Proceedings of
Umeå's 15th Student Conference in Computing Science

USCCS 2012

Suna Bensch, Frank Drewes, Håkan Gulliksson, Thomas Mejtoft (editors)

UMINF 12.02
ISSN-0348-0542

UMEÅ UNIVERSITY
Department of Computing Science
901 87 UMEÅ, SWEDEN

Preface

Umeå's Student Conference in Computer Science is the highlight of the conference course in our Computer Science curriculum. The idea and objective of the course is to give the students a forum where they can develop their own ideas in a scientific manner, thereby improving their understanding of how research is conducted and how the achieved results should be presented according to scientific standards. The conference format was chosen to provide a realistic environment in which the research results can be presented.

A student who participates in the course first selects a topic and a number of research questions that he or she is interested in. If the topic is accepted, then the student outlines a paper and composes an annotated bibliography to give a survey of the field. The main work consists in conducting the actual research that answers the questions asked, and convincingly reporting the results in a scientific paper. Each submitted paper receives two or more reviews written by members of the department. If the reviews are favourable, the paper is accepted, meaning that the student is given the opportunity to present his/her work at the concluding conference, and to submit a final version that will be included in the proceedings. The course thus gives an introduction to independent research, scientific writing, and oral presentation.

This year was the fifteenth offering of the course, resulting in a total of 16 submissions, of which 13 were accepted for presentation at the conference. Except for one, all of these are included in these proceedings. We are grateful to the reviewers who helped us to evaluate the submissions within a very short time frame.

Umeå, 9 January 2012

Suna Bensch
Frank Drewes
Håkan Gulliksson
Thomas Mejtoft

Organizing Committee

Suna Bensch
Frank Drewes
Håkan Gulliksson
Thomas Mejtoft

Reviewers

Ahmed Aley
Suna Bensch
Martin Berglund
Andrii Dmytryshyn
Frank Drewes
Patrik Eklund
Karin Fahlquist
Håkan Gulliksson
Johanna Högberg
Francisco Hernandez
Thomas Hellström
Stefan Johansson
Thomas Mejtoft
Dipak Surie
Shafiq Urrehman

Table of Contents

Using Databases in Virtualized Environment

Syed Usman Ahmad

Department of Computing Science
Umeå University, Sweden
mcs10usd@cs.umu.se

Abstract. Virtualization is widely used in many different areas of computer technology and is very popular in the area of databases. This paper describe the use of databases in a virtualized environment and discuss the problems which are discovered in recent years, such as problems for DBA due to lack of supported tools, live migration of multitenant databases and privacy problem for the DBaaS in the cloud. This paper also describe the solution to these problems as what tools should be used to help the DBA, different low-cost migration techniques and privacy solution for the DBaaS in the cloud. This paper also provides the idea to make the use of database in a virtualized environment more better and easier.

1 Introduction

The paper describes the use of databases in a virtualized environment, the consequent problems which could occur and the possible solutions to these problems. Also what further improvements to be used to make it more stable and efficient for the upcoming technologies (e.g. Cloud Computing). Section 1 gives an overview on the concept of virtualization, tenant (mulitenant) databases, live migration and usage of database as a service (DBaaS) for the cloud. The current problems such as lack of virtualization aware tools for the DBA, live migration for multitenant databases and privacy for DBaaS in the cloud are discussed in Section 2 and the solution to these problems are explained in Section 3. Section 4 present the overall paper and suggestion about the new ideas that can be used for further improvements so as to overcome most of the problems are given in Section 5. Section 6 describes the conclusion and future developments.

1.1 Virtualization

"Virtualization is a way to abstract applications and their underlying components away from the hardware supporting them and present a logical or virtual view of these resources. This logical view may be strikingly different from the physical view" [1]. In a simpler way virtualization can be explained as a server which is exactly same as a physical server but not available in a physical environment. The difference between a physical server and a virtual server is shown in Fig. 1.

Bensch, Drewes, Gulliksson, Mejtoft (Eds.): Umeå's 15th Student Conference in Computing Science USCCS 2012, pp. 1–13, January 2012.

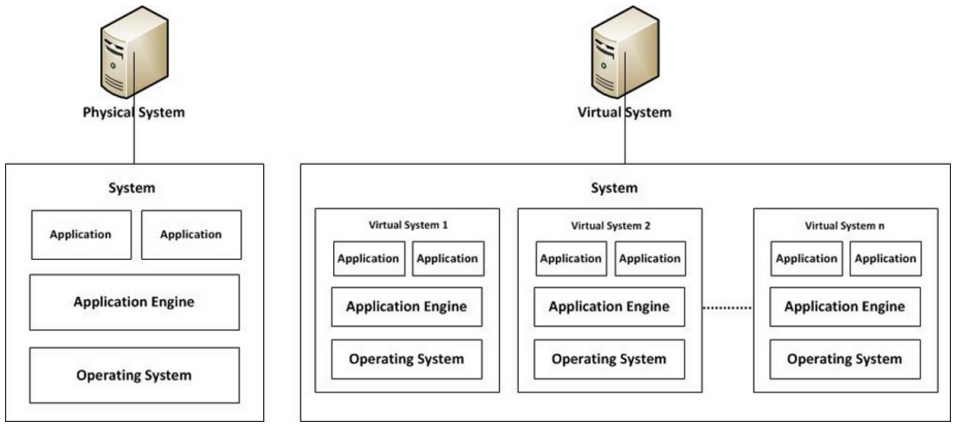


Fig. 1: Difference between Physical and Virtual System; the idea of this image is taken from the source[2]

It is very clear in the figure that in a physical system there is only a single instance of each hardware component is working, but in a virtual system multiple hardware components are working simultaneously on the same physical server.

There are many Virtualization Technologies available but some of them are very popular such as VMware, Microsoft Hyper-V, Xen, Oracle VM etc. Different companies and institutions are using these technologies according to their requirements.

Background: The concept of Virtualization was thought-up in 1960 and was first implemented by IBM [3]. In 1999 VMware introduced the first ever virtualization application which can solve their problems [3].

Growth in the industries are directly proportioned to the demand of additional resources. As IT Industry started to grow, their demands such as additional hardware requirement, increased database storage capacities, maximum resource utilization and the extra cost of maintenance keep increasing which leads to the problems e.g. live migration, privacy for DBaaS in the cloud, additional challenges for the DBA to configure, test and analyze the database.

Many different products have been developed to overcome these problems such as virtualization aware tools for the DBA, different tools, techniques for the live migration (for e.g. Albatross [4], Zephyr [5]) and other commercially available tools introduced to make it more efficient, economical and productive.

Advantages: Virtualization technology comes with various advantages. Few of them are described as below [3]:

- It increases the usage of hardware more efficiently.
- It minimizes the cost of management and need of additional resources.

– It improves the business activities for the companies and establishments.
– It increases the data availability and reduces the processing time.

1.2 Databases

Databases in Virtualizied Environment: Databases (especially traditional databases e.g. RDBMS like Oracle, MySQL, IBM DB2 etc.) are now widely used on virtual servers by many big companies and organizations. Also tenants database systems (*which serves a large number of applications*) on virtualized and even in cloud environments are now widely used which support a large number of applications [4]. "Multitenancy allows effective resource sharing among these tenants" [4], as it joins multiple tenants databases to a single node, which saves a lot of cost and increases performance especially for the large enterprises [4].

A cell (or Tenant cell) is used to represent all the necessary general and important information present to serve a tenant [6].

Database Migration: Database migration means to migrate (move) databases from one server to another server. It is needed to upgrade, test or deploy new systems while keeping the database intact and allowing it to process transaction requested by users with minimum downtime of the service. In paper [6], there are three different forms of migration explained which are listed below.

1. **Asynchronous Migration:** "It is an immediate, blocking migration which relies on a coordinating process to copy the *cell* from a source host to a destination host" [6]. It is an immediate due to prompt action for copying, either by a file or via a backup and restore process. All the traffic is redirected to the destination by the coordinating process when migration gets completed. This form of migration is good for the large *cell*, but it also adds the period of inactivity since it blocks all the client's requests during the migration.
2. **Synchronous Migration:** This form of migration is an event based which do not block the client requests during the migration process. The source and the destination work together as a tightly coupled cluster (i.e. share some or all of the system's memory and I/O resources). If there is any change happens in the source, it also take place in the destination synchronously. Many popular RDBMSs have the ability to run in a master-slave mode in order to efficiently replicate data across hosts in a cluster. A minimal amount of downtime and interruption of service may occur while switching the primary master to the destination.
3. **Live Migration:** Live migration is an immediate, non-blocking (i.e. client requests are not blocked during the migration process) which migrate the *cell* from the host to the destination seamlessly without any downtime and interruption of the service. Live migration is the ideal form of migration but it is the hardest to implement.
   There are two common architecture used in the cloud for the live migration of the databases. The first one is the **Shared Storage Architecture** in

which the persistent image of the database is stored in NAS (*NAS stands for Network Attached Storage is a device which is used to store files*). This device is directly connected with different nodes through which data can be accessed, which is shown in Fig. 2 [7]. Shared Storage Architecture provides high availability and gives best performance results under read environment.

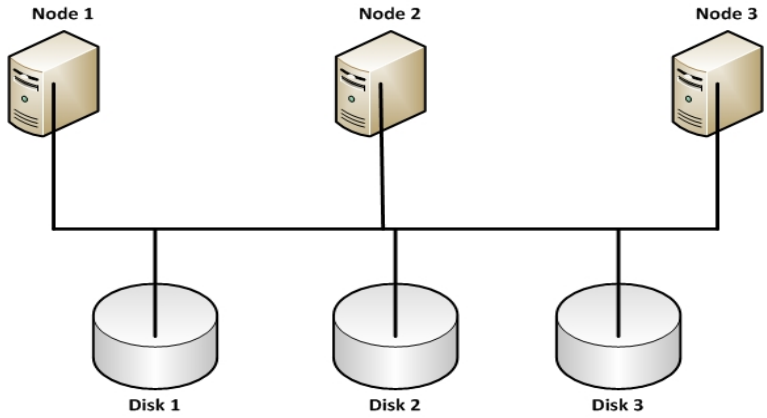


Fig. 2: Shared Storage Architecture; the idea of this image is taken from the source[8]

The second is called **Shared Nothing Architecture** in which the database are partitioned into separate disks. Each of these disks are connected to a node, but the nodes are independent from each other as there is no connection between them. The nodes are connected together with a high speed connection which is also shown in Fig. 3 [5]. Shared Nothing Architecture works well for high-volume databases (both read and write) and provides better environment for scalability.
The use of any one architecture depends upon the technical and business requirement for an application. For example, if there is an application on which scalability is more preferred than consistency then Shared Nothing Architecture would be preferred. Similarly if there is an application which prefers more on the high availability of the data then Shared Storage Architecture would be preferred.

Following are the four cost metrics which are evaluated to find the effectiveness during the migration of the database [4].

(i) **Service unavailability:** The time duration when the service is unavailable.
(ii) **No. of failed requests:** The number of failed requests, failed operations and aborted transactions.

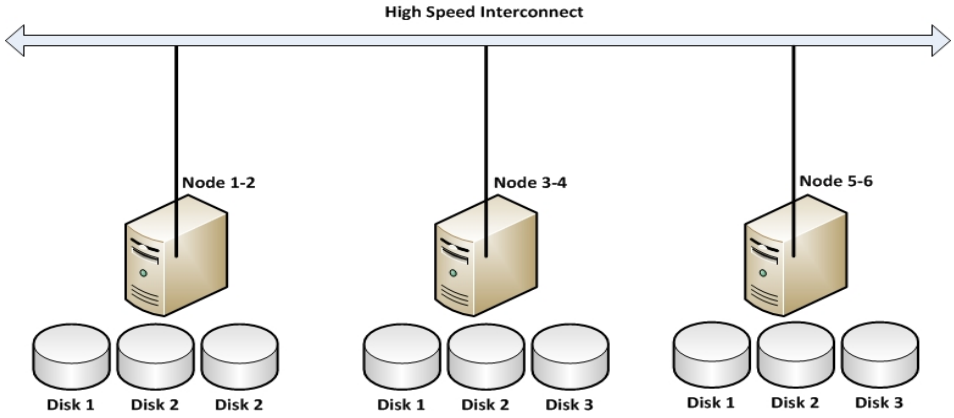


Fig. 3: Shared Nothing Architecture; the idea of this image is taken from the source[8]

(iii) **Impact on response time:** The change in the result time.
(iv) **Data transfer overhead:** The amount of data transferred.

Table 1 describes the overall summary of these three forms of migration [6]

Forms of Migration	**Downtime**	**Service Interruption**	**External Coordination**	**Operation Overhead**	**Migration Overhead**
Asynchronous	Medium	Medium	High	None	High
Synchronous	Low	Low	Medium	Medium	Medium
Live	None	Low	Low	None	Low

Table 1: Forms of migration and the associated costs; the idea of this table is taken from the source[6]

As we can see from the Table 1 that live migration is the most suitable form of migration which can be use to achieve best results. Asynchronous migration is the base line for the live migration but it has too much cost factor. While synchronous migration is in the middle of these two migration.

Database and Cloud: For Cloud Computing technology, database comes in the form of Database as a Service (DBaaS). DBaaS in the cloud comes in the form of service which is available on demand by the user request without any installation and configuration of hardware or software up-front [9]. DBaaS is popular for two reasons. First, it reduces the overall cost of hardware and save significant amount of management overhead. Second, it allow users to pay-per-use which applies to both software licensing and administrative costs [9].

Advantages: Using database in a virtualized environment provides a lot of advantages for the IT Industry. Some of them are listed below [10]:

- It provides high availability.
- It provides a flexible backup or duplication.
- It provides the ability to consolidate server which reduce costs and make the use of available resources more efficiently.
- It provides a better approach for disaster recovery.

2 Current Problems

The problems, listed below are discovered in recent years. These problems are related to the use of databases in a virtualized environment such as configuration, testing, analyzing of database for DBA, live migration and cloud computing.

2.1 Problems for DBA

A **database administrator (DBA)** is a person who is responsible to manage and control both the data and the programs that access those data running in a DBMS [11]. DBA is responsible for user access privileges, security, resource management, overall maintenance etc. [12].

Virtualization helps the DBA to reduce the additional hardware resources and increase the availability with accessibility of the database. But there are still some major problems for DBA which are described below.

- **Invisiblity:** When a database runs on a virtual server it only view as this is the only server. This create the problems for DBA as they are not sure what exactly is going in that virtual server, for example if the virtual server shows the status that the current database is taking 60% CPU power which might not be exactly accurate, as DBA have lack of visibility and the statistics shown by the virtualization tools are not always accurate [13].
- **Resource Managment:** Management of resources play a vital role when running the databases in a virtual server. In a database there are many SQL queries which are very critical and require a good amount of resources. The DBA is responsible for managing the resources, but in a real world scenario there are many cases where running queries on virtual server takes more resources than expected because these queries need to be specifically made and optimized to run on a virtual server. Since DBA does not have the full control of the virtual server [14, 15], due to this it is hard to determine what is causing the queries for having additional resources and this affects the overall performance.
- **Troubleshooting:** Troubleshooting gets complex for the DBA in a virtualized environment as one needs to check what cause the huge amount of CPU usage, why a particular database query keeps crashing under high workloads, why scripts are not giving the same result as they supposed to be etc. Even though DBA uses tools and utilities to monitor and manage databases [14], but these tools cannot tell the difference whether the server is physical or virtual [16].

2.2 Live Migration for Multitenant Databases

In Section 1, we have discussed about the live migration of the databases and see the two most commonly used cloud architectures (i.e. shared storage and shared nothing). For both architectures the multitenancy *shared process model* will be used, as this model provides a good balance of resource sharing, performance, independent schema and better isolation [4, 5]. We will discuss what problem arises when using these architectures.

For Shared Storage: We will now discuss the problems for the Shared Storage Architecture from the paper [4].

When using the traditional straightforward migration techniques for the shared machine multitenancy model where each tenant running its own database instance inside a virtual machine independently has two major drawbacks. First, it creates additional overhead of copying the OS (Operating System) and VM (Virtual Machine) state when migrating the entire virtual machine.

Second, in recent study [9], the comparison between the shared machine model and shared process model shows the shared process model is better in performance especially in the case where VM migration cannot be used to migrate individual tenant database from a database process shared by multiple tenants.

Also most of the traditional DBMS do not support the feature of live migration or not given as the main important feature, due to which a *cell* is migrated by first stopping it from the source node to the destination node of the DBMS. During this migration process all active transitions are aborted and flushing the changes to the NAS device. Once all this is completed the service is resumed at the destination. This approach makes a major impact on the migration cost as the tenant becomes unavailable during this entire operation and all active transactions also get aborted. "Also the entire database cache is lost when the *cell* is restarted at the destination DBMS node" [4], which overall effects the tenant SLA's.

For Shared Nothing: For the live migration of the database in Shared Nothing Architecture, the following two techniques are widely used which are taken from the paper [5].

1. **Stop and Copy:** This is the simplest technique to migrate a database. In this technique, it first stop all the updates in the tenant and makes a checkpoint to the current state which is moved to the persistent storage and then restarts the tenant at the destination. The advantage of using this technique is that it is very simple and efficient in terms of minimizing the amount of data transferred. However, the main drawback of this technique is that it takes too much time. Also it causes service interruption and takes time to warm up the cache at the destination. But this technique is still in use as many current RDBMS (e.g. MySQL, HBase etc.) only support this technique to migrate a tenant to the persistent storage.

2. **Iterative State Replication (ISR):** This technique does not allow the long waiting time and service disruption. "ISR uses an iterative approach where checkpoints are created and iteratively copied" [5]. It creates the checkpoints of the source tenant database and start the migration of the checkpoints to the destination while allowing the database to serve the requests. When the destination start loading the checkpoints, the tenant database are copied iteratively until a small amount of differential changes (which is maintain by the source) remains or the maximum number of iterations are reached. Once this process get completed, a final stop and copy is performed.
   "The iterative copy can either be performed using page level copying or shipping the transaction logs" [5]. Even though this technique does not allow any unavailability of the service and do not take too much time as in the *Stop and copy* technique, but there are some drawbacks of this technique.
   When migrating a database during heavy workload (for e.g. updating a database which make more changes to its state), the ISR can lead to replicate more data being transferred as expected during the migration. Also ISR create multiple checkpoints during the migration so the result is the high disk I/O usage at the source, even though this overhead is minimized by the use of log replay at the destination which will start with a warm cache but it will still effect during heavier workloads.

