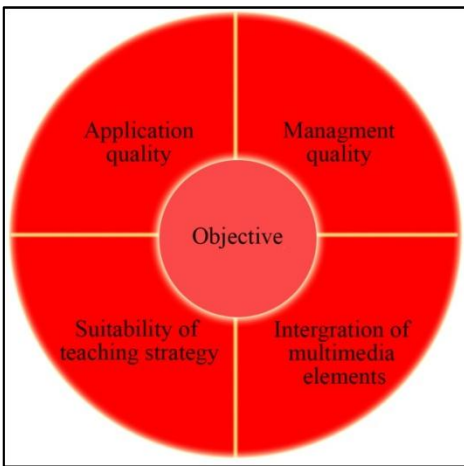
Figure 1. Evaluation elements

5.1 The overall quality of the application

Quality evaluation of a prototype involves criteria such as user interface designs that involve several aspects such as usability, consumerism, consumer satisfaction and interactivity. Past studies showed that good prototype quality as a whole involves the aspect that is being emphasized in a prototype [7, 8]. As for usability, it refers to how the feature works well and does not have any errors in the application system [9]. Consumerism refers to how the application is used as their daily needs. In the context of education, the application refers to the need for students to help them in their lesson.

While customer satisfaction involves the feeling of the user which depends on the advantages of the application that can provide benefits and thus create a satisfaction. Digital learning is a learning that has a better effect as it is more interactive. This aspect of interactivity is also assessed to obtain positive feedback on the application. Interactivity is aimed at two-way communication between users and applications where users can give directions to the applications and subsequently applications can respond to users. It sees a systematic data processing.

5.2 Quality of application management

Management quality refers to how each component is organized neatly, where the layout is more systematic and neat. And this makes it easy for users to use the app. And it affects to the user's age in which the learning process takes place faster and easier. Previous studies by [2] say mobile applications practiced in education have complex layouts and have hidden choices. Hence researchers in this study measure the quality of management and look at end-user acceptance and make this application acceptable to other users in deaf and mute communities.

5.3 The suitability of teaching and dissemination strategies

Looking at the technology of today's mobile technology is a necessity not just for parents but students also have their own smart phone. Smartphone are not only limited to their functions for communication but they can be utilized in broader contexts. Among the areas that can be used is education where many mobile applications have been created to help students. As mobile apps it is now a culture of teaching and learning as it can be done at anytime and anywhere according to the ability of learning [2] [10].

5.4 Multimedia elements in mobile applications

In this application the designer integrates four multimedia elements as follows,

5.4.1 Text

It is a basic element of conveying information. Put emphasis on the content you want to convey. Have a variety of types and forms of writing that are capable of attracting information delivery. Most multimedia systems use texts as it is an effective way to communicate, an easy way to provide instructions to users, to deliver information in various fields and to be information mediation

medium. Text consists of four printed texts, scanned texts, electronic texts and hypertext. In multimedia applications, the text will clarify the required information. The existence of the text will be further clarified with the use of other multimedia elements and it is an important element. Text can be presented in various forms to attract users like graphics, animations, titles and others. Text presentation to users depends on the purpose and number of users.

5.4.2 Graphic

Graphic is a presentation of immovable images such as pictures, drawings, sketches, illustrations and others. It is very important to emphasize in the process of delivering information. Graphic is able to make a presentation more interesting and able to focus the attention of the users with the information they want to convey. According to [11] Info graphic becomes one of a good platform as a visual communication in promoting the knowledge in an efficient manner. The importance of graphics in an app is to increase the emphasis of information, attract users' attention, illustrate a concept, and act as a backdrop for a concept.

5.4.3 Audio

Audio is one of the most frequently used media in any interaction between computer and user. Music elements, background sound and sound effects are some examples of audio usage that can be used in a multimedia application. As with other multimedia elements, audio is able to deliver information more effectively, creating a more prominent atmosphere and attracting attention to what it wants to convey. Audio works to help deliver information more effectively. It also helps to increase the attractiveness of a presentation as well as to increase motivation among consumers to be more interested in following a process of information delivery and creating more attractive atmosphere and will focus on the content presented.

5.4.4 Animation

Animation is a fast show of the sequence of 2-D or 3-D static images to create the illusion of movement. Animation will leave the effect of optical illusion of movement due to the phenomenon of vision retention, and can be demonstrated and done in several ways. The most frequently used method is to present animation as a movie or video program. Animation works to deliver feedback and reinforce activities, describing the structure, function, sequences between components within a specific domain. Animation is capable to convey huge information in successive images. Moreover, animated info graphic can make the audience to be more understandable with the things that are not clear enough [11]

6 RESEARCH METHODOLOGY

The researcher in this study uses a survey method whereby, the respondents are conducted to four respondents who are classified as experts in the field. Each respondent is selected based on their background. It involve difference field of background, they who have been taught more 15 years of teaching and expert in the sign language. Also involved in this study are designers who have served as designers for more than 10 years. This study also involved 2 students and 2 lecturers in which the selection was made to identify the needs of students in an application.