2.3 Privacy Problems for DBaaS in the Cloud

Privacy allow users to increase their level of trust and can allow them to use their database without any worries. *CryptDB* comes as a sub-system of the DBaaS in the cloud that guarantees the privacy of the stored databases [9]. It allows DBA to perform tasks on the databases without having any visibility to the actual stored data. But this create problems as if a user wants to encrypt all of the databases stored in a cloud, then how the queries will be executed over the encrypted data [9].

For example, if there is a SQL query that ask for a specific student id from student table including the number of credits he passed, which also includes a join with another table or even more complex query that includes more than one join. For such kind of query if the entire database or each record is separately encrypted then the back-end DBMS will not be able to process these kind of queries.

3 Related Work

In Section 2, we have discussed some of the important problems regarding the use of database in a virtualized environment and see how these problems impact the overall system working. Now in this section we will discuss what work has been done so far to overcome these problems:

3.1 Virtualization Aware Database Tools for DBA

There is a need for such tools which are fully aware of the virtualization that allows the DBA to get the visibility to the server without exploiting any security. By having such tools the DBA will be able to solve many problems and be able to improve the overall system performance and stability. Currently there are many commercial tools available which are specifically designed for this kind of purpose[14].

3.2 Better Ways for Live Migration of Multitenant Databases

Albatross for Shared Storage: Albatross provides a technique to overcome the problems for the live migration of a database in Shared Storage Architecture. Albatross allows the migration of the *cell* in shared storage environment with minimal impact on performance of transaction latency and throughput both before and after migration [4].

Albatross focuses on migrating the database cache and the state of active transactions. First the source takes a quick snapshot of a *cells* cache and once that is done the destination warms up its cache with this snapshot [6].

As the destination initializes its cache, the source continues executing transactions at the same time, thus minimizing the impact on transaction latency [4]. "Albatross uses an iterative phase where changes made to the source node's cache are iteratively copied to the destination" [6]. Transactions are blocked once the maximum number of iterations are reached for copying the data and an atomic handover completes the migration process [6]. The state of active transactions which are copied in the final handover phase allows them to resume execution at the destination that already has a warmed cache [4].

By using the Albatross technique, it is guaranteed that:

- The SLA's of the tenants are not violated.
- The serialized execution of the transactions.
- The migration will be fail safe.

Zephyr for Shared Nothing: Zephyr is the first complete solution for the live migration in Shared Nothing Architecture [9]. Zephyr uses the synchronized *dual mode* which allow the source and destination to start the execution of the transaction simultaneously on the *cell* [5]. The migration process start by transferring the meta-data of the tenants to the destination which then start serving the new transactions,while the source completes. All the active transactions which occurred during the start of the migration does not allow any structural changes to the indices [5, 6]. During the start of migration process the source holds all the access (read/write) to the database pages and allow access on demand to the transaction at the destination [5]. A transaction might get aborted by the source node if a page is already migrated which then needs a structural changes to the indices at both nodes [6].

Once all the active transactions are completed by the source node then the migration gets completed by pushing all the pages (with proper read/write access) to the destination [6]. Zephyr guarantees no system downtime, interruption of the service, fail safe migration and avoiding additional replication overhead [5].

3.3 Solution for the Privacy of DBaaS in the Cloud

The solution for the privacy of DBaaS in the cloud is taken from the paper [9].

Privacy can be solved if *adjustable security* be provided to the client as Crypts have different levels of encryption for different types of data. The client should be allowed to use minimum encryption capabilities as needed which will allow minimum decryption time for any query.

It is observed that many different cryptographic techniques can be used to execute SQL queries. For example randomized encryption (RND) provides maximum privacy, deterministic encryption (DET) provides weak privacy, homomorphic encryption (HOM) allows operation over encrypted data, etc.

To main idea to implement the adjustable security is to encrypt all the values in each row independently into an *onion* and start with the most private technique (for e.g. RND). The client has access to the keys for all the onion layers. and make connection to the server by using the database driver (*such as JDBC, ODBC etc.*) and receives the SQL queries. It calculates the onion keys which is needed by the server to decrypt (decrypting higher levels for e.g. DET) specific columns and allow it to execute on the server.

The server decrypt the columns after receiving the keys and write the values on the disk and discard the original ciphertext once the entire column is decrypted. Thus the only overhead for the server is working with expanded tuples. Using this adjustable security technique avoid additional encryption, decryption of the queries and improve overall performance.

4 Discussion

We have discussed about the problems which currently exist in the use of databases in virtual machines, cloud environment and explained the solution to these problems. We discussed how virtualization aware tools can help the DBA to test, configure and analyze the database in virtualized environment. Also Albatross provides better results for the live migration of the multitenant databases in Shared Storage Architecture without any overheads and interruption of service [6]. Similarly Zephyr provides a complete live migration solution for the multitenant databases in Shared Nothing Architecture and avoid the use of traditional techniques (e.g. Stop and Copy, ISR) which are not very efficient for high workloads and takes too much time [5]. The privacy problem for the DBaaS in the cloud environment and how the adjustable security technique provide the solution for the back-end DBMS regarding encrypted data without exploiting any security with minimum performance overhead [9].

5 Suggested Ideas

Most of the Database Management System (DBMS) do not support the feature of live migration or not given as the main important feature [4, 5]. It should be able to support the feature of live migration for all DBMS, especially the Open Source DBMS as the companies would not need to buy commercially available DBMS (that include the support for live migration). This will significantly cut the cost of the project for a company.

Similarly as we have discussed that there are problems for DBA to test, configure and analyze the databases which have been moved from physical server to a virtual server[13, 15]. The developers of the DBMS should add some extra functionalities to give enough information about the virtual system on which database is currently running. This will help the DBA as they do not need to learn special tools and can use the built-in functionalities given in the DBMS. Also this will save the cost for a company to buy such tools.

6 Conclusion and Future Work

The use of the databases in virtualized environments in growing day by day and as cloud computing is getting popular these days, new challenges are arising. For example, current live migration techniques are not supported by all the DBMS. Also the privacy solution for the DBaaS reduce some amount of performance [9]. To make the usage of databases more stable and secure in these environments, these challenges need to be solved.

In recent months there are many new developments to improve the virtualization for databases such as VMware launches new products to improve the live migration and make it very simple [17]. Also different companies are introducing specialized tools for the DBA which can help them in testing, configuring, and analyzing the databases running on a virtual machine. Similarly Oracle has launched its own Oracle VM 3.0 which comes as a whole package that includes the DBMS with full support of virtualization [18]. Presently VMware has launched its product, vFabric Data Director for Enterprise Database as a Service which provides the support of the database on the Cloud for the Enterprise and offers many benefits and features [19].

But these products come with a very expensive cost or need specialized training in order to use them and not all the companies can afford it. There is still a lot of need to improve these problems and providing the solution within Open Source tools and DBMS.

References

1. Kusnetzky, D.: Virtualization: A Manager's Guide. O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472 (2011)
2. Kusnetzky, D.: Virtualization: A Manager's Guide. First edition edn. O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472 (2011)

3. Menken, I., Blokdijk, G.: Virtualization: The Complete Cornerstone Guide to Virtualization Best Practices: Concepts, Terms, and Techniques for Successfully Planning, Implementing and Managing Enterprise It Virtualization Technology. Emereo Pty Ltd. (2008)
4. Das, S., Nishimura, S., Agrawal, D., Abbadi, A.E.: Albatross: Lightweight elasticity in shared storage databases for the cloud using live data migration. PVLDB (2011)
5. Elmore, A.J., Das, S., Agrawal, D., Abbadi, A.E.: Zephyr: live migration in shared nothing databases for elastic cloud platforms. In: SIGMOD Conference. (2011) 301–312
6. Elmore, A.J., Das, S., Agrawal, D., Abbadi, A.E.: Towards an elastic and autonomic multitenant database. In: Proceedings of NetDB 2011, Department of Computer Science, UC Santa Barbara, CA, USA `http://research.microsoft.com/en-us/um/people/srikanth/netdb11/`.
7. Agrawal, D., Abbadi, A.E., Das, S., , Elmore, A.J.: Database scalability, elasticity, and autonomy in the cloud. Technical report, Dept. of Computing Science, University of California at Santa Barbara, Santa Barbara, CA 93106, USA (2011)
8. Lurie, M.: Simulating massively parallel database processing on linux (March 2002) `http://www.ibm.com/developerworks/data/library/techarticle/0203lurie/0203lurie.html`.
9. Curino, C., Jones, E., Popa, R.A., Malviya, N., Wu, E., Madden, S., Balakrishnan, H., Zeldovich, N.: Relational cloud: A database service for the cloud. In: 5th Biennial Conference on Innovative Data Systems Research, Asilomar, CA (January 2011)
10. Khafagy, M.H., Salem, A.A.: The cost of migrating dbms from a conventional environment to virtual machines. In: Proceedings of the 10th WSEAS international conference on Telecommunications and informatics and microelectronics, nanoelectronics, optoelectronics, and WSEAS international conference on Signal processing. TELE-INFO'11/MINO'11/SIP'11, Stevens Point, Wisconsin, USA, World Scientific and Engineering Academy and Society (WSEAS) (2011) 41–46
11. Silberschatz, A., Korth, H., Sudarshan, S.: Database System Concepts. Sixth edition edn. McGraw-Hill, 1221 Avenue of the Americas, New York, NY 10020, USA (2010)
12. Elmasri, R., Navathe, S.B.: Fundamentals of Database Systems. Fourth edition edn. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA (2003)
13. Bergal, D.: Overcoming the challenges of managing databases in virtual environments. `http://www.dbta.com/Articles/Editorial/Trends-and-Applications/Overcoming-the-Challenges-of-Managing-Databases-in-Virtual-Environments-76778.aspx` (July 2011)
14. N.N.: Confio white paper, five trouble spots when moving databases to vmware, a guide for it managers (October 2011) `http://marketo.confio.com/WhitePaper_MonitoringDatabasesOnVMware.html`.
15. Kline, K.: Database - abstracted. `http://www.dbta.com/Articles/Columns/SQL-Server-Drill-Down/Database---Abstracted-78183.aspx` (October 2011)
16. N.N.: Virtualization basics. `http://www.vmware.com/virtualization/virtual-machine.html`
17. N.N.: Virtual machine migration comparison: Vmware vsphere vs. microsoft hyper-v, a test report by principle technologies, commissioned by vmware, inc (October 2011) `http://www.principledtechnologies.com/clients/reports/VMware/vMotion_vs_Live_Migration_1011.pdf`.

18. N.N.: An oracle white paper, oracle vm 3: Application-driven virtualization (August 2011) http://www.oracle.com/us/technologies/virtualization/ovm3-app-driven-459334.pdf.
19. N.N.: Vmware introduces new platform for enterprise database as a service. http://www.dbta.com/Articles/Editorial/News-Flashes/VMware-Introduces-New-Platform-for-Enterprise-Database-as-a-Service-77383.aspx (August 2011)

Evaluating iPad Note Taking During Lectures

Peder Boberg

Department of Applied Physics and Electronics
Umeå University, Sweden
peder.boberg@gmail.com

Abstract. iPads are becoming increasingly popular both for private use and in learning environments. This paper primarily investigates the use of the iPad as a note taking tool. Especially high speed note taking is examined, as in the situation of a lecture. A secondary goal of this paper is to see whether there are any guidelines for what a productive note taking application should contain. The findings in this paper are based on a literature review as well as on a survey sent out to a swedish upper secondary school.
The results of this paper show that it is difficult to find a note taking tool or method that suits each and every one. This sets high requirements on the flexibility of the chosen tool and is in fact one of the advantages with an iPad. When it comes to guidelines for note taking applications, the main finding is that it should be as versatile as possible.

1 Introduction

In the summer of 2011 the total number of iPads sold exceeded 20 million units [1] and educational environments, such as schools and universities, have started to use iPads as part of learning environments [2]. There are, of course, both advantages and disadvantages with implementing this kind of new technology into learning environments. Valstad [3], e.g., states that the iPad has "great pedagogical potential" and Kennedy [4] states that "tablet computers are on their way to being a major part of education - from kindergartners to collage students". However, there are obstacles with implementing iPads as well. An eBook project [2] performed with sixth grade students showed that although the students engagement increased when using the iPad, their comprehension scores decreased.

Several studies have been conducted on topics such as *iPad in educational environments* [1, 3, 5–7] but none of these has focused on taking notes. This topic deserves attention as an iPad is more portable and lightweight than a laptop computer [5, 8], which makes it a more accessible tool to carry around. Another strength of the iPad is its battery life [4, 8], which can last up to ten hours even when watching videos [3]. Apple states that the iPad is "poised to change the learning landscape" [9] as they consider it to be a unique kind of device.

The primary goal of this paper is to examine whether there are sufficiently productive ways of using an iPad for taking notes. In other words, is it possible

Bensch, Drewes, Gulliksson, Mejtoft (Eds.): Umeå's 15th Student Conference in Computing Science USCCS 2012, pp. 15–23, January 2012.

to replace pen and paper with an iPad? Especially high speed note taking is examined, as in the situation of a lecture. There are three main methods of taking notes with an iPad:

- Using the visual keyboard.
- Using a stylus pen.
- Using an external keyboard.

These methods are examined and evaluated based on an extensive literature review and a survey sent out to *Stenforsaskolan*. The survey provides data to investigate how the iPad is being used in different note taking situations today as well as giving the opportunity to compare the different note taking methods.

A secondary goal of this paper is to see whether there are any guidelines for what an effective note taking application should contain. An important fact is that this paper focuses on *note taking* and not *document creation*. This paper considers *note taking* as writing meeting or lecture notes in short form, while *document creation* is more towards writing essays or reports. Thus, the best suited application should not be expected to create complex documents but instead facilitate document creation. Therefore, copying and pasting text and figures between this note taking application and other applications should be quick and simple.

The outline of this paper is divided into four main parts, *Method* explains how the study was conducted, *Literature Review* and *Results* states the results of this study, *Discussion* analyzes and discusses the result and *Conclusions* summarizes the results and presents guidelines based on the discussion.

1.1 Applications

The range of note taking applications on *App Store*, Apple's digital application distribution platform for *iOS*, is extensive and it keeps expanding. A search for *note taking* in early December 2011 resulted in 70 iPad applications with varied functionality, everything from simple note taking applications to word processors to audio recording. The phrase *note taking* gives accurate results as it limits the scope of the search and excludes some annotation and document creation applications. To be able to choose an application the user first has to decide its purpose [10]; should the application just function as a note paper or should it be able to create complex documents? Some primary functionality of note taking applications for educational environments was identified by Valstad [3]:

(V1) Write text quickly.
(V2) Ability to draw sketch, input equation and diagram.
(V3) Offer note organization and searching.
(V4) Automatic backup, perhaps in the 'cloud'.
(V5) Bullets and indentation formatting.

1.2 Other Work

Earlier studies [6] claim that it is easier to take notes using a keyboard in humanity- and social science courses than in mathematics courses because of the difficulty of writing complex formulas quickly. This leads to that mathematic students continue to use pen and paper as their favoured note taking method. A study [5] conducted on geology students showed that after using an iPad for a semester as many as 82% still preferred using pen and paper as they, e.g., found the iPad to slow to annotate and take notes with. An important note regarding that study was that students did not get the opportunity to use a *stylus pen*. The stylus offers both a more familiar note taking method [3, 5] as well as the fact that the user does not have to be restricted by the limitations of a keyboard when it comes to drawing figures and equations [6]. Another study [7] conducted on engineering and computer science students showed that 45% of those who participated benefitted from using a Tablet PC for note taking while as many as 31% said they experienced a negative effect. This vast difference is interesting and shows that people adapt differently to new technology.

Valstad writes about how important it is to use the iPad in situations where it is useful and not to force students into an "awkward usage pattern" [3]. Another way to look at this is how to create a usage pattern that better suits the specific situation. Users seldom want to learn, e.g., a new keyboard layout simply to increase their typing speed [11], this statement probably concerns note taking methodology as well. In other words, the device, or perhaps more suitably the application, should be designed with regard to how we use pen and paper today.

Murphy states that the iPad was primarily introduced as a *content delivery tool* while content generation, and similar functionality, was considered a secondary function [1]. Although it is common that unexpected effects, or usage areas, emerge when a device like this is introduced [1, 3], this view could have made the iPad into a less optimal *content generation tool*.

2 Method

Two kinds of data gathering methods were used in order to investigate the goals of this paper. First, an literature review was carried out and based on the result of this review, a survey was sent to *Stenforsaskolan*. *Stenforsaskolan* is, according to them, the first upper secondary school in Sweden that provides every new student with their own iPad.

The literature review was primarily conducted within the fields of usability and computer science, while the pedagogical and social science aspects were in some extent left out. The most frequently used keywords, which were combined with each other, during the search for relevant literature were *iPad*, *note taking*, *school* and *keyboard*. As mentioned in the introduction, most of the earlier studies of iPads in educational environments focused on its pedagogical qualities. However, some of these [3, 5] included a short section about note taking. The problem with those papers is that they do not distinguish between taking notes and document creation.

2.1 Survey

At the start of the autumn semester of 2011, *Stenforsaskolan* provided each first year student with their own iPad, which makes this group a good reference point for this paper. What makes this particular school even more interesting is that they have been using external keyboards during some, but not all, lectures. This provides the opportunity to compare the different note taking techniques to each other in the specific note taking situation of focus in this paper. The latter part of the survey will cover how the note taking applications is being used and how the students perceive them.

The respondents of this survey are 15-16 years old and social- or natural science students, a total of 22 students answered the survey. As *Stenforsaskolan* is a relatively small school, all students were asked to participate, regardless of education type, which resulted in a response rate of 37%. *Google Form* was used to create the survey due to its simplicity of delivering the survey to the students as well as giving them the opportunity to answer with their iPads. Some of the questions the survey hopefully would answer was:

- Supposing all note taking methods have been tested, which is preferred?
- Why is this method preferred?
- Which note taking application is preferred?
- What is the benefits, and possible weaknesses, with this application?