Using the assessment form a survey was conducted against the six respondents to complete the assessment form. The researchers conducted a summative assessment of the prototype. Summative evaluation is more appropriate for the evaluation of a prototype it seeks to obtain results based on user responses [11]. Before the respondents completed the evaluation form they were given a prototype to use it, and subsequently they completed the assessment form. Analysis is done based on accepted assessment forms. Figure 1

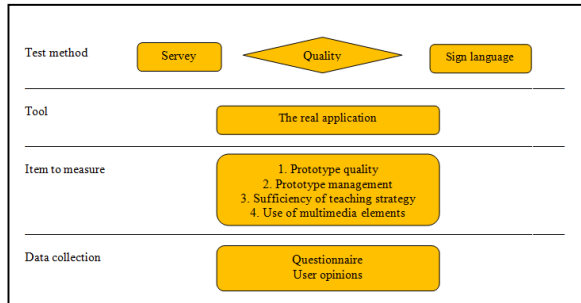shows the prototype testing framework to identify the usability and functionality of the mobile application.

Figure 2. Test method

App review using test form by [12]. There are two divisions that are objective and subjective. On the objective part of each test item uses four likert scales as shown below. In the subjective section, users need to provide comments and suggestions to improve the quality of the app.

Table 1. Likert scales

Score value	Scale
Excellent	1
Good	2
Modest	3
Poor	4

7   APPLICATION DESIGN

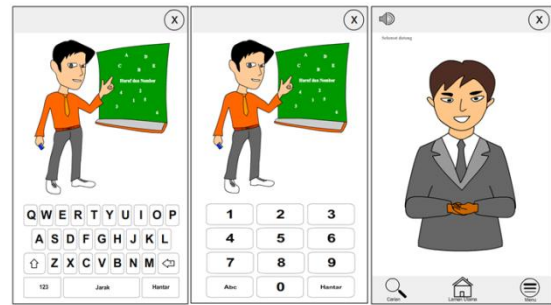**Figure 3.** Loading/home/menu design

Figure 4. Letters/numbers/words

8   DATA ANALYSIS

Respondent demographic

The assessment was given to three groups of teachers or lecturers, designers, and students each group has to respond. Respondents for this test involved a woman and five men. However the respondent's gender did not influence the results of the test.

Number of User	Application quality	Management quality	Pedagogical strategies	Integration of multimedia elements
1	Excellent	Excellent	Good	Good
2	Excellent	Good	Excellent	Excellent
3	Good	Good	Good	Good
4	Good	Good	Good	Good
5	Excellent	Excellent	Excellent	Good
6	Excellent	Excellent	Good	Excellent

Figure 5. User feedback

Results showed that the quality of the app showed that 4 persons (66.67%) stated that the quality of the application was excellent and 2 (33.33%) were good. On the other hand, in the application management aspect, 3 person (50.00%) were excellent and 3 person (50.00%) were good. The pedagogical strategy aspect shows that 2 person (33.33%) are excellent and 4 person (66.67%) are good and the aspect of multimedia element integration is 2 person (33.33%) are excellent and 4 person (66.67%) are good.

Based on the expertise made by the expert it shows that this mobile application prototype can be used and applied in learning because it works well like real applications. However, from the teacher's perspective and the suggestion it suggests that the

background of each word reflect its purpose to make it clearer. Additionally, users also recommend that this app adapt augmented reality (AR) technology to provide more effective learning experience. From the designer's perspective, it also suggests that animation should be streamlined to make it attractive.

9 CONCLUSION

Dependence on technology in life is helpful in any field and has been proven by many earlier researchers. Technological contribution in education also has a tremendous impact. Based on the results of the study it shows that the acceptance of users towards technology in education is very good, this is clearly supported by [13]. However it should be based on the characteristics and needs of the students to meet the needs and to complete a learning process.

Design suitability affects the frequency of application usage so the design aspect should include the user's features to meet the needs of users in order to more user friendly technology. Good design improves the students academic achievement and more critical thinking skills. It also makes students active learning and improves their professional knowledge and application of knowledge [14]. Thinking ability can make students more visionary and make students more innovative. This finding is clear and aligned with [15].