3 Literature Review

As mentioned in the introduction, a problem with taking notes on an iPad is that it is too slow and might be limited by, e.g., the restrictions of the visual keyboard or an application. A way to become less restricted by these limitations is to use a stylus pen, in addition to offering a more familiar note taking technique it might also attract less attention than writing on a regular keyboard [12]. There are a small number of applications today which supports translation of analog iPad notes, in other words text written in free hand on the screen, to digital text. Huettel [7] has identified some benefits with digital text compared to analog text:

> "However, electronic notes can be more comprehensive and organized ... can be searchable, and can facilitate the integration of concepts across lecture and laboratory sessions when all notes are accessible ... from any location."

Laptop keyboards, e.g. the *Acer Netbook*, have been proven to be more effective, with respect to words per minute, than the visual keyboard on an iPad [13]. However, the participants in that study enjoyed using the iPad and might consider to use it, for shorter notes rather than essay writing, if they had more time to practice. Another study [8] showed quite the opposite, with regard to user satisfaction, while writing with the visual keyboard:

> "Our study participants found that the iPad's greatest shortcoming as a tool for academic work was its keyboard ... they found the soft keyboard to be awkward to use ... and reported difficulty typing efficiently with it. Most students used the keyboard only to annotate texts outside of class, not to take notes in class or to write papers ... Several students expressed a desire to write directly on the screen of the iPad, rather than using a keyboard."

As the last part of the quote implies, neither a stylus nor an external keyboard were evaluated in this study.

Before any kind of Tablet PC was released, some research was conducted on, e.g., the MicroTouch touchscreen. Research on capacitive touch screens like this was conducted in the early 1990s and investigated for instance which angle on the screen that lead to the highest performance rating, with regard to words per minute. One study [14] showed that 30 ° gives the highest performance which, in fact, is twice the angle of Apple's *Smart Cover*, which is a protective cover as well as a keyboard stand [15], folded in the horizontal position. Another study [16], performed on an identical touchscreen, suggested that the users themselves should be able to specify the size of the keyboard. This coheres with other research showing that some users think the visual keyboard on the iPad occupies to much space [11].

4 Results

Among the respondents, 23% stated they did not use their iPad for note taking at all. The comments regarding why they preferred other methods were, e.g., that it is easier, quicker and looks better to write with pen and paper. Some of the participants had not gotten the opportunity to try an external keyboard and therefore would have preferred a laptop computer. Of those who used their iPad for taking notes, a summary of preferred method can be seen in Fig. 1. The most interesting fact with this statistics was that around 30%, of both those who preferred using an external and visual keyboard, had not used an external keyboard yet. Interpreting the comments about this, it is safe to say that more of the participants would have chosen to use an external keyboard given the opportunity.

The participants were also asked to rank the different note taking methods that they used, Fig. 2 shows how they answered. These answers gave an average rating of 4 for the external keyboard, 3.43 for the visual keyboard and 3.33 for free hand note taking on a scale from 1, bad, to 5, good. Worth mentioning is that only one respondent had been using a stylus pen and rated free hand note taking highly.

The respondents of the survey primarily used two different applications for note taking and document creation. It is easy to quickly describe these applications by fitting them into the functionality list in the introduction. Apple's *Pages* is an document creation application and its wide range of functionality supports everyday note taking well [17]. This application fully supports V1 and

V3-V5 but lacks support for free hand writing and therefore do not fully support V2. *Notes* is a built in note taking application which focuses on simplicity. This application provides none of the functionalities V1-V5 as it has no support for free hand writing, sharing is limited to e-mail, organization is limited to date and search and there is only one font without support for bullets and indentation.

Questions regarding note taking applications showed that all of the respondents use *Pages* and 41% use *Notes*. None of the respondents had tried any other application and a majority answered that they used these applications since their teacher recommended them. *Pages* got good comments regarding its organization and mail functionality while some respondents prompted *Notes* as being better suited for short notes. None of the students could think of any disadvantages with the applications.

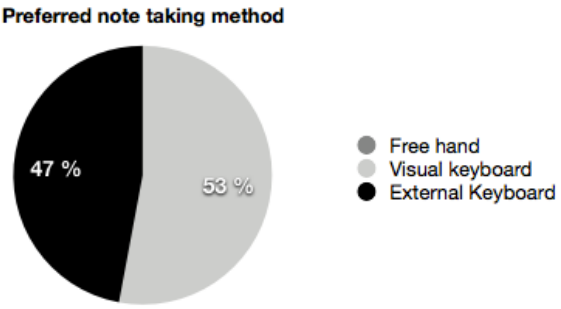


Fig. 1. Preferred note taking method.

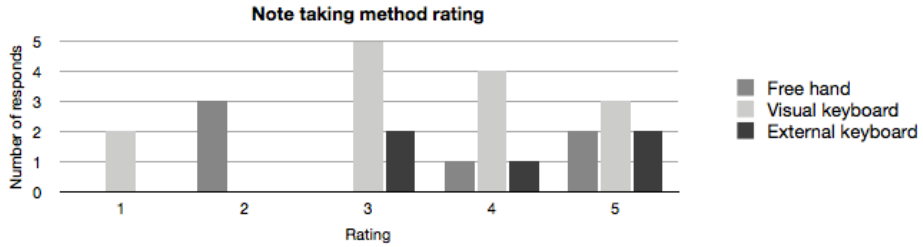


Fig. 2. Rating of note taking methods.

5 Discussion

An interesting result of the survey is that many prefer an external keyboard or even a laptop computer. After reading studies and articles in this area, it is clear

that some considers an iPad with an external keyboard as a laptop computer in two parts [13, 17]. Also, carrying an external keyboard around makes the iPad less portable which is one of the advantages with an iPad compared to a laptop computer. For a school, such as *Stenforsaskolan*, this is not a problem as they intend to have external keyboards in every classroom. However, for a university student this is a disadvantage.

Regarding the situation of a lecture, both the survey sent out *Stenforsaskolan* and a study conducted by Evenstuen, Dalen and Midtbö [5] found the iPad to slow to take notes with in the quick pace of a lecture. However, as earlier mentioned, other studies [5, 6, 8] have found, or suggest, the use of a stylus pen for annotation since it offers a more well known methodology. To be able to take notes quickly with an iPad during lectures it is important that it offers flexibility which, the more familiar technique, pen and paper does. A problem with comparing an iPad for taking notes with pen and paper is that students have been using this latter technique for a long time. Adapting to a new study situation takes time, how long is difficult to say. This makes it hard to draw conclusions about how usable and suitable an iPad is for taking notes. A way to increase flexibility is, as mentioned in the introduction, to let the user specify the size of the visual keyboard.

Murphy's [1] discussion regarding iPad as a *content delivery tool* might have been spot on. As several studies show [5, 6, 8], students enjoy using the iPad for annotating *pdf*-files, preferably with a stylus pen, while note taking is less appreciated or even disliked in other cases [8]. Worth mentioning is that in this paper, annotation is not considered as *content generation* but more towards *content delivery*, as the user is viewing a complete document. Why does an iPad seem to support annotating better than note taking? Perhaps because annotation consist largely of comments and highlighting, which might not demand a high detail level, compared to taking notes. Annotation will probably become increasingly important as more and more material is delivered via internet.

Analyzing Valstad's [3] suggested functionality, V1-V5, for note taking applications and considering the survey did not produce any negative feedback regarding the applications, they still are very relevant. Considering that these were written for the situation of using iPads in a group, they are relevant for the situation of a university student. In the situation of iPads in groups, the possibility of sharing documents within a class should be considered important. This is based on the situation at *Stenforsaskolan* where many of the students thought it was good that they could mail documents to the teacher. If documents were automatically shared, mailing the teacher would be a thing of the past as well as that sharing notes could improve collaboration. This coheres with Huettels [7] quote about digital notes being accessible anywhere.

This study has a few limitations which made it hard to draw conclusions regarding, e.g. the advantages and disadvantages with the different applications. A problem with sending out surveys is that it is hard to formulate questions that results in good answers. An interview or a workshop would probably have made this easier. To get as good result as possible from this study, with respect to

note taking methods, it would have been better if all of the students had tried all of the methods. As mentioned earlier, only one student had used a stylus and about three quarters of the students had not gotten the opportunity to try an external keyboard. The low number of survey answers sadly restricted the possibility of statistically testing the different note taking methods.

Another important issue is that the group of interest for this paper is broader than the student group in the survey. Part of the intended usage group is, e.g., university students that own their iPad and use it exclusively for their own purpose. In other words, they are students who might be the only one in their class using an iPad and evaluates different note taking applications by themselves. A problem with trying to use these results as guidelines for another group of users is that, at least in Sweden, university students have a higher level of personal responsibility than lower grade students. This fact may lead to that university students take more notes as well as demanding more of the application with regard to organization.

6 Conclusions

It is hard to come up with general guidelines regarding the use of an iPad as a note taking tool, since almost each and every one of us take notes in their own way. Although, this requires that a new note taking method should be as flexible as possible, which makes the iPad a good alternative. Compared to hand written notes, digital notes can, as mentioned earlier, be searchable, organized, accessible anywhere and so on. While a laptop is limited to a keyboard, if the owner does not bring an extra tablet board or mouse, the iPad offers more flexibility and portability even if the owner brings an external keyboard and a stylus pen. Both this survey and the literature read, implies that if a user wants to get the most of its iPad a stylus pen is needed and if the user wants to write essays, an external keyboard probably is a wise purchase.

When it comes to the secondary goal of this paper regarding note taking applications, Valstads [3] suggested functionality covers the most aspects. Although, if the application should cohere with the desired flexibility, the user should be able to make some limitations. In other words, the user should be able to strip the application of unwanted functionality. This paper suggests the following guidelines:

(V1*) Write text quickly, regardless of method.
(V2*) Ability to draw sketch.
(V3) Offer note organization and searching.
(V4) Automatic backup, perhaps in the *'cloud'*.
(S1) Switch note taking method easily.
(S2) Facilitate essay writing.
(S3) Limit functionality.

These are very similar to the primary functionality requirements suggested by Valstad except with focus on flexibility. Only V5 was discarded although V1 and

V2 was altered, S1-S3 is new suggested functionality. As some test subject of different studies still prefer using pen and paper, the focus should be on imitating hand written notes. Therefore, the functionalities to create equations, diagrams and bullets should be regarded as secondary and are not as important as those mentioned above. Another functionality worth considering, which exists in some applications today, is the possibility to translate notes drawn with a stylus pen to digital text notes. This, together with a similar functionality which allows browsing stylus drawn notes, could make the iPad into a intelligent note taking tool.

References

1. Murphy, G.D.: Post-pc devices: A summary of early ipad technology adoption in tertiary environments. e-Journal of Business Education and Scholarship of Teaching **5**(1) (2011) 18–32
2. Sheppard, D.: Reading with iPads – the difference makes a difference. Education Today (2011) 12–15
3. Valstad, H.: iPad as a pedagogical device. Master's thesis, Norwegian University of Science and Technology (December 2010)
4. Kennedy, M.: Learning tools. American School and Univerity (June 2011)
5. Evenstuen, A., Dalen, J.T., Midtbö, Ö.H.: The use of iPad as a learning tool. Technical report, Human-Computer Interaction (2010)
6. Romney, C.A.: Tablet pcs in undergraduate mathematics. In: Frontiers in Education Conference (FIE), 2010 IEEE. (October 2010)
7. Huettel, L.G., Forbes, J., Franzoni, L., Malkin, R., Nadeau, J., Nightingale, K., Ybarra, G.A.: Transcending the traditional: Using tablet pcs to enhance engineering and computer science instruction. In: 37th ASEE/IEEE Frontiers in Education Conference. (2007)
8. Marmarelli, T., Ringle, M.: The reed college iPad study. Technical report, The Reed Institute (2011)
9. Apple: http://www.apple.com/education/ipad/ (November 2011)
10. McElhearn, K.: The iPad at work: Taking notes. Macworld (November 2010)
11. Li, F.C.Y., Guy, R.T., Yatani, K., Troung, K.N.: The 1line keyboard: A qwerty layout in a single line. In: ACM Symposium on User Interface Software and Technology (UIST) 2011. (October 2011) 461–470
12. Kissel, J.: Take Control of Working with Your iPad. TidBITS Publishing Inc. (2010)
13. Chapparo, B., Nguyen, B., Phan, M., Smith, A., Teves, J.: Keyboard performance: ipad versus netbook. Usability News **12**(2) (November 2010)
14. Ahlström, B., Lenman, S., Marmolin, T.: Overcoming touchscreen user fatigue by workplace design. In: CHI '92 Posters and short talks of the 1992 SIGCHI conference on Human factors in computing systems. (1992)
15. Apple: http://www.apple.com/ipad/smart-cover/
16. Sears, A., Revis, D., Swatski, J., Crittenden, R., Shneiderman, B.: Investigating touchscreen typing: The effect of keyboard size on typing speed. Technical report, Dept. of CS and HCI Lab, University of Maryland (October 1992)
17. Powers, K.: http://www.viget.com/blog/note-taking-on-the-ipad/ (November 2011)

How to Combine Computer Games with Teaching to Improve the Educational Effectiveness in Adolescents

Anna Enoksson

Department of Computing Science
Umea University, Sweden
id07aen@cs.umu.se

Abstract. Learning is acquiring or modifying knowledge, behaviours and skills. Every individual learns through different ways and there are several models and theories that describe how people learn. To fully embrace the education motivation and commitment is needed from the pupils. Through our technical society, children and adolescents are exposed to the technical environment and by integrating well-known gaming elements in virtual learning tools it is possible to create more motivating and stimulating teaching methods in addition to traditional education. This paper will discuss how it is possible to integrate computer games in teaching to increase the effectiveness of the education and how an educational computer game should be designed in order to be efficient enough to be used in class.

1 Introduction

Learning is the lifelong process of transforming information and experience into knowledge, skills, behaviours, and attitudes [1]. How people learn is described by different learning theories and these are the foundation for today's teaching [2]. For good pedagogics appropriate teaching methods must be employed during education. Fleming's VARK model describes four types of learning techniques that are common to almost all people [3]. According to the VARK model people either learn through images (Visual), talking and listening (Aural), reading and writing (Read/write) or via practical experience (Kinesthetic). Teachers have difficulty adapting the instructions for each individual pupil, therefore pupils should learn how to best gather knowledge during education [4]. Learning is a cognitive, emotional and environmental process [5] which requires a lot of work and effort. To cope with new impressions, motivation and commitment is key. Therefore education should be interesting and stimulating, to make it easier for the pupils.

Today's digital communication technology changes continuously and affect the entire society. The advent of computers changed how people will act in every situation. What would mankind do without the help from computers? During the 19th century the first computer was built in order to analyse numbers [6].

Bensch, Drewes, Gulliksson, Mejtoft (Eds.): Umeå's 15th Student Conference in Computing Science USCCS 2012, pp. 25–38, January 2012.

Today the digital society has grown copiously through the significantly increased access of computers and internet [7]. Many children and adolescents are familiar with the technical environment in that they use it as a leisure activity [8]. So why not combine business with pleasure? Because of their exposure to the characteristic gaming environment these children could benefit from using a more unconventional teaching method like educational computer games. Using games in education can improve knowledge, pupil enjoyment, commitment and interest in the learning process [9, 10]. Studies show that the use of educational games in learning can produce better learning outcomes than non-gaming methods [9–11]. But using commercial computer games in education has both gotten positive and negative criticism. Research findings shows that gaming has a positive effect on spatial abilities and problem solving [12] which are important in technical and artistic contexts. The study also showed that children that were playing violent computer games had tendencies to be more aggressive towards other children. These findings was consistent with another study which in addition showed that playing commercial computer games lead to negative school performance [13].

This paper will look into the question of how it is possible to integrate computer games in adolescents education to increase the effectiveness of teaching and how an educational computer game should be designed in order to be efficient enough to be used during teaching.

The outline of this paper is divided into four parts. Section two introduces learning theories, qualities needed for learning, and today's teaching methods. Section three gives information about computer games and the impact they can have. There is also information about educational computer games and design guidelines to these. This is followed by discussion in section four.

2 Teaching Methods

A teaching method consists of different principles and methods to use for instruction. Traditional teaching methods is where the teacher is seen as the source of knowledge. The knowledge is seen as a product which consists of knowledge material conveyed from the teacher to the pupil and the pupils are often seen as passive recipients of the knowledge. The most commonly used teaching methods include lecturing, demonstration, recitation, class participation and memorization, and these are all based on different learning theories. According to Cooper [2] there are three main categories for learning theories: behaviourism, cognitivism and constructivism [14].

Behaviourism is a philosophical framework that operates on a principle of "stimulus-response". The learner is seen as a clean slate and all behaviour is caused by external stimuli. Both positive and negative reinforcement increases the probability that previous behaviour will repeat itself. In contrast punishment decreases the likelihood of previous behaviour. All behaviour can be explained without the need to consider internal mental states or consciousness. A well known example of behaviourism is Pavlov's dogs [15].

Cognitivism focuses on the learners inner mental activities. The learner is viewed as an information processor and mental processes such as thinking, memory, knowing and problem-solving need to be explored. Cognitivism consider that people are rational beings that require participation in order to learn and that their actions are a consequence of thinking, not merely a response to environmental stimuli as suggested by behaviourism.

Constructivism is a paradigm that states that learning is an active, constructive process. The learner is an information constructor that actively creates their own subjective representations of objective reality. New information is linked to prior knowledge on personal experience and hypotheses of the environment.

These theories attempts to describe how people learn and is a foundation of the teaching methods that are used in classrooms today [14]. The choice of appropriate teaching methods depends on which information is being taught and to whom, since there are different pros and cons with every method and some of them are summarized below [16].

Lecturing is the process of teaching using oral explanations of the subject. It is a powerful tool since the information is presented in a direct manner and is applicable for both small and large groups of pupils. The lecture could be inspiring to the pupils, but if the expert, the source of knowledge is not a good teacher the learning could be absent. The communication is a one way process which makes the pupils passive and not involved.

Collaboration allows pupils to participate in their own learning process by discussions and group assignments. These kind of methods establishes a personal connection between the pupil and the subject and working together with others may help the pupils to think in a more unbiased way. Although pupils are more committed than during a lecture, the group collaboration could lead to getting side tracked.

Individual assignments allows pupils to think for themselves and to not be influenced by others. This could lead to unbiased opinions.

For good pedagogics appropriate teaching methods must be employed for each individual situation. When deciding which teaching method to use the teacher must consider the pupils background, previous knowledge, environment and learning goals [14]. The fact that pupils have different ways of learning must be taken in consideration when choosing teaching methods in order to get motivated and committed pupils in the classroom [3].