The need for technology in education makes the design aspect should be the focus of the prototype development because learning depends on the modules provided. And it is also proven by [16] that is suitable for learning. Self-learning, informational presentation needs to be more creative so that it has more impact on users. Aspects of application management are also an important issue in development, so users can communicate with the system so as not to lose purpose. Hence some of

these aspects are for mobile application development and it is evident that mobile applications can enhance student motivation in learning [17].

REFERENCE

1. Rio Sumarni Shariffudin, Abdul Hafidz Omar, and Dayang Hajjah Tiawa Awang Haji Hamid, *Reka Bentuk Dan Keberkesanan Perisian Multimedia Membaca Faham Berasaskan Gambaran Visual Bagi Kanak-Kanak Prasekolah* 2007, Universiti Teknologi Malaysia: Institutional Repository. p. 7.
2. Hocutt, A.M., *Effectiveness of Special Education: Is Placement the Critical Factor?*, in *The Future Of Children*. 1996.
3. Htun, K.S., *Performance Evaluation of Cloud Computing for Mobile Learning over Wireless Networks*, in *The International Conference on E-Technologies and Business on the Web*. 2013, SDIWC: Thailand.
4. Love, S., *1 - Introduction to mobile human-computer interaction*, in *Understanding Mobile Human-Computer Interaction*. 2005, Butterworth-Heinemann: Oxford. p. 1-13.
5. Pavlovic, V.I., R. Sharma, and T. Shuang, *Visual Interpretation Of Hand Gesture For Human Computer Interaction* Journal of Engineering Research and Applications, 2014. Vol 4(1).
6. Rempel, D., M.J. Camiller, and D.L. Lee, *The design of hand gestures for human–computer interaction: Lessons from sign language interpreters.* Int. J. Human-Computer Studies, 2014. 72: p. 728-735.
7. Ali, A., Alrasheedi, M., Ouda, A., & Capretz, L. F. (2014). A Study Of The Interface Usability Issues Of Mobile Learning Applications For Smart Phones From The User's Perspective. International Journal on Integrating Technology in Education (IJITE), Vol.3(No.4).
8. Feras Tarawneh, Fauziah Baharom, Jamaiah Hj. Yahaya, & Faudziah Ahmad. (2011). Evaluation and Selection COTS Software Process: The State of the Art. International Journal on New Computer Architectures and Their Applications, 1(2), 344-357.

9. Zulkefli Mansor, Zarinah Mohd Kasirun, Saadiah Yahya, & Noor Habibah Arshad. (2012). The Evaluation of WebCost Using Software Usability Measurement Inventory (SUMI). International Journal of Digital Information and Wireless Communications, 2(2), 197-201.

10. Mohd Hanafi Mohd Yasin, Hasnah Toran, and Mohd Mokhtar Tahar, *Tinjauan Terhadap Tahap Kesesuaian Infrastruktur Program Integrasi Pendidikan Khas Bermasalah Pembelajaran.* Asia Pacific Journal of Educators and Education,, 2011. Vol. 26(No. 1): p. 159-171.

11. Nur Nazihah Rahim, Nik Zulkarnaen Khidzir, Anuar Mohd Yusof, & Aznan Zuhid Saidin. (2016, 22 November 2016.). *An Assessment of Quality on Animated Infographics in an Islamic Context.* Paper presented at the International Conference on Teaching and Learning in Education.

12. Azlina bt Hashim, et al., *Multimedia Principles*. 2003, Kuala Lumpur: Cosmopoint Sdn Bhd. 391.

13. Jeng, Y.-L., et al., *The Add-on Impact of Mobile Applications in Learning Strategies: A Review Study.* Educational Technology & Society, 2010. 13(3): p. 3-11.

14. Yi, W., *Study on Implementation and Evaluation of Curriculum Based on Blended Learning -Take Course "Investments" Teaching Practice as Example*, in *The Third International Conference on E-Learning and E-Technologies in Education (ICEEE2014)*. 2014. p. 125-132.

15. Norasikin Fabil, et al., *Aplikasi Reka Bentuk Sistem Visualisasi Maklumat Berasaskan Teori Persepsi Visual Dalam Ilmu Hadis.* Journal Of Islamic And Arabic Education, 2011. Vol 3(1): p. 12.

16. Johnson Lim Soon Chong, Jailani Md Yunos, and Ghazally Spahat, *The Development and Evaluation of an E-Module for Pneumatics Technology.* Malaysian Online Journal of Instructional Technology, 2005. Vol. 2(No. 3).

17. Maszuraini Miswan and Hamedi bin Mohd Adnan, *Pembangunan Aplikasi Peranti Mudah Alih untuk Kemahiran Membaca Kanak-Kanak: Aplikasi Literasi LINUS (LiLIN).* Jurnal Pengajian Media Malaysia, 2015. Vol. 17, (No. 2).