2.1 Qualities Essential to Learning

The ability to learn is what made humanity evolve and create the innovative world that is today. There are several approaches of learning, of how to allow an individual to learn best. One of the most commonly used learning styles is the Fleming's VARK model [3] . According to Fleming there are four different types of learners and the VARK model is a tool to measure the input of information, V (visual), A (aural), R (read/write) and K (kinaesthetic).

Visual learners have a tendency to think in images, so the best way for them to learn is usually when information is associated with images, diagrams and techniques. By representing the data spatially pupils are able to focus on meaning and make better use of their visual memory. To increase the efficiency of the education, teachers could think about color, layout and spatial organization. Mind maps and diagrams is examples of how to show information to the pupils.

Aural learners depends on listening and talking as the main way of learning. By listening, discussing and repeating the information the pupil is able to memorize the data. The teacher could include music and rhythm during lectures to help the aural learners during education.

Read/Write pupils prefer to learn from printed text, they often want to rewrite the text that was given to fully understand and memorize the information.

Kinaesthetic people learns best via experience, to actually carrying out the physical activity. Sometimes it can involve the other perceptual modes. In this case the teacher could use physical objects, touch, action and hands-on work in the learning activities.

By making use of various techniques for studying for the different modes the outcome of teaching could be improved. But how should a teacher be able to meet each individual pupils needs? By varying different techniques during class more pupils could be reached out to. According to Bruner [4] the pupils should learn how to best gather knowledge during education, and thereby get a better understanding, analyse and interpret reality. But in order to learn the pupils must have an interest in the learning process and be committed to the education. Ryan and Deci [17] describe two kinds of motivation - intrinsic and extrinsic motivation. Intrinsic motivation occurs when an individual places value on an activity. Then motivation is authentic and the pupil has more interest and confidence in the task, which often leads to better endurance and performance. The pupils need to feel that they have chosen to perform the task themselves, otherwise the task could be seen as a punishment, which reduces independence and thereby also reduce the motivation. Positive feedback could also lead to greater intrinsic motivation. Extrinsic motivation is when other people affect an individuals choice. For example, when a pupil get a homework. The homework must be done, but the pupil might not have the intrinsic motivation to do it. But maybe the pupil have an extrinsic motivation e.g. getting a certain grade, which could motivate the pupil even though he or she might not find the task amusing or interesting. But to be able to perform one need to be focused on a specific task. The ability to concentrate is built on the process that evaluates and grades the expressions to be includes or excluded according to Kadesjo [18]. He claims that pupils need to "address the perception, thoughts and feelings towards the task and exclude all external stimuli and get started on, hold on to and finish the task". Kadesjo consider that there are three factors that control a pupils concentration:

- The pupils knowledge and intellectual level (whether or not the task has any meaning for the pupil)
- Feelings that arise from the task and context (reminds the pupil of success or failure)
- The pupils motivation towards the task.

From these factors it can be concluded that motivation and concentration are strongly linked and together contribute to the effectiveness of learning [18].

2.2 Technology in Education

New technologies will change how and what people learn, there new sources and ways to manipulate information. But Resnick [19] claims that today's education, the technologies are used simply to reinforce outmoded approaches of learning. To fully take advantage of the new technology we must rethink our approaches to learning and education and how technologies is provided for pupils. Instead of using computers for basic literacy and numeric skills Rusten [20] mean that education must help to build higher-order cognitive abilities, strengthen processes of inquiry and enable collaborative problem solving. New approaches and strategies is needed and educational technologies is an important part of this. So can computers in school improve learning? Research show that pupils behaviour and attendance improved, along with their attitude towards themselves and toward learning. Their performance also improved i several ways [21]:

- "Test scores indicated that, despite time spent learning to use the technologies, pupils were performing well and some were clearly performing better than before"
- "The pupils wrote more, more effectively, and with greater fluidity"
- "Some classes finished whole units of study far more quickly than in past years"

The research [21] found that technology in education encouraged the pupils to collaborate more than in traditional classrooms. Another positive outcome was that the pupils motivation to learn grew over time while using computers during education.

3 Computer Games

Today's digital communication technology changes constantly and affect the entire society. Along with computers games were created. The worlds first computer game was "Tennis for two", created by Willy Higinbotham in a laboratory in the United States [22]. The game was made for play and the idea was not that it would be exploited commercially. Two decades later the next game was created, "Pong" was built on the same idea as "Tennis for two" and became the first commercially sold computer game. The first personal computer came first in 1980, when there was very large scale integration on the printed circuit board [6].

A long with the personal computer the game industry escalated. During 1975-1985 "Space Invaders", "Astroids", "Donkey Kong" and "Pac Man" were games that were important for the development [22]. As the computers improved, the games also improved and evolved continuously [8]. As the digital environment spread the access to computers, internet, TV and mobile phones increased significantly [7]. Today there are more children and adolescents than adults who uses the computer for listening to music, watching TV, playing games, chatting and downloading files [8]. These activities plays an important role in their everyday lives and computer- and video games have become a popular leisure activity and is played by both children and adults and is sometimes used as sport and work.

3.1 Impact of Gaming

A game is categorized based on the game play interaction and is defined by a set of game play challenges [22]. Every genre requires different skills from the player, e.g. action games requires quick reflexes, accuracy and timing from the player, compared with adventure games which focuses on problem solving by interacting with other people or the environment. A research study conducted by Lager and Bremberg [12] shows that gaming have a positive effect on spatial abilities, i.e. spatial awareness and the ability to mentally deal with shapes and patterns so-called mental rotation. These skills is important for problem solving and especially in more technical and artistic contexts. The study also showed positive effects of computer gaming on the players reaction time. It was also discovered that the test persons became more aggressive when playing with other children after playing violent games. To prevent certain messages from violent computer games to be directed towards young children many games uses video game classifications. One of classifications is called PEGI (Pan European Game Information) and is used in most European countries. PEGI uses a rating to clearly mark the contents of a video game [23] . The rating is a guide for consumers to judge for whom the game is appropriate for. The system is supported by several major console manufacturers e.g. Sony, Microsoft and Nintendo. The rating is easily understood and consists of a number appearing on the game packaging. The number stands for what age the game is recommended from. PEGI also uses descriptive symbols on the games back of the packaging to show why a game has a certain rating. These symbols describe the contents of the game and stands for violence, bad language, fear, drugs, sexual, discrimination, gambling and online game play with other people. This classification method can help parents draw conclusions about how appropriate a particular game is for their child and prevent consequences like aggression tendencies. Hastings study [13] on young children's game use related to behaviour and school performance shows that the relationship between game playing and outcomes is that a more amount of gaming lead to behaviour problems, like aggression, attention problems, delinquency and external behaviour, and negative school performance. Educational games on the other hand was associated with positive outcomes [13]. Gaming does not only effect the mental health but also the physical. Studies indicates that playing computer-/video games is related to children's weight [24]. Children

with higher weight also chose to participate in more sedentary activities than children with lower weight. But this is assumed to be a consequence of that children that are heavier is more likely to turn to computer games instead of other more active activities.

3.2 Using Games in Education

When using computer games in education they can be seen as a tool with which children can get educational outcomes in a well-known characteristic gaming environment. Since children growing up today already is exposed to a society depending on digital technology, they can benefit from these kinds of educational environments and it is a good way to combine education and entertainment, so called edutainment, to make education more enjoying [25]. But along with computer games are being adapted to the education they need to be more precise in the balance between entertainment and the educational needs. According to Egenfeldt-Nielsen edutainment is mostly inspired by behaviourism and to a lesser degree cognitivism and constructivism [25], and the different theories has different view of using games in education:

Behaviourism claims that children learn by practising skills and contents through reinforcements and conditioning and that edutainment is straightforward information delivered to the user using a simple computer game. Through practice children will learn the correct response to a certain stimuli.

Cognitivism think that the learner is the center of attention. The theory claims that educational computer games must present information in appropriate ways for every specific learner and that the game is a tool for facilitating perception in order to achieve meaningful learning.

Constructionism has a strong focus on the learner while involving the setting. The paradigm looks at the broader process of educational use of computer games, and is less interested is the actual content of the educational game and focuses more on the pupils engagement while interaction with the game.

There has been many studies about using commercial computer games as a digital learning tool during education [26]. The findings concludes that the motivational significance in games is important, but the core of the game is critical to create an intrinsic relationship with the learning content of the game. It was noticed that intrinsic integration in educational games have benefits in both motivation and learning outcomes [26]. A study by Sandford, Ulicsak, Facer and Rudd shows that using games in a meaningful way within lessons depended far more on the effective use of existing teaching skills than it did on the development of any new, game-related skills [27]. Far from being sidelined, teachers were required to take a central role in scaffolding and supporting pupils learning through games.

3.3 Educational Computer Games

Educational computer games are considered to be a type of more serious game than regular computer games and this is because the goal of the game is not pure

entertainment, there is also an aim for educational outcomes. Making learning fun by using edutainment as learning tools, helps to motivate pupils to study [9]. Learning from a computer game compared to a similar non-application teaching method is more effective in promoting pupils knowledge and is more motivating than the non-gaming approach [9–11]. Papastergiou concluded that educational computer games can be exploited as learning environments, given that they can considerably improve both knowledge of the embedded subject matter and pupil enjoyment, engagement and interest in the learning process [9].

Moreno-Ger [28] claims that video games in virtual educational environments are a complement to traditional teaching for the pupil. With digital learning tools as educational games teachers may assist pupils on social aspects such as knowledge-based communication, effective interpersonal skills and critical learning. According to Wiklund [29] educational games is not meant to replace the teacher, it is just a tool that facilitate learning. A study made by Virvou, Katsionis and Manos [30] shows how educational computer games affected the pupils performance. The talented pupils did not increase their performance as much as the pupils who were not considered so good in school. It may be that those pupils in most cases are not interested in the school subjects and therefore are not motivated to learn. Because of this unconventional methods, e.g. educational game, could be a good way to make poorer pupils to become more interested in their education [30].

Educational games for children is often simple games which cover a few concepts. There are several types of games: online games, computer games and games for hand-held devices. Studies shows that children is particularly demanding regarding multimedia and denotes that pupils expects to find the same elements in the educational games that they encountered in games they used on their spare time [9]. So given that educational computer games is a good tool for learning it is necessary to understand how a successful educational game should be designed on a pedagogic framework and created to be a motivating gaming environment [31].

3.4 Designing Educational Computer Games

One of the reasons that certain commercial computer games are so popular is because they teach the player how to play the game in an effective way [32]. There are several guidelines on how to design a good game and the points in the guidelines are similar to a great extent [33, **?**,**?**,**?**]. According to Linehan, Kirman, Lawson and Chan [34] the following features appears important to most commercial games:

- "Games typically present the player with a series of short, medium and long-term goals. To have goals of different levels help motivate learners to continue playing"
- "Games typically require the player to take some actions or decisions in order to reach the goals in the game"

- "Games typically excel at providing immediate, appropriate and specific feedback to players. This feature is at the heart of the motivation, sustained attention, learning and fun experienced by game players"
- "Games often have a complex system for presenting players with rewards for achievement"
- "Games methodically teach players the skills needed to meet complex challenges. Long, complex tasks are broken down into short, simple components. These components are trained individually before being chained together"
- "Generally, players are expected to demonstrate excellent performance of a skill before they can advance to using that skill in a more challenging environment. Complex tasks, then, simply require the chaining together of these previously learned simple skills"
- "Where games present the player with options for taking action, no one action should be obviously correct, while others are obviously incorrect"

These features have been observed in several commercial games and thereby implies that they could be useful if employed in educational games as well [34].

And in the same way, educational games should incorporate easy, understandable elements that instructs the user how to play the game so it can be used as a successful tool in virtual learning. A reason that games are proposed to be a good educational tool is because of the way they motivate the users [9–11]. But only turning a school book into a computer programme is not good enough. Researchers have found that learning outcomes from educational computer games are affected by the instructional methods and strategies that are implemented in the game and not necessarily the game itself [31]. To integrate game learning and domain learning in an educational game, the game designer and instructional designer need to work closely during the design process. In order to design educational games an educational framework need to be adopted. This framework should then be used together with the goals of game design. According to Wainess, Kerr and Koenig [35] instructional designers must determine which instructional methods and strategies are best suited for the school subject which the game designer must use to build a efficient and attractive game environment. To achieve integration between instruction and domain instruction, the same set of instructional methods must be used for learning the game as well as for learning the domain. The instructional methods should use the games properties to adapt the performance to each individual player and to deliver specific feedback to motivate the players to continue playing and thereby learn more.

Just as for commercial games, there are many different design guidelines for educational games. Designing educational games is based on how to design regular games but also focus on learning. Pereira and Roque [36] used gaming and simulation design principles as an inspiration for how to create design guidelines for educational games and then used these designing a game, called "The Living Forest", with educational outcomes. Kiili presents an experiential gaming model that can be used to design educational games [37]. The model is created to link game play with experiential learning in order to facilitate flow experience. It is

descriptive and explains how the player uses the game and what a designer should create for the player to get the best possible experience. The model suggest that there are some specific points that are important while designing educational computer games:

- Sustain the motivation and engagement of the player by pumping appropriate challenges to him or her.
- Give the player an opportunity to think of innovative solutions to the given problem.
- Game should be usable and provide clear goals and appropriate feedback to the player in order to facilitate flow experience.
- The game should provide a player with challenges that are matched to his or her skill level in order to increase the likelihood of experience flow.
- Adaptation in games should be transparent to the player in order to ensure that the player does not change his or her normal behaviour in the game world
- Adaptation should not change the user interface of the game
- Bad usability, inappropriate challenges and objects that break the harmony in the game world decrease the likelihood of experience flow.

Beyond these points Kiili means that the game needs to be created with an engaging storyline, appropriate graphics and sounds, and game balance [37]. Storytelling integrates the challenges into a larger task and engages the player to continue with the story. Game balance provides that the game is fun and engaging, and consistent and fair at the same time. And by optimizing the cognitive load for the player, by presenting information in an appropriate way, he/she can assimilate the learning outcomes of the game in the best way possible.

Most of the existing design guidelines for educational games share the same thoughts on what is important for games used in classrooms. In summary they tend to focus on clear goals, understandable interactions, feedback on actions, motivating challenges and have an underlying meaning in the form of teaching subjects [9–11, 31, 35–37].

4 Discussion

New technologies in the digital environment has given opportunities we never seen before. The digital society affects us and our everyday lives radically. The younger generation has grown up with technology and rely on them to be around. Children are more accustomed to the digital culture and often uses it on their free time. This could mean that the younger generation is ready to use technology more in their everyday lives, also in the classrooms. There are many advantages of using technology during education, but it is important to focus on learning and not the tools that are being used as teaching aids. By integrating well-known gaming elements in virtual learning tools, as discussed in earlier chapters, more motivating and stimulating teaching methods are being created.

Using computer games as a virtual teaching aid could be a good way to reach out to more pupils. Everyone is different and learns in different ways. Fleming's VARK model describes the different ways of learning, visual, aural, read/write and kinaesthetic. Computer games might be a good way to combine the different models. By using different elements from the various models or to specialize games for a specific model, it is possible to create tools that are applicable and useful for many users. For visual learners games could be designed to present information through images and animations. Perhaps this could be used to help the pupil to rehearse and practice for examinations by bringing together the right information with the right image. There are many ways of using images for teaching if they are a part of a computer game. For aural learners games could be used for listening and repeating. This may be a good way to rehearse before an examination. For read/write learners the game could let the pupils read along and select the words that needs to be heard spoken. And there are numerous ways of incorporating writing in computer games. E.g. there could be gaps in a text that need to be filled in or maybe a texting competition with a high score. For kinaesthetic learners games could be developed to perform practical exercises. The game could perhaps be used as an example of how it works in real life. E.g. in history class, a game like "Age of Empires 3" could be used to inform the pupil of how people lived in other centuries and cultures.

Teachers should think carefully about the games that are going to be used in education and how they are going to be used for best possible outcome. By incorporating computer games in the regular education there can be cycles of increased motivations of learning. The teacher should have a supervising role and be a source of knowledge so that he/she can help if needed and also deepen the pupils knowledge. The game could be seen as a way to practice the pupils skills or to be another input source of information. Since computer games motivates the pupils to learn it is a great compliment to the regular education. So what are the differences between using educational vs. Commercial games during education? An educational game like the ones imagined above has a more serious goal that an usual commercial game would have. Since the focus lies on the subject to be taught and not to be pure entertainment, like a commercial game is, it is appropriate to use for younger children. Educational games are designed to be simple and with few impressions which makes the children embrace the key elements of the game. These types of games does not have a long storyline, which could be seen as more suitable for classes. Commercial games on the other hand could be suitable to use for specific school subjects. The pupils can learn, E.g. history, in a more practical and fun way than reading about it in a textbook. These often has a long story which leads to greater playing time which could be more appropriate for the pupils free time. These kinds of games has a lot of impressions which can affect the pupils in both negative and positive ways, as discussed in an earlier chapter. This suggests that commercial games are more suitable for adolescents, which can better absorb the underlying sentences of the games.

How should educational games be designed for best positive learning outcome? The research in earlier chapters suggests that educational games should be adapted to the guidelines for commercial games, but with a strong focus on the subject to be taught and the interaction between the player and the game. An important aspect of game design is to standardize. With the technical environment that is today there are several platforms to make use of. But in order to reach as many as possible, both schools and pupils, the games need to be cross-platform to work on different digital units. Another crucial aspect of game design is the balance between entertainment and learning. This balance is essential for the game, too much entertainment affects the learning outcomes and too much learning affects the players perception of the game. To summarize previous research, clear goals, feedback, consistency, and competition are key game characteristics in order to increase motivation and have been found significant in game development.

According to studies on the use of computer games in school, talented pupils did not increase their performance as much as the pupils who were not considered so good in school. It may be that those pupils in most cases are not interested in the school subjects and therefore are not motivated to learn. Because of this, unconventional methods like computer games could be a good way to make poorer pupils to become more interested in their education. Both commercial and educational games have been shown to increase motivation and the willing to learn. Games could also be used at home to help enhance the skills needed for success in school and to keep learning after school hours. This article finds that it would facilitate pupils to use computer games during education, but there should always be a focus on the actual content of the game, learning and teaching.

References

1. Cobb, J.: A definition of learning. http://www.missiontolearn.com/2009/05/definition-of-learning/ (December 2011)
2. Cooper, P.A.: Paradigm shifts in designing instruction:from behaviourism to cognitivism. Educational Technology **33**(5) (1993) 12–19
3. Fleming, N.: Teaching and learning styles: Vark strategies. (2001)
4. Bruner, J.: The process of education. (1977)
5. Illeris, K.: Three dimensions of learning. (2004)
6. Wikipedia: Computer. http://en.wikipedia.org/wiki/Computer (December 2011)
7. InternetWorldStats: Usage and population statistics. http://www.internetworldstats.com/stats.htm (December 2011)
8. Ungdomsstyrelsen: New game, om unga och datorspel. Ungdomsstyrelsens skrifter **2** (2005)
9. Papastergiou, M.: Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation. Computers and Education **52** (2009)
10. Ke, F., Grabowski, B.: Game playing for maths learning: Cooperative or not? British Journal of Educational Technology **38**(2) (2007) 249–259
11. Liu, T.Y., Chu, Y.L.: Using ubiquitous games in an english listening and speaking course: Impact on learning outcomes and motivation. Computers and Education **55** (2010) 630–643

12. A. Lager, S.B.: Halsoeffekter av tv- och datorspelande. Statens Folkhalsoinstitut **18** (2005)
13. Hastings, E.: Young children's video/computer game use: Relations with school performance and behavior. Issues in Mental Health Nursing (30) (2009) 638–649
14. learningtheories.com: http://www.learning-theories.com/. (December 2011)
15. Pavlov, I.: Conditional reflexes: an investigation of the physiological activity of the cerebral cortex. (1927)
16. McCarthy, P.: Getting the most out of your aids/hiv trainings. East Bay AIDS Education Training Center (1992)
17. Ryan, R., Deci, E.: Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist. **55**(1) (2000) 68–78
18. Kadesjo, B.: Barn med koncentrationssvarigheter. Stockholm: Liber (2002)
19. Resnick, M.: Rethinking learning in the digital age. The Global Information Technology Report 2001–2002: Readiness for the Networked World (GITR) (2002) 32–37
20. Rusten, E.: Using computers in schools. Digital Opportunities for Development 207–247
21. Apple Computer, I.: Changing the conversation about teaching and learning and technology. (1996)
22. N.N.: Pc game (December 2011) Wikipedia, http://en.wikipedia.org/wiki/PCgame.
23. PEGI: http://www.pegi.info/en/. (December 2011)
24. Vandewater, E., Shim, M., Caplovitz, A.: Linking obesity and activity level with children's television and video game use. Journal of Adolescence **27** (2004) 71–85
25. Egenfeldt-Nielsen, S.: Making sweet music: The educational use of computer games. (2009)
26. Habgood, M., Ainsworth, S.: Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. Journal of the Learning Science **20**(2) (2011) 169–206
27. Sandford, R., Ulicsak, M., Facer, K., Rudd, T.: Teaching with games: Using commercial off-the-shelf computer games in formal education. Futurelab (2006)
28. Morenoger: Educational game design for online education. Computers in Human Behavior **24**(6) (2008)
29. Wiklund, M.: http://dsv.su.se/forskning/datorspel/wiklund. (December 2011)
30. Virvou, M., Katsionis, G., Manos, K.: Combining software games with education: Evaluation of its educational effectiveness. Educational Technology and Society **8**(2) (2005) 54–65
31. Garris, R., Ahlers, R., Driskell, J.: Games, motivation, and learning: A research and practice model. Simulation and Gaming **33**(4) (2002) 441–467
32. Kekeljevic, I.: How to make addictive game. http://apps-on-android.com/ (April 2011)
33. Desurivive, H., Wiberg, C.: Game usability heuristics (play) for evaluating and designing better games: The next iteration. OCSC '09 Proceedings of the 3d International Conference on Online Communities and Social Computing: Held as Part of HCI International 2009 (2009)
34. Chan, G.G., Linehan, C., Kirman, B., Lawson, S.: Practical, appropriate, empirically-validated guidelines for designing educational games. CHI 2011 - Session: Learnin (5 2011)

35. Wainess, R., Kerr, D., Koenig, A.: Improving the way we design games for learning by examining how popular video games teach. Technical report, Graduate School of Educat ion and Information Sciences (7 2011)
36. Pereira, L., Roque, L.: Design guidelines for learning games: the living forest game design case. Proceedings of DiGRA 200 (2009)
37. Kristian, K.: Digital game-based learning: Towards an experiential gaming model. Internet and Higher Education **8** (2005) 13–24

Sequential Learning From Demonstration Based On Semantic Networks

Benjamin Fonooni

Department of Computing Science
Umeå University, Sweden
fonooni@cs.umu.se

Abstract. Most of the humans day to day tasks include sequences of actions that lead to a desired goal. In domains which humans are replaced by robots, the ability of learning new skills easy and fast plays an important role. The aim of this research paper is to incorporate sequential learning into Learning from Demonstration (LfD) in an architecture which mainly focuses on high-level representation of behaviors. The primary goal of the research is to investigate the possibility of utilizing Semantic Networks in order to enable the robot to learn new skills in sequences.

1 Introduction

Learning from Demonstration (LfD) is a technique that allows robots to extend their capabilities by observing human or robot teacher performing sequences of actions. For instance, teaching a robot how to find an object, go towards it and push it to the desired location can be time consuming and require programming skills. In case of LfD, even a non-roboticist tutor is able to teach such a behavior. This method is inspired by humans and animals natural learning ability which is more intuitive rather than explicit programming [1]. Most of the behaviors can be divided into low and high level representations. The low-level representation of a behavior is regarded as a set of sensory-motor events that form an action. The high-level representation of a behavior is formed based on connection of the concepts that are represented by labeled graphs or Semantic Networks. During the past years, several LfD algorithms have been proposed [2–4] which mainly focused on low-level representation of behaviors. In this paper we are focusing on the other aspect of learning behaviors from demonstration which less explored by the other researchers. Our goal is to investigate high-level representation of behaviors and find the solution for integrating it with low-level representation. Therefore, the approach proposed here is not an alternative for any LfD algorithm band should be considered as a complement.

In our previous work, we aimed for solving one of the well-known problems in LfD, namely how to generalize a demonstrated behavior such that it can be performed also in new, previously unseen situations [5]. The purpose was to introduce a technique that integrates high and low-level learning and control in

Bensch, Drewes, Gulliksson, Mejtoft (Eds.): Umeå's 15th Student Conference in Computing Science USCCS 2012, pp. 39–47, January 2012.

a way that supports generalization. At the end, we showed the learning results of our novel technique with the real robot and in the simulated environment.

The high-level controller deals with concepts represented and processed in Semantic Networks. This controller is interfaced to a low-level controller that learns and performs behaviors defined at the sensory-motor level. The connector, interfacing the two levels, is learned *contexts*, describing behaviors using high-level concepts within the Semantic Networks. A context is a definition of necessary conditions for a low-level behavior to be performed. Therefore, in this research our robot is learning contexts that are connected to the behaviors which tutor demonstrated. Learning low-level representation of the behavior is also an important task, but we do not investigate it in this paper.

Sequential behaviors are the necessity to intelligence and an inseparable part of human daily activities. One of the most prevalent forms of human and animal learning is sequence learning [6]. Normally, a demonstration consists of several behaviors performed in sequence by the teacher. Sequencing refers to the arrangement of behaviors into a sequence. The robot should recognize demonstrated behaviors and connect them to a single sequence. The main topic of this paper is to propose a method based on our previous architecture and Semantic Networks [5] in order to sequentially teach the robot new behaviors controlled by environmental conditions. This enables the robot to later recognize the sequences and execute the sequenced behaviors.

2 Semantic Networks

Semantic Networks are often used to represent abstract knowledge in a human-like fashion. In robotics, Semantic Networks can be used for concept forming and situational awareness [7]. The structured way of representing knowledge can in combination with visualization tools [8] help humans to understand the internal state of the robot and what is happening in the robot's cognitive system. This may for instance help a tutor to put the robot back on track when it is distracted during learning or performing phases. In our usage of Semantic Networks, high-level concepts such as object types (Place, Furniture, ...) and properties (Color, ...) are represented as nodes while relations between concepts are represented as links. The initial Semantic Network is pre-defined. Nodes are activated through perception of their corresponding object, person or location in the environment. This will be done by the perception unit [5] which is responsible for collecting information from various sensors. Figure 1 depicts the initial Semantic Network which is used during learning and performing phases.

A common reason for using a Semantic Network as a model of the environment is its ability to generalize [9]. For instance, after a demonstration in LfD, the robot will be able to extend the learned context to other, related, contexts. Assume for instance that the robot learns how to clean the table if there are empty cups on it. By generalizing the cup concept to all the drink wares, it will also perform the cleaning behavior when perceiving a mug on the table. The generalization is done by spreading and decaying activation which are funda-

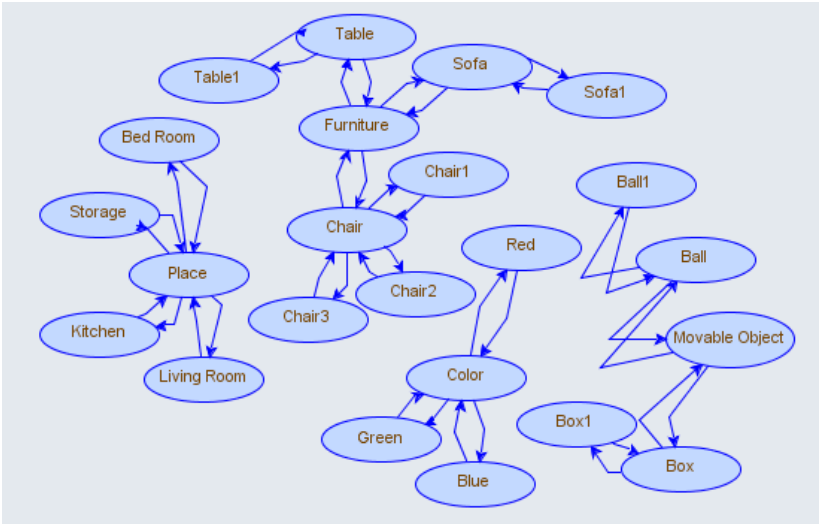


Fig. 1. Sample of Initial Semantic Network.

mental functions in Semantic Networks [10]. In our approach, each node has an activation level; therefore, *Spreading* can be defined as a mechanism by which activation spreads from one node to another in proportion to the strength of their connection. The strength is determined by the weight value obtained by the learning algorithm. *Decaying* is defined as a mechanism by which the activation levels of nodes gradually decrease over time.

3 The Proposed Approach

For simplicity, we assume that all the required low-level behaviors are pre-programmed and callable from the high-level controller. Therefore, no learning for low-level controller is required and focus is only on the high-level representation of the behavior. In the following sections, learning and performing mechanisms are explained and a technique for weight calculation is elaborated.

3.1 Learning Phase

Learning will be started by tele-operating the robot, observing the environment and forming a context. The predefined Semantic Network shown in Fig 1 is used as a base and a new context node *(Move Object)* which robot starts to learn, will be added. The whole demonstration will be performed by the tutor at once, while segmenting the behavior into sub-behaviors and determining start and end of each one for the robot. Thus, termination of the first sub-behavior leads to the start of second one. Therefore, each sub-behavior node that represents each segment of the behavior, will be added and connected to the context node.

Furthermore, sub-behavior nodes are connected to each other from subsequent to the preceding ones with the weight value equal to -1. This assures that by activating the subsequent sub-behavior node, the preceding is disabled. Figure 2 depicts a sample of learned network.
In addition to context and sub-behavior nodes, set of other nodes which represent objects, persons or places in the environment are exist in the network. These nodes are activated by various sensors from the perception unit. All activated nodes are linked to the corresponding sub-behavior node with a weight value that can be obtained by several learning modes. In our previous work, we introduced *Novelty Detection* technique [5]. In this paper another approach, *Multiple Demonstrations*, will be explained in section 3.1.1. Arrows depicted in Fig 2 show direction of activation spreading through their connected nodes. Thicker arrows have higher weight values which result in transferring more energy from one node to another. Each node has an energy parameter that limits the number of link levels for spreading and controls the degree of generalization.

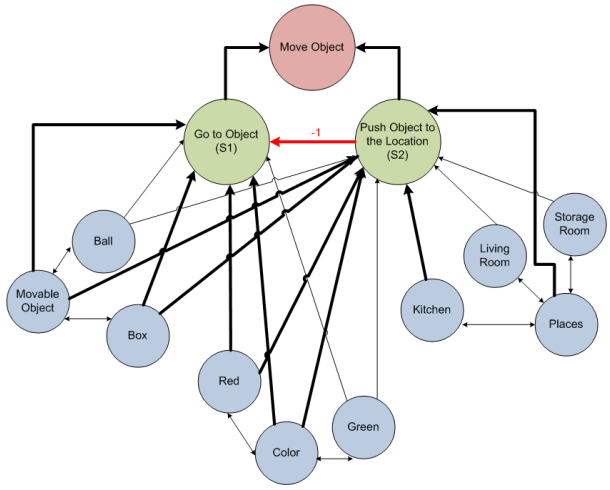


Fig. 2. A Sample of Learned Network.

As an example, assume that the tutor wants to teach the robot how to move a red box to the kitchen. Therefore, there should be a *Move Object* behavior that has two sub-behaviors. The first sub-behavior is *Go to the Red Box* and the second one is *Push the Red Box to the Kitchen*. As mentioned earlier, Semantic Networks have the ability to generalize one concept to another. Thus, by generalizing *Go to the Red Box* behavior to *Go to the Object (S1)* and *Push the Red Box to the Kitchen* behavior to *Push the Object to the Location (S2)*, there is no need to demonstrate the same behavior with different objects or locations. In this research, we assume that the tutor is responsible for determining start and end of each sub-behavior. The tutor indicates start of the first sub-behavior *(S1)*

with a designed user interface and executes tele-operation. After reaching the red box, the tutor indicates start of the second sub-behavior *(S2)* and continues the demonstration with pushing the box to the kitchen. As shown in Fig 2, at the end of the demonstration and terminating the second sub-behavior, both sub-behavior nodes will be connected to the *Move Object* context node. Meanwhile, the nodes that were active during each segment of the behavior will be connected to their respective sub-behavior nodes.
As mentioned before, a single directional link from *Push Object to the Location* node to *Go to the Object* node will be established and weight value equal to -1 will be assigned. The more strengthened links have higher weight values, meaning that by activation of respective nodes, the chance of activating sub-behavior nodes is relatively higher.
At the end of demonstration, the last sub-behavior node forms and the learning phase terminates.

3.1.1 Multiple Demonstrations Algorithm One of the important tasks of the learning mechanism is to obtain the weights for each connection between the nodes. In our previous work, *Novelty Detection* technique introduced and tested [5]. Our novel approach is *Multiple Demonstrations* that has similarities to the *Novelty Detection* but with changes in number of demonstrations and the way each eliminates the irrelevant nodes from the network. In *Novelty Detection*, system checks for the significant changes in the environment, but in *Multiple Demonstrations* system checks for significant invariances.
The approach is to read the sensor values and sample the activation levels of all activated nodes at a given frequency while demonstrating the behavior. At the end of each demonstration, the learned network is stored and labeled as same as the context. Statistical tests will be run to determine whether the data for the same node in different sets (demonstrations) are from the same distribution or not. In case of having two demonstrations, Unpaired T-Test and for more than three demonstrations, One Way ANOVA test will be run [11]. In this section, formulation for T-Test is described.
The purpose of running t-test is to compare mean node activation of all nodes.

$$t_x = \frac{\mu_{A_{1x}} - \mu_{A_{2x}}}{\sqrt{\frac{Var_{1x}}{n_1} + \frac{Var_{2x}}{n_2}}}. \tag{1}$$

where
$\mu_{A_{1x}}$ is mean activation of node x in the first demonstration
$\mu_{A_{2x}}$ is mean activation of node x in the second demonstration
n_1 and n_2 are number of samples for first and second demonstrations
t_x tells whether the samples for the two nodes are drawn from the same distribution or not. In other words: did the node change significantly between two demonstrations. If it did not, the connection between the node and the context node should be disconnected.
Confidence Interval (CI) of the test is given by the t-distribution with α value set to 0.05. Degree of Freedom (DF) is calculated as follows:

$$DF = (n_1 + n_2) - 2. \tag{2}$$

According to equation 1, t_x will be computed and nodes which fulfill equation 3 remain connected.

$$-CI \leq t_x \leq CI. \tag{3}$$

After disconnection of the irrelevant nodes, weight values for each remained node should be calculated:

$$w_x = \frac{N_x \mu_{A_x}}{P}. \tag{4}$$

where N_x is the number of samples for which node x has activation value above 0 during the learning phase, μ_{A_x} is the mean activation of node x in both demonstrations. P is the weighted sum for all nodes, calculated as follows:

$$P = \sum N_i \mu_{A_i}. \tag{5}$$

Finally, the learned sub-behavior nodes will be connected to the context node. As mentioned earlier, *Multiple Demonstrations* and *Novelty Detection* techniques have features in common. Determining which technique is suitable mostly depends on the learning scenario. Also, number of demonstrations (datasets) is important while choosing the best technique. The main difference is about the approach they disconnect the irrelevant nodes.
Multiple Demonstrations technique is well-suited for learning behaviors that sufficient presence of an object during demonstrations is necessary for connecting the concept of that object with the context. At the other side, in the dynamic and highly changing environment, *Multiple Demonstrations* may not work properly. Even though, increasing number of demonstrations can solve parts of the problem, but is not the best solution. Therefore, other learning modes like *Novelty Detection* that looks for the changes in the environment is more suitable choice.

3.2 Performing Phase

After the learning phase, the robot is ready to recognize similar environmental conditions in which it started to learn the behavior. In the given example, environmental conditions are starting positions, box location, color or any other feature that can be perceived by the sensors. Due to the activation spreading in Semantic Networks, a node's activation propagates to all of its connected nodes and causes the linked sub-behavior and behavior nodes to be activated. Therefore, even by perceiving objects or locations other than the ones perceived during the learning phase, system can correctly recognize and execute proper sub-behavior. The execution of the sub-behavior is done by evaluating its node's activation level. After each perception, current activation level of all sub-behavior nodes is checked according to the selection threshold defined by the user. If the

activation level exceeds the given threshold, system executes that sub-behavior. The next sub-behaviors execute accordingly while deactivating their preceding sub-behavior nodes. Figure 3 depicts the performing phase for a given example. Suppose we replace the red box with a green ball in the same location and let the robot to move around and observe the environment. By perceiving the green ball, the corresponding nodes in the learned network is activated. Activation level of nodes spread through their connections based on the weights value. Depending on degree of generalization which corresponds to the energy level, other nodes in the network may get activated. Therefore, activation of *Ball* node activates *Movable Object* and to some degree, *Box* nodes. The same situation happens by activation of *Green* node which activates both *Color* and *Red* nodes. As illustrated in Fig 3, the nodes with yellow color are activated. The activation levels are shown by light or dark yellow color.
As a result of nodes activations, *Go to Object (S1)* sub-behavior node is activated and if its activation level exceeds the selection threshold, it will be executed by the robot. As mentioned earlier, we assumed that all the required low-level behaviors like moving and pushing are pre-programmed and do not require learning. Thus, robot moves toward the green ball and observes the environment again. At front of the green ball, the robot recognizes the same conditions for performing the second sub-behavior *(S2)* of the *Move Object* behavior. This activates *Push Object to the Location (S2)* sub-behavior node and due to the weight value equal to -1; it deactivates *(S1)* automatically. Finally, if the activation level of *(S2)* exceeds the selection threshold, the robot performs the second sub-behavior.

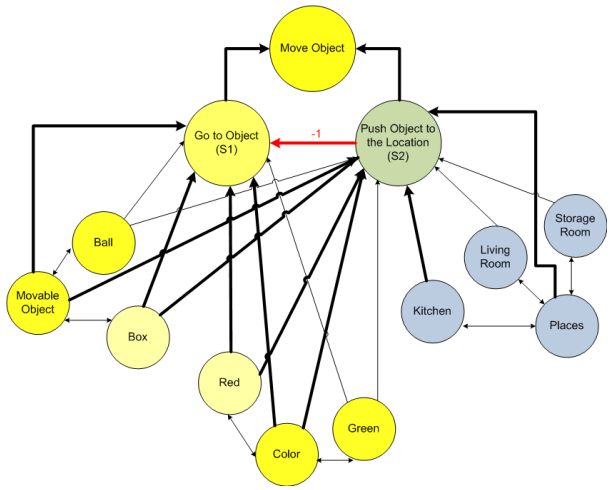


Fig. 3. Nodes which are activated during the Performing Phase.

4 Conclusion and Future Works

Sequence Learning is playing a key role in the task domains like planning, reasoning and robotics [6]. It is inspired by the humans and animals natural learning skills. In this research, incorporation of sequential learning and our previously developed architecture is discussed. By introducing Semantic Networks as a core element, its usage in sequentially learning and performing high-level representation of behaviors are elaborated. Also, the technique for disconnecting irrelevant nodes from the learned network is introduced. We believe the proposed approach enables the robot to focus on right aspects of the demonstration and incrementally teaches the robot new behaviors from demonstration.
Currently, our approach is incapable of handling quantities and negations. In our future work, we are going to define new link types in the Semantic Networks and design the high-level control in a way that can learn and perform more complex behaviors.
In this research we assume that sub-behaviors are determined by the tutor during the learning process. Therefore, an algorithm can be introduced to automate identification of sub-behaviors.

Acknowledgments

This work has been financed by the EU funded Initial Training Network (ITN) in the Marie-Curie People Programme (FP7): INTRO (INTeractive RObotics research network), grant agreement no.: 238486.

References

1. Oztop, E., Kawato, M., Arbib, M.: Mirror neurons and imitation: A computationally guided review. Neural Networks **19**(3) (2006) 254–271
2. Billing, E.A., Hellström, T.: Behavior Recognition for Segmentation of Demonstrated Tasks. In: IEEE SMC International Conference on Distributed Human-Machine Systems, Athens, Greece (2008) 228–234
3. Billing, E.A., Hellström, T., Janlert, L.E.: Behavior Recognition for Learning from Demonstration. In: Proceedings of IEEE International Conference on Robotics and Automation, Anchorage, Alaska (2010)
4. Chao, C., Cakmak, M., Thomaz, A.L.: Towards grounding concepts for transfer in goal learning from demonstration. In: Development and Learning (ICDL), 2011 IEEE International Conference on. Volume 2. (2011) 1–6
5. Fonooni, B., Hellström, T., Janlert, L.E.: Learning high-level behaviors from demonstration through semantic networks. In: to appear in proceedings of 4th International Conference on Agents and Artificial Intelligence (ICAART). Vilamoura, Algarve, Portugal. (2012)
6. Sun, R., Giles, C.L.: Sequence learning: from recognition and prediction to sequential decision making. IEEE Intelligent Systems **16**(4) (2001) 67–70
7. Coradeschi, S., Saffiotti, A.: An Introduction to Anchoring Problem. Journal of Robotics and Autonomous Systems **43** (2003) 85–96

8. Hartley, R., Barnden, J.: Semantic networks: Visualizations of knowledge. Trends in Cognitive Science **1** (1997) 169–175
9. Mugnier, M.L.: On generalization specialization for conceptual graphs. Journal of Experimental and Theoretical Artificial Intelligence **16** (1995) 325–344
10. Crestani, F.: Application of spreading activation techniques in information retrieval. Artificial Intelligence Review **11**(6) (1997) 453–482
11. Noreen, E.W.: Computer-Intensive Methods for Testing Hypotheses : An Introduction. Wiley-Interscience (1989)

How to Make a Taximeter Less Distractive

Anna Hortell

Department of Computing Science
Umeå University, Sweden
anna.hortell@gmail.com

Abstract. One major cause of car accidents is driver inattention while focusing on other objects in the car. As new technology is a necessity for many reasons the question arises if and when the new technology goes from being an assistance to a distraction? The trend with more electronic equipment in taxi cars increase distraction. Taxi drivers in Sweden caused 1184 accidents in 2009 and to prevent these accidents it is important to reduce the type of distraction objects in the car environment. By changing the interaction form, i.e. by not having to press on the taximeter, the sources of distractions will decrease. This article list the problem a taxi driver have today from a interaction point of view and gives gives suggestions on how to change it.

1 Introduction

New technology may be necessary for many drivers because it provides assistance in vehicles [7]. But when does the new technology become more of a distraction rather than an assistance? It is important that any negative effects on driving performance that technology might induce, such as distraction, are minimized before using the system [16].

Increased use of car technology, such as mobile phones, GPS and different entertainment technology increases the distraction factor in the traffic [14]. A taximeter is also a car technology, therefore increasing the possible of distraction.

The problem for a taxi driver today is that it is easy to shift focus from the traffic because a lot of information is shown on the taximeter and it is necessary to interact with the taximeter through touch [13].

It requires only small changes in traffic for a dangerous situation to develop, like a pedestrian walking over the road or the car in front breaking. If the driver then is distracted by other things in the car the reaction time increases. An experienced driver has a better idea of the potential dangers on the roads because they have the experience to interpret more clues in traffic [3]. A taxi driver is a professional driver and typically belongs to the category more experienced driver, this is shown in section 4. But, by having more technical gadgets in the car that experience might be overridden.

Sweden's biggest insurance company Trygg Hansa [13] went out with the recommendation to avoid pressing on the taximeter while driving, to reduce the accidents. The worker should instead park the car if he/she wants to interact

Bensch, Drewes, Gulliksson, Mejtoft (Eds.): Umeå's 15th Student Conference in Computing Science USCCS 2012, pp. 49–57, January 2012.

with the taximeter. The most common accidents caused by a taxi is rear-end collisions and collision with a parked vehicle [13]. Rear-end collisions has decreased with 23% to 369 collisions in four years, while collisions with parked vehicle has increased with 2% to 233 accidents during the same period. The number of accidents for a taxi driver in total 2009 was 1184, that is a decrease with 24 % from 1569 accidents in 2006. AAA Foundation for Traffic Safety recommends no reading while behind the wheel [1] thus no multitasking. More statistics are addressed in section 2 of this paper.

The main goal of the paper is to research guidelines and get insight into what makes a driver lose focus in the traffic, section 2. Also evaluate what a typical interface for a taximeter looks like and how the interaction between a driver and the interface is. This is shown in section 5. In section 6 some guidelines of what constitutes a good design is presented. By taking all these things into consideration, an idea of a new form of interaction with the interface will be proposed, so it takes up less attention from the driver.

The information that was gathered for this article is drawn from own experience as a taxi driver and a literature review.

The taximeter that is investigated in this paper is of the brand Frogne [5]. This article is written from the perspective of a taxi driver and by a taxi driver who works in Umeå and drives a normal sedan car, not a bus.

2 Statistics

About 90 percent of all traveling in Sweden is made on the roads. In 2007 there were almost 4.3 million cars in the country. In addition, there were over 500 000 trucks, 14,000 buses, 290,000 motorcycles, 325,000 tractors and about 150 000 EU mopeds in traffic. Until the early 1970's, the number of fatalities in traffic increased as the vehicles became more. Then began a downward trend, with the exception of the years 1983-1989 and 1999-2000. In 2007 471 people were killed in the traffic [15] this is shown in Figure 1.

A drivers inattention can be caused by a number of sources. Figure 2 lists the most common distraction sources [12]. It is shown in the figure is that "Using other device or object" and "Other" has a combined percentage off 28,5% and that is a great impact on the driver in a negative way. In a taxi, GPS and taximeter can be counted as "Using other device or object". Taximeter can also be counted under "Other" because it also make noises and the communication central may talk through it. All this is more described in section 5.

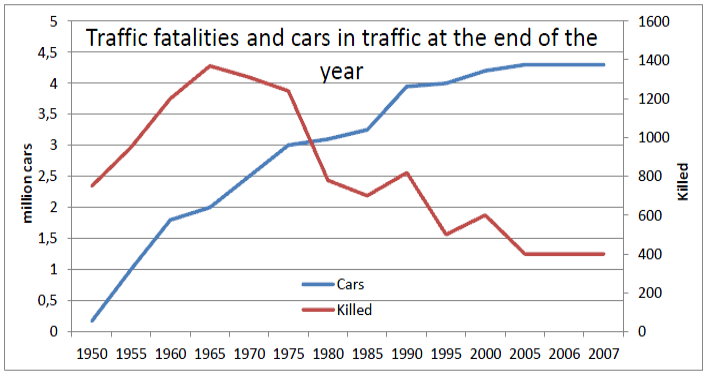


Fig. 1: How the development has been the last 50 years of car use and accidents [15]

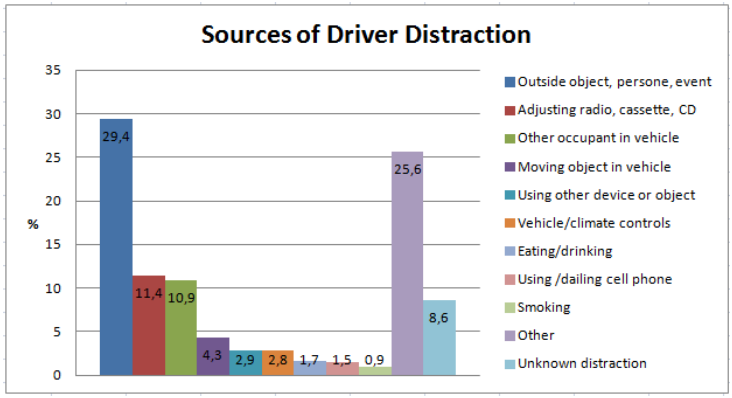


Fig. 2: Different distractions sources [12]

3 Method

As mentioned in the introduction is this article is written from a perspective of a taxi driver in Umeå and by a taxi driver. The information that was gathered for this article is drawn from own experience as a taxi driver and a literature review. The method that was used to investigate the taximeter was to list all the main features. There are many features on a taximeter some of them are used so rarely that there are no necessity to list them. From these main features selecting the three major tasks for a taxi driver, these are selected from the author's own experiences and is shown in section 5.2. Then mention the shortcomings of these functions from an interaction point of view. Investigate what makes for a better design to then give suggestions about what could make for a better design, interaction ways.
Another way of doing this investigation could be riding with a driver during a

working day. Taking notes of what is troublesome with the taximeter. As this paper was written from by a taxi driver the idea felt unnecessary.

4 The Taxi Occupation

Taxi occupation may vary to a certain degree because the companies have different agreements. This section will take up a relatively broad view of how it can be being a taxi driver.

4.1 A Workday

The workday starts with the driver logging in with the individual driver card. By doing this the driver gets in to the system so the communication central can see in which district the driver is.

A taxicab is four times longer in traffic then the average car, over a year. The average mileage for a taxi in Sweden is 65720 km [11]. A normal workday for a taxi driver includes driving between 200 km and 400 km. There are of course some exceptions such as when one trip is over 400 km one way. Those days the driver can easily drive up to 1000 km in one shift.

On a normal day the driver has between 15 to 35 trips, depending on how busy the workday is and how many taxis are in service.

4.2 Driving Assignment

When a worker gets a driving assignment an inquiry will appear on the taximeter asking if they want accept or deny. If the driver chooses to deny the assignment then nothing happens aside from that the driver get cut off from the system for three minutes. If the driver chooses to accept the mission then it will appear on the taximeter. Usually the address where the customer is supposed to be picked up is written first. Then what kind of trip it is, for example, driving schoolchildren, disabled or just an ordinary trip. The text usually ends with some information where the customer is getting dropped off.

On an ordinary trip there are two factors that determine the price. First, the time it takes for the whole trip so if the car is standing still it will still cost money. Second, the distance of the trip. Those two factors are then combined and calculated by the taximeter [10]. Other trips only have the distance as a factor for calculating the price, not the time it takes [6].

5 Interaction Devices

The display covers the entire workspace and tapping on the screen serves all the features. The built-in magnetic card reader extends the application possibilities for example, the use of the individual driver card and bankcards [5]. The taximeter has a built-in speaker and microphone in case the worker wants to

communicate with the communication central and vice versa. The display foot is positioned on the dashboard where usually all taximeters are installed. With the very flexible neck of the monitor the user can always place within the drivers field of vision.

5.1 Technical Specifications

There are two types of taximeters. The new version has a touch screen with an 140 mm TFT color display, 16 bit colors[1]. They both have a dimension of 170 x 129 x 24 mm(WxHxD). It has also a built-in speaker and microphone [5].

5.2 Interface

The interface consists of a tab system with five tabs.

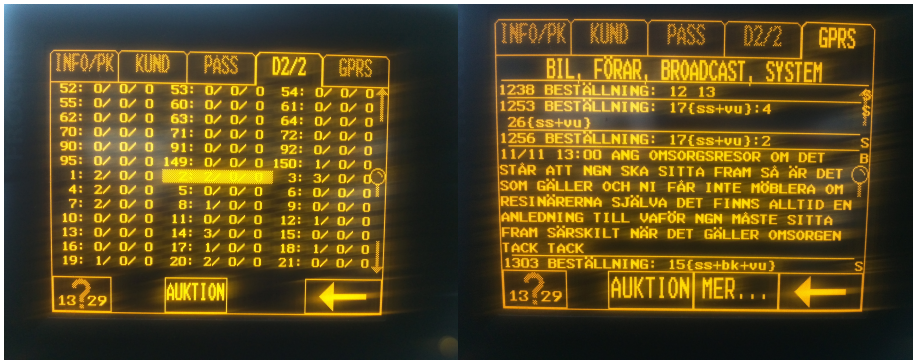
1. Info/kp. Here is where the driver can adjust the settings for the interface like light and sound. The pressure sensitivity and the precision sensitivity can be regulated.
2. Trip. When the driver gets an assignment here is where it shows. When an assignment has been received, usually, three buttons appears on the interface when on trip tab. Start journey, cancel journey and callback. The callback button appears only when there are a register cellphone number to the customer.

(a) When a trip has been initiated

(b) Here the driver gets all the information about the shift

Fig. 3: Screen shots of the interface

[1] The colors is not shown in the pictures in this paper due to an older version off the interface



(a) Overview of the districts (b) Messages

Fig. 4: Screen shots of the interface

When a journey has started new buttons appear on the screen (see Figure 3a). Partial Pay, Choice, Change the tariff, Card check, Cash register. Choice button is for example, to call for another car to the address without involving the communication central. Card check button is for those customers who want to pay in beforehand.

3. Pass. Here all the information about the shift is shown. How far the worker has been driving, both in paid distance and the total distance and how much money the driver has collected during the shift as well as how many trips that have been made. Other information is how long time the shift has been and how long break has been made (see Figure 3b).
4. District. Shows which district the car is in and also an overview of all the districts and how many cars that is in the different districts. Here the driver also registers the district he or her will end up in when a trip is over.
   The districts appear like: 2: 1/0/1 which means: district two: one car in the district/ no direct bookings/ one pre booking (see Figure 4a).
5. Messages/channel. The communication central sends out messages to all the taxicabs if there, for example, is car shortage when a plane is landing (see Figure 4b). They also send private messages to drivers, it can be a customer's telephone number, or changing the pick up address. The drivers can send personal messages to each other but a messages only consist of numbers.
   The driver can change the channel if necessary. Sometimes channels get stuck, the communications center can not communicate with the taxis and the channel has to be switch to another channel.

When the driver does not have an assignment the two tabs that are most used are the district tab and message tab. By switching between these two, the driver has a good overview of what is happening on the roads.

The three most important tasks of the taxi driver are:

1. To get an assignment - the diver have to press a button to accept
2. Lock at which districts are empty - the driver have to change tab to lock up the districts
3. Check the messages for important information - the driver has to change tab to lock up the messages

These three things are repeatedly check up throughout the working day.

5.3 Heads up Display

A heads-up display (HUD) can be used in vehicle to observe various information [8]. HUD is any transparent display that presents icons and texts, usually found on the dashboard of the car [2], without requiring users to look away from their usual viewpoints.

5.4 The Steering Wheel

The cars that are on the market today often have the majority of controls and information of the systems, in the middle console and a small set of additional physical controls on the steering wheel. A study has shown that using a multi-touch steering wheel reduces the visual demand by a large degree [4]. By moving controls to the steering wheel the visual demand reduces with an averages of 67%.

6 A Good Design

A very important part of designing a taximeter is that it should not be distracting to the driver. While designing things people use everyday there are certain things to take in consideration [9].

Such as:

- It should be easy to determine which actions are possible to do at any time.
- Making things visible, including the conceptual model of the system, the alternative actions and the result of documents.
- Making it easier to evaluate what state the system is in.

A major role of technology should be to make tasks simpler [9]. Tasks should be simple in structure, overcomplicated structures only makes for more errors and frustration. That can be done by keeping the task much the same but also providing a mental aid.

How smart design is the new competitive frontier. How and why some products satisfy customers while others only frustrate them.

7 Results and Conclusion

From the literature study it is clear that more technology would increase the drivers distractions and thereby the number of car accidents. Taxi drivers caused 1184 accidents in 2009 [13]. The trend with more electronic equipments in taxi cars gives an enhanced distraction situation. To prevent accidents it is important to reduce the type of distraction objects in the car environment. As previously mentioned the three major tasks of a driver is assignments, districts and messages. What they all have in common is that a driver has to interact with the interface by touching it. As it is known [4] using the steering wheel reduces distraction. By changing the interaction form, from buttons on the taximeter to putting a button on the steering wheel instead, it leads to the driver never has to leave the grip off the wheel. This button should not control the entire interface, as that would get overly complicated, just the main parts. This should be switching from the different tabs and accepting assignments, which is the biggest part when it comes to interaction between the taximeter and the taxi driver. As technology should make tasks simpler [9] it is important to not let technology take over. For example, in Figure 4b the messages are written with a small text, they can be difficult to read long messages if not concentrated. If the driver instead has the option to let the taximeter read long messages could give for a easier interaction. There is no need for the driver to concentrate on reading what is on the taximeter instead listening to what it has to say. This is possible because it has built-in speakers [5]. From how a good design should be e.g. making things visible, one could argue that having HUD in taxis would diminish the problem with looking away from the traffic. By having the interface in sight of the driver, i.e. not having to look to the right, would provide a better focus. An overview of the changes that are suggested is seen here in Table 1.

Important tasks	**Get an assignment**	**Overview of the districts**	**Check messages**
Solutions	Button on the steering wheel to accept	Button on the steering wheel to scroll between tabs	Possibility to hear long messages
Guidelines	Making things visible - with a button	What state the system is in - which tab the taximeter is in	Which actions are possible - choose between hearing or reading

Table 1: Table of the solutions

The limitations of this work is that this suggestions, moving the interaction, is not evaluated by more taxi drivers. Maybe then new insights might have come up. Other limitations is that the layout of the interface was not investigated more only the interaction. By investigate the layout, more shortcomings might have

come up. Another interesting information to have would be to compare different types of taximeters to each other and list advantages and disadvantages.

References

1. AAA Foundation for Traffic Safety: Distractions in Everyday Driving (2011)
2. Benoit, A., Bonnaud, L., Caplier, A., Jourde, F., Nigay, L., Serrano, M., Damousis, I., Tzovaras, D., Lawson, J.: Multimodal signal processing and interaction for a driving simulator: component-based architecture. Journal on Multimodal User Interfaces 1(1), 49–58 (2007)
3. Borowsky, A., Shinar, D., Oron-Gilad, T.: Age, skill, and hazard perception in driving. Accident Analysis & Prevention 42(4), 1240–1249 (2010), http://www.sciencedirect.com/science/article/pii/S0001457510000357
4. Döring, T., Kern, D., Marshall, P., Pfeiffer, M., Schöning, J., Gruhn, V., Schmidt, A.: Gestural interaction on the steering wheel: reducing the visual demand. In: Proceedings of the 2011 annual conference on Human factors in computing systems. pp. 483–492. ACM (2011)
5. Frogne: Färgdisplay. http://www.frogne.dk/sv-SE/SE/Produkter/Produktinformation.aspx (2011), accessed 2011-10-19
6. Gerst, W.: Taximeter (12 1980)
7. Jahn, G., Krems, J., Gelau, C.: Skill acquisition while operating in-vehicle information systems: Interface design determines the level of safety-relevant distractions. Human Factors: The Journal of the Human Factors and Ergonomics Society 51(2), 136 (2009)
8. Matsumoto, T., Watanabe, R., Eguchi, S., Aritake, H., Morihara, T., Yamagishi, F.: Heads-up display (11 1993)
9. Norman, D.: The design of everyday things. Basic books (2002)
10. Rodriguez, O., Ramos, S.: Standard taximeter using gps technology (29 2009)
11. Statistiska centralbyrå: Transporter och kommunikationer statistisk årsbok 2011 (2011), http://www.scb.se/statistik/_publikationer/OV0904_2011A01_BR_11_A01BR1101.pdf
12. Stutts, J., Stutts, J.C., Reinfurt, D.W., Staplin, L., Rodgman, E.A.: The role of driver distraction in traffic crashes. AAA Foundation for Traffic Safety Washington, DC (2001)
13. Trygg Hansa: Taxichaufförer vållar 25 procent färre olyckor nu jämfört med 2006 (2010), http://www.mynewsdesk.com/se/pressroom/trygg-hansa/pressrelease/view/taxichauffoerer-vaallar-25-procent-faerre-olyckor-nu-jaemfoert-med-2006-388890
14. Underwood, G.: Visual attention and the transition from novice to advanced driver. Ergonomics 50(8), 1235–1249 (2007)
15. Vägverket: Alkohol, droger i trafiken (2008), http://publikationswebbutik.vv.se/upload/1745/88294_alkohol_droger_trafik_6.pdf
16. Young, K., Regan, M., Hammer, M.: Driver distraction: A review of the literature. Distracted driving. Sydney, NSW: Australasian College of Road Safety pp. 379–405 (2007)

User Perception of Reliability in Mobile e-Commerce Applications

Annika Jonsson

Department of Computing Science
Umeå University, Sweden
id07ajn@cs.umu.se

Abstract. The increasing acceptance and adoption of e-commerce in general, along with the availability to the internet of mobile devices, have created a market for mobile e-commerce. Applications aimed for the mobile e-commerce industry place demands on design since they handle money and personal and financial information of the user. Problems to attend to are how to design in order to make the user perceive the application as reliable. A literature review is conducted covering the areas trust, privacy, security and perceived information quality which are shown to affect the perceived reliability in web-based e-commerce sites. In order to apply the information gathered from the literature review, a study of what affect the user acceptance in mobile applications is conducted. Based on information gathered from the literature review and the study of user acceptance, this article presents guidelines on how to design a mobile e-commerce application to be perceived as reliable.

1 Introduction

E-commerce is a way to electronically conduct business through the internet [14]. The adoption of e-commerce in Sweden has continuously increased since the year 2000. In 2010 81% of the people, 18 years or older, had used an online vendor for shopping and more than a third of the Swedish population above 12 years purchase online at least once a month. That equals 4.5% of the total retail sales. On top of this the use of internet through mobile devices has increased from 42% of the population 2009 to 64% 2010 [4] and today many of the larger banks and shopping sites have mobile applications available. This shows that the potential market for e-commerce using mobile devices is increasing, thus do the need of knowledge of how to design mobile e-commerce applications in a successful way.

The purpose of this article is to provide knowledge about how to design a mobile e-commerce to be perceived as reliable. Since an online service of this type handles the money of the customers as well as personal and financial information, it is of importance to establish what key values a user associate with a reliable web-based e-commerce service. The two notion reliability of an application and the perceived reliability might not necessary mean the same thing. Reliability of an application is thought of, in this article, as how reliable an application actually is, while perceived reliability refers to how the user think of an application.

Bensch, Drewes, Gulliksson, Mejtoft (Eds.): Umeå's 15th Student Conference in Computing Science USCCS 2012, pp. 59–69, January 2012.

However, to design an application to be perceived as reliable is probably useless, unless it reflects the actual case.

The goal is to apply the results from a review of articles, covering the area of user perception of reliability in web-based e-commerce applications, to mobile e-commerce applications. The findings are summarized in a number of guidelines.

To be able to analyse why something is perceived as reliable, the concept of reliability need to be defined. In addition to reliability, also security is defined below. The perception of security is related to the perception of reliability and is therefore included in the review. The definitions serve as a starting point when searching for relevant literature to base the article on.

Reliability has been defined by many different disciplines. The standard definition of software reliability is the probability of a failure-free operation of a computer program [16]. The corresponding definition from a psychological perspective is "The quality of being trustworthy or dependable" [2].

Security can be summarized as the protection and prevention against unauthorized access and/or destruction of information [3]. The more general description is freedom from risk or danger [9].

A literature review of existing research of the area is presented in Section 3 to provide a theoretical foundation. Concepts covered in the review are trust, privacy, familiarity, perceived information quality and security, which all show to have impact on the perceived reliability. To be able to apply the literature review to mobile applications, a section is dedicated to describe mobile services and user acceptance along with benefits and drawbacks compared to using desktop computers. In the last two sections a discussion of how to apply the results to mobile applications are presented and summarized in a number of guidelines to consider when developing a mobile application aimed to the e-commerce market.

2 Method

A considerable amount of research of reliability in web-based e-commerce sites has already been done, but the corresponding research for mobile applications is not as extensive. Because of this the literature review is based on research conducted of web-based e-commerce rather than mobile applications. The literature used are mostly articles found by using Google Scholar during the autumn of 2011.

Designing for reliability is important not only for e-commerce applications but for other applications as well, e.g. banking applications. However, in this article e-commerce applications were chosen to delimitate the topic and to simplify the search for relevant literature.

3 Literature Review

The above mentioned concepts serve as a base for the literature review presented in this section. The definition of reliability includes trust, which because of that

also is viewed as an important concept to take into account when striving for reliability. In addition to trust, the concepts of privacy, familiarity, perceived information quality and security were shown to have an impact on decision done by the users and the trust users have in online vendors, and were therefore included.

3.1 Trust

Several definitions of trust exist and are used usually depending on the area of interest. Psychologists define trust as a willingness to trust others, while social psychologists might define trust as cognition about the trustee and their attributes. Because trust is multidimensional, one single overall definition that stretches over all disciplines is hard to find.

Previously conducted research about the adoption and intention to use e-commerce, shows trust as a factor that affects the decisions of the customer. Consumers often hesitate to interact and purchase from online vendors because of the perceived risk of providing financial information and personal identity as well as an uncertainty about the intentions of the web-based vendor [5, 8]. As stated above, the idea that trust is multidimensional in nature, the concept affects e-commerce and the relationship between vendors and customers in several ways. For example, trust in that the internet is technically secure is not necessarily the same kind of trust that a person feels when trusting a vendor to use provided personal information ethically and morally correctly or when trusting a web vendor to actually fulfil his part of an agreement. All these trust-related behaviours influence the process of building trust and should be assessed when web-based vendors are about to develop a strategy for trust between themselves and their potential clients [8].

Researchers have found that several factors affect whether a customer engage in trust-related behaviours or not [5, 8, 10, 13]. These factors will be described below.

Disposition to trust or propensity to trust refers to what extent a person shows willingness to depend on other people and situations. Some people are more open to engage in behaviors that require some level of trust between the parties, while some might be suspicious to all attempts on building trust. *Faith in humanity* and *trusting stance* are two subconstructs of this category. Faith in humanity means that people assume that others are reliable, benevolent and dependable. Trusting stance refers to when a person thinks that better outcomes will be obtained by dealing with people as if they are reliable and benevolent, regardless whether one really believe people are reliable or not [8].

Institution-based trust means that the structural conditions needed to achieve a desirable outcome are in place. Institutional-based trust originally comes from sociology and deals with structures, for example, legal conditions that make a specific environment appear trustworthy. Technological safeguards are examples of this in e-commerce and other web-based business. Institution-based trust has two dimensions, *structural assurance* and *situational normality*. Structural assurance is where one believes that contextual conditions, such as

contracts, promises and guarantees are in place, and therefore success is a likely outcome. This helps lower the risk users might experiences while taking part in an online transaction [10]. If the user believes that success is likely because the situation seems normal and well-ordered, then it is the case of situational normality [5].

Trusting beliefs is that a truster perceive the trustee to possess attributes that are beneficial to the truster. Some of the main attributes that researchers agreed on are: [8, 5, 13]

- *Competence*, the trustee possess the ability to do for the trustor what the trustor needs to have done and fulfill their goal effectively.
- *Benevolence*, the trustee cares about the trustor and are motivated to act in the interests of the trustor.
- *Integrity*, the extent to which the trustee is honest, promise keeping as well as adhering to principles that the trustor believe is acceptable.
- *Predictability*, the trustee's actions are consistent enough that the trustee are able to foresee the outcome of a given situation.

The extent to which these attributes are consistent with how the trustor see the vendor have impact on whether the trustor finds the vendor trustworthy or not.

Perceived information quality represents a reaction, of the user, to the information versus the requirements the user have on the information. The information must be accurate, complete and timely to provide a relevant and reliable appearance to the vendor and website. Another important aspect is that the information must be easy to access and be placed in a relevant context [10]. Nicolaou and McKnight [10] propose that the two notions, *control transparency* and *outcome feedback*, influence the perceived information quality because they provide users with information and cues by which the users can form opinions about the quality. Control transparency means the availability of such cues and information that makes the users aware that the transaction or data exchange, when discussing the area of e-commerce, is taking place. Outcome feedback are denoted as the information available to control the outcome of the exchange, e.g, transaction data validation. According to Nicolaou and McKnight's [10] study, control transparency does affect the perceived site quality. This quality was substantially higher when the control transparency was high. This notion has an positive impact on trust and on the intention of users to use a system because of when the transaction is performed online, the user cannot obtain assurance of the accuracy of the exchange from the vendor in person and must rely on provided information. With relevant and accurate information, the perceived risk with the purchase is decreased, and whereby the trust in the specific vendor is increased [10].

3.2 Privacy

When people use services online or from a mobile device, the user might not know the identity of the service provider in the same way as if they were dealing with

someone in person. This fact might lead to an uncertainty of the intentions of the vendor or service provider. The level of concern for privacy depend on which sort of information the users are asked to provide. The primary concern of the user is about revealing identity and financial information and having it stolen by unauthorized individuals [8, 1]. However, studies show that the usefulness of the service has a high impact on whether people are willing to take the risk or not.

3.3 Familiarity

Familiarity with a vendor and its processes affect the intention of a customer to use a website or a vendor for its needs. Gefen [5] explains that as familiarity helps reduce uncertainty and simplify the relationship because it is built from previous interactions with the other party. In this case, familiarity is defined as the knowledge and understanding of why, what and where a vendor or website does what they do. One example is that familiarity with the online vendor Amazon.com could be to know how to search for books and how to order them. If the previous experience of the vendor is positive, then the familiarity increases the trust as well. Familiarity and trust are closely connected when discussing what is influencing a user to engage in e-commerce, to use a service from a specific vendor, and to make a decision to purchase. Although Gefen [5] argues the importance of familiarity when discussing trust in a vendor and the intention to use a vendor, he also states that the disposition to trust, of a user, is primarily the reason that affect the trust in the vendor. Familiarity supports e-commerce and increases users willingness to purchase by helping to build trust. However another researcher, Kaasinen [6], also states that familiarity to a brand or service from another type of environment such as television or news paper also helps the user accept a service or a vendor.

3.4 Security

The importance of security is natural when discussing reliability and trustworthiness of applications and services handling purchases, personal and financial information of customers. As mentioned by McKnight, Choudhury and Kacmar [8] customers often hesitate to deal with online vendors because of the risk of having their identity stolen by hackers as well as uncertainty concerning the behaviour of the vendor. Ka-Ping Yee [15] brings up two ways of establishing security.

Security by admonition relates to notifying the user of a problem or a performed action in order to prevent a program from carry through actions which are undesirable. This approach might be illustrated as when a warning is displayed that enables the user to take action to prevent the undesirable action to happen. A problem with this approach is that the user might not have all the necessary knowledge about the context because the warning is initiated by the program and not by the user [15]. The prompt might seem unjustified or unintelligible by the user, which makes it complicated for the user to take a good

decision. This kind of prompt might not be a desirable way to inform the user because it is interrupting the workflow of the user and teaches them to dismiss warnings carelessly. The use of proximity and relevance are more preferable ways to ask for attention than by aggressiveness [15]. An example given is the way the web browser firefox use to inform the user about that a site is trying to install software. It is done by adding a transient bar to the browser window. This way is preferable because it does not interfere or prevent the users to perform their work as a prompt would have done.

The second way is denoted as **security by designation**, this means that a user authorizes the program to perform a certain action at the same time as a command is given. An example of designation is the action to send an email with an attachment. The user do not have to both select a file to attach and then separately assign reading permission to the receiver. By attaching a file, the action performed defines the file to be sent, the receiver it should be sent to and assigns the authorization to read the attachment in one single action. The benefit with this approach is that it is straightforward and a smooth way of integrating security decisions with the primary task [15].

4 Mobile Devices and Services

One can see from everyday life that mobile phones are today not only used as a device for communication. They are equipped with a wide variety of applications and services and the internet can be accessed from almost everywhere. The past years have the number of applications and services available on the market increased [7, 11]and more and more advanced and security demanding services are developed. One example is that many banks have an application that enables customers to access their accounts and transfer money with their mobile phone. To design applications it is useful to know how users are thinking about mobile applications and services [12]. This section will cover what influence users to use a mobile application. It will also try to find benefits and drawbacks that mobile devices have versus desktop computers and fixed internet connections.

4.1 User Acceptance

In the dissertation written by Eija Kaasinen [6] four factors were defined as the foundation to user acceptance regarding mobile services.

- Perceived value of the service
- Perceived ease of use
- Trust
- Perceived ease of adoption

Perceived value refers to the key features which are the main reason why users appreciate a particular service. By defining these values, the design can be adapted to focus on the most essential issues. The study conducted by Kaasinen [6] also concluded some characteristics which were generally appreciated and

valued by the test subjects. The service content should be comprehensive, topical and familiar. The user should be able access all relevant information but it is important that the information is structured in a way that enable the user to access it in small portions. The acceptance of the mobile service increases if the user already is familiar with the brand or service from another environment, such as television[6].

The users also found service entities, which were seamlessly integrated, valuable. An example of such an entity was if the user had the possibility to directly call a nearby restaurant which was recommended in an application showing locally located restaurants. The possibility of personalisation was also an appreciated feature, which enhanced the willingness to use an application.

Perceived ease of use. An important feature was a clear overview of all information and services available. As when meeting another person for the first time, the first impression is of great importance. It can both end up with appreciation of the other person as well as the opposite. The same can be applied to the first impression of an application. A fluent and well thought through navigation was also something the test subjects favored. Whether the navigation works smooth or not, affect how fast a user was able to reach the information of interest. It also had an impact on the willingness to pay for the service [6].

Trust is of value not only to web-based sites and services but also when dealing with mobile applications. Mobile devices are often able to collect information of the location and environment surrounding the user, which makes trust a more and more important aspect.

Perceived ease of adoption is the last of the four factors described. Ease of adoption refers to how smoothly and effortless a new user can take a service into use. First time users must be carefully designed for.

4.2 Benefits and Drawbacks with Mobile Devices

The differences between designing for something used on a desktop computer and a mobile device are many. Physical attributes such as smaller screen size, type of interaction as well as multiple types of devices must be considered. When the focus is on e-commerce it can be assumed that one of the goals, in many cases, is to reach out to as many customers as possible no matter what kind of mobile device the customer uses. This concludes that the application must be runnable on a variety of different devices. Further, one benefit for the users is the availability of mobile services, with connection to the internet almost anywhere, services and information can also be accessed at almost any time and anywhere [6]. This fact can be utilised by companies to provide locally and topical information to the user depending on the current location. E.g. providing travel guidance to the nearest office or store.

This benefit of mobility in itself creates approaches that should be considered when developing and designing an application. The environment in which an application is used can vary greatly, both between different sessions but also during a single usage session, depending on the location of the user. When deciding about how to interact, designers should, among other things, consider

how high the possibility of noise from the surrounding is. An aspect that might cause problems are the connection to the internet, some have connection at all times and some only use the internet when wireless networks are available. The internet connection might also be disrupted by other reasons, such as empty battery, which causes the mobile device to shut down and results in an interrupted application. These variations should be taken into account, especially when considering applications which frequently handle transactions of money. The application should be designed in such a way that it handles this type of unexpected interruptions in a secure and reliable way.

Momentary usage is a result of the mobility of the device. People might use a service or application to fill up empty moments, for example, while sitting on the bus or having a break at school. This may lead to expectations and wishes to be able to leave the service for different amount of time and then be able to pick up the usage session at a later moment.

5 Discussion

This section will discuss each of the concepts covered in the literature review and how to apply them to mobile e-commerce applications.

The literature study shows that trust is of importance when the goal is to develop a reliable and successful service which handles personal data and financial information. The areas explained in the literature study are not independent from each other. For example, trust can be strengthened in the same way as reliability can, by taking familiarity, privacy, security and information quality into account.

As mentioned in Section 4, trust is also a perspective that affects the willingness of a user to accept mobile services. It can therefore be assumed that it also is of importance when discussing mobile e-commerce applications. Because of information is more difficult to convey on mobile devices, e.g. because of smaller screen size compared to desktop computer screens, trust might have an even more important role. Trust should therefore be viewed as a main issue to address when making design decisions.

The disposition to trust that a user has can be problematic to design for because it is subjective and vary from person to person depending on personality and prior experiences. Although, by addressing other approaches, such as institutional based trust, trustworthiness may be achieved. Technical safeguards can be used to increase the perceived institutional based trust, the problem is how to assure the user of the presences of these. The small screen of a mobile device restricts the amount of information to be displayed and also this kind of information might not be appropriate to display on the first view of the application. The main interest of the user are the usefulness and perceived value [6]. To display a full description of the technological security that are present can have a discouraging effect, and in that way it might counteract the reason why the information was put there in the first place. Because the user acceptance is higher among services that are found useful, this must be a main focus. How

should the features be displayed in the best way? If the user find the application valuable they might be more motivated to actively find out which technological and legal safeguards that are present, this information can then be displayed in a secondary view to where the users navigate on their own.

In a similar way as the security information, the privacy policy and information about ethical considerations can be available for those who are interested, instead of taking valuable space from the main functions.

The main advantage with mobile devices is its mobility. Because of its appreciation the security of the application must be able to handle interrupted connections. By using security by admonition, a confirmation can be delivered that the transactions are handled in a safe way, also when the application is interrupted. By doing this the application may be perceived as more reliable. The user must be aware of whether a transaction was successfully completed or not.

Logout are also important in e-commerce applications to avoid unauthorized usage. It must be obvious to the users if they are logged in or not. Again, because of the mobility and changing environment, the security must handle if a user decides to interrupt a session without turning the application off or log out. Momentary usage can be handled by automatic logout after a certain amount of time to avoid unauthorized usage. In other, less urgent, situations the security can be handled with security by designation.

The concept of familiarity in a mobile application may be reached by using a consistent design and layout that will help the user to remember how the application works and how the processes are built. If the design is consistent with the graphical profile of the provider or designs from other disciplines or products the user might recognize the similarities and easier learn how to use the interface. Because of the simplified learning, the familiarity will also have a positive impact on the perceived ease of use when the user learns the navigation and structure of the application faster. By being clear with who the provider of the application is and where to turn when a problem or question arises, the familiarity is supported. Familiarity may also help to gain trust from those who have low disposition to trust. If trust to services of a specific provider in other areas already exists, the disposition to trust the provider might be higher than if the provider was not recognized at all. Because of that familiarity support building trust between the provider and the user, familiarity also supports the perceived reliability.

With high user acceptance of an application follows a higher amount of people that actually uses the application. As mentioned in the previous section, integrated features, which simplified for the user, were of value to the user. By making support or contact directly available by integrating call service or email client from the application itself, the user acceptance may be enhanced. In the same time, by having a way of assisting the users when problems are encountered, the perceived reliability is supported by personating the provider as benevolent and helpful.

Why the application is useful and which purpose it fills for the user, are among the more important things to design for, because the application otherwise will not be useful. Without consistent structure, well thought through navigation and a first view which in an obvious way show the main functions of the application, it might be hard to make the user start and keep using the application. This, along with promoting the service owners competence and willingness to provide the user with desirable results, increases the chance of developing a successful and useful application.

6 Conclusion

When designing e-commerce applications for mobile devices there are numerous concepts that should to be considered. To develop an attractive and well accepted application by the user, the designer may consider some important areas which increases the probability of the application being perceived as reliable. The areas that are of value, showed by the literature review, can be summarized in the following guidelines.

1. Use several well thought through trust building strategies to create a good relationship to the user. The ethical considerations regarding, among other areas, personal and financial information must be available to the users who wishes to take part of it. Create a clear structure of the information and process which convince the user that all necessary knowledge and structures are in place to reach the desired result. Enable functions which assists the user when encountering problems or questions, in order to make the user perceive the application and the provider as benevolent and helpful.
2. Aim for designing a familiar environment which is easy to use and adopt. If the provider appears in other disciplines, take advantage of prior experience and knowledge about the provider by using similar graphical and structural language. Make it obvious who is providing the service and how to use the application.
3. Smoothly integrate functions which assists the users when trying to reach their goal, such as easy access to support and relevant feedback.
4. Use technological safeguards and structures to ensure the safety of the identity and money of the user. The information of the user should be kept safe also during interruptions or if problems with transactions arises. Implement automatic logout to handle momentary usage and avoid unauthorized usage. Demand confirmation for important decisions but use security by designation in situations where the actions and expected output are obvious.
5. Provide relevant, accurate and timely information about functions and the situation at hand, to prevent contradictions and confusion by displaying expired or incorrect information.

References

1. Barkuus, L., Dey, A.: Location-based services for mobile telephony: a study of users' privacy concerns. In: Proc. 9th IFIP TC13 International Conference on Human-Computer Interaction (2003)
2. Colman, A.M.: reliability n. A Dictionary of Psychology, Oxford Reference Online (2009), http://www.oxfordreference.com/views/ENTRY.html?subview=Main&entry=t87.e7109
3. Daintith, J., Wright, E.: security. A Dictionary of Computing, Oxford and Reference and Online (2008), http://www.oxfordreference.com/views/ENTRY.html?subview=Main&entry=t11.e4659
4. Findhal, O.: Svenskarna och internet 2010 (2010), https://www.iis.se/docs/SOI2010_web_v1.pdf
5. Gefen, D.: E-commerce: the role of familiarity and trust. Omega 28, 725–737 (2000)
6. Kaasinen, E.: User acceptance of mobile services – value, ease of use, trust and ease of adoption. Ph.D. thesis (2005)
7. Lawson, S.: Android market needs more filters, t-mobile says. PC World (2009), http://www.pcworld.com/article/161410/android_market_needs_more_filters_tmobile_says.html2012-01-04
8. McKnight, D.H., Choudhury, V., Kacmar, C.: Developing and validating trust measures for e-commerce: An integrative typology. Information Systems Research 13, 334–359 (2002)
9. Nationalencyklopedin: säkerhet (2011), http://proxy.ub.umu.se:2067/sve/s%C3%A4kerhet
10. Nicolaou, A.I., McKnight, D.H.: Perceived information quality in data exchanges: Effects on risk, trust, and. Information Systems Research 17, 332–351 (2006)
11. Paul, I.: Android market tops 400,000 apps. PC World (2012), http://www.pcworld.com/article/247247/android_market_tops_400000_apps.html2012-01-04
12. Stone, D., Jarrett, C., Woodroffe, M.: User Interface Design and Evaluation. Morgan Kaufmann series in interactive technologies, Morgan Kaufmann Publishers (2005)
13. Sutcliffe, A.: Trust: From cognition to conceptual models and design. In: Dubois, E., Pohl, K. (eds.) Proc. CAiSE 2006. Lecture Notes in Computer Science, vol. 4001, pp. 3–17 (2006)
14. Wu, F., Li, H.H., Kuo, Y.H.: Reputation evaluation for choosing a trustworthy counterparty in c2c e-commerce. Electronic Commerce Research and Applications 10, 428–436 (2010)
15. Yee, K.P.: Guidelines and strategies for secure interaction design. In: Security and usability : designing secure systems that people can use illustrated edition, chap. 13. O'Reilly (2005)
16. Zhang, J., Zhang, L.J.: Criteria analysis and validation of the reliability of web services-oriented systems. In: Proceedings of the IEEE International Conference on Web Services (2005)

Analysis of Interaction with Graphical Interfaces for Music Creation

Carl Larsson

Department of Applied Physics and Electronics
Umeå University, Sweden
dit06cln@cs.umu.se

Abstract. The number of implementations of software that represent real world items are increasing and this paper discusses how graphical interfaces for music creation are designed with regards to specific design models. One of the problems with designing for new media is that it is difficult to have an innate knowledge of how to use the medium. This article evaluates further how this has been managed when implementing possibilities for users to use applications for music creation. In order to not make the evaluation too general the software used in the analysis have been divided into subsets based on their distinctive traits. Two design models regarding emotional design and usability where chosen and tried to correlated the attributes of these to the subdivision of different interfaces.

1 Introduction

A difficult interaction design question, as the physical world is becoming more virtualized, is how to make a satisfactory transition from an objects traditional medium to a different medium. The focus of this article concerns the transition to a medium that acts in a different medium- and domain space.

This particular area has been chosen because it addresses a situation where the traditional media is a 3D, tactile, *"real world"* instrument which is represented in, or projected onto, a 2D graphical user interface. Considering that the interface defines the connection between interaction and cognition [15] it is interesting to further investigate if this connection remains intact when moving the interface into a new medium. The motivation of moving the instrument into a virtual world is not to entirely remove the need for real world devices, but to enhance the use of these artifacts. When discussing music creation the opportunities of moving to a 2D world is that it gives the user possibilities that he or she would not have when playing an instrument. Even though some measure of *affordance* could be lost, it provides greater areas of use. For instance if someone where to record their composition using an instrument one would need some sort of recording device, normally found in a studio. By setting the creation of music in a digital context, this becomes a minor issue as it can be provided in the same system as the music creation service is provided.

Bensch, Drewes, Gulliksson, Mejtoft (Eds.): Umeå's 15th Student Conference in Computing Science USCCS 2012, pp. 71–83, January 2012.

The idea that the demands on actions in the physical world are lowering because of the increasing use of actions in virtual worlds [16] seems to imply that the future for interfaces will be less situated in a real-world context. This in turn makes it necessary to solve the problem of how to narrow the entire virtual domain space of music, and in particular music creation, to a medium space that (in comparison) is quite specified.

Interaction with software made for music creation widely varies and consists of a vast number of different solutions on how to let the user act in the system, and this paper will research different solutions to the constraints and opportunities in this musical design process.

The main purpose of this research paper is to define different traits in the graphical user interfaces with regards to specified models. This is done so one could apply these attributes to the implementation of such a software with regards to what sort of interface that is sought to create. This could also in the future lead to creating a design model for the area and find out whether or not there are any guidelines that can be found when implementing for a new media. The results does not only affect designing interaction for music creation, but can be applied to the interaction design of a number of different physical artifacts that are transposed to a *"virtual"* world.

2 Method and Interfaces

2.1 Choice of Method

The softwares that have been used in this evaluation are programs made explicitly for music creation. Because of this, systems using external devices to represent and play a particular instrument, e.g. Rock Band [7], have been excluded from the analysis.

To be able to make distinctions between the different graphical interfaces the evaluation of each program has been made from asserting specific traits in a particular design area of the system and investigating if these traits correlate to certain design models. These traits, or attributes, are the parts of the system that is a part of one of the divided areas of the design models, e.g. the importance of the users ability to learn the system is a trait/attribute that correlates to Paternós usability evaluation. All the examined applications have been divided into subsets to simplify the definition of what kind of music software that has applied the certain attributes of the models used.

The design areas that has been chosen are emotional design and usability. By comparing these, quite extensive, models the functionality of the interface will be defined.

2.2 Interfaces

The softwares that have been tried are those that were frequently appearing while doing the research and softwares that I personally have used. To be able

to draw any conclusions on differences the chosen interfaces were those who were different from each other but still gave the user with the same desired possibilities; music creation. The interfaces that have been evaluated are:

- **Ableton Live** A music sequencer made for live performances. It can also be used to produce and mix finished arrangements [6].
- **Guitar-Pro 4** A tablature editor for guitar playing and arranging [12].
- **Cubase** Music arranging, editing and recording software [5].
- **Garageband for iPad** Podcast and music creator for OS X applications [4].
- **Tuxguitar** Like Guitar-Pro Tuxguitar is a tablature editor initally created for UNIX users [11].
- **Bloom HD for iPad** An experimental software specifically for iPad created by Brian Eno and Peter Chilvers [3].
- **Fruity Loops** Software that brands itself as a digital audio workspace, similar to *Cubase* and *Ableton Live* [1].

2.3 Models

To comprehend the basics behind a concept of *interaction design* one must consider not only the specific usage of an application and its end results. The cognitive process which takes place on the user side must also be taken in account, since it is the user that will experience the entire interaction.

A design model is a structure from which one could draw relevant observations when designing similar artifacts or systems. These are created either by trying to create an optimization within an area or by defining the main specific attributes that a system or artifact may possess.

When choosing models that are appropriate for the research it is important to specify exactly what traits you want to evaluate and against what criteria. Donald Normans emotional design model has been chosen because he advocates a user-centered design [17] which is an important aspect when analyzing traits in graphical interfaces for *specific users*. The main reason for acknowledging this model is that it is widely cited in books and articles regarding cognitive science[21], interaction design and user interface design[14].

Fabio Paternós usability evaluation has been chosen because it addresses the area of *specific use* in contrast to a *specific user*. It is not by definition a design model as it is a evaluatory tool for determining usability, but by combining these two the functionality and distinction between different music interfaces will be well grounded. The acknowledgment of Fabio Paternós usability evaluation is similarly to the emotional design model based on his work in the field of usability and that his book *Model-Based Design and Evaluation of Interactive Applications* is well cited in articles and books regarding interface design[19].

Donald Normans Approach to Emotional Design Norman defines a framework for establishing different kinds of emotional designs [18], and how these manifest themselves in levels of processing by people. These three levels are defined as:

- **Visceral** The level of initial impact of the design, the referral to the intended objects appearance.
- **Behavioral** Used to describe the overall experience the user feels when using a product or a system. It mainly deals with the look and feel of the system, and is possibly the most important aspect in relation to interaction design.
- **Reflective** What the user feels and thinks after using the desired object or system, and focuses on the part of how the user reflects after using said artifact.

Fabio Paternós Usability Evaluation In the book Model-based design and evaluation of interactive applications [20] Paternó describes how to determine usability and what it means to actually create such an application.

- **Relevance** How well the system serves the users needs, i.e. if the design reflects the specific purpose.
- **Efficiency** In what capacity the user can efficiently carry out preferred tasks in the system.
- **Learnability** Making it easy for the user to learn the system, and be able to remember how to use said system.
- **Attitude** The users personal feelings about using the application or system. This differs from the *reflective* aspect of Normans design model, since it addresses the users feelings about *using* the system and not the emotional impact *after using* the system.
- **Safety** Giving the user the opportunity to undo undesired actions.

3 Results

3.1 Division of Interfaces

Because of the range of applications in the market of music creation, there will undoubtedly be a fair share of similarities between the systems. By dissecting the main focus of the applications into smaller subsystems the differences will become more apparent.

To simplify the process to set up a design model for this specific area the interfaces have been divided into four separate groups. These distinctions have been defined solely for this research article, but could be applicable for future research in the area.

Three of these, *mimicking interfaces*, *theory based interfaces* and *advanced interfaces* are built up of a hierarchical structure where the advanced interface envelops the theory based ones as well as the mimicking ones. And similarly the theory based interfaces contains the mimicking. The fourth type is *experimental interfaces* and is included for the specific purpose not to be a orthodox representation of musical instruments. I have made the subdivision of these interfaces with regards to the philosophical argument that "The limits of my language mean the limits of my world" [22] meaning that a person will not be able to

make distinctions, choices or fully comprehend symbols which he or she cannot interpret or describe. This means the areas cannot be fully understood without the user knowing about the subdivisions and being able to interpret these specific parts.

3.2 Hierarchy of Interface Implementations

Mimicking Interfaces These are the most basic interfaces as they have a direct graphical imitation of the instrument which it is supposed to represent. Mimicking interfaces usually have very high usability as the user most likely already has a firm understanding of the instrument that it imitates [20]. This makes for easily created applications that are efficient in the sense that the user will have no problem in understanding the system.

Fig. 1. Screenshot of an iPad application with a mimicking interface[2].

Theory-Based Interfaces Software made explicitly for users who possess an understanding of the theory behind the musical instruments the theory must be possible to represent. These interfaces tend to have less of a graphical representation than the other systems, mainly presenting information to the user. These are very hard for users without any theoretical knowledge of the area to use, as it demands a prerequisite knowledge of the user, e.g. being able to read notes. A theory based interface is very often designed for a specific purpose, e.g. *Guitar-Pro* [12] and *Tuxguitar* [11] which are stand alone tablature editors.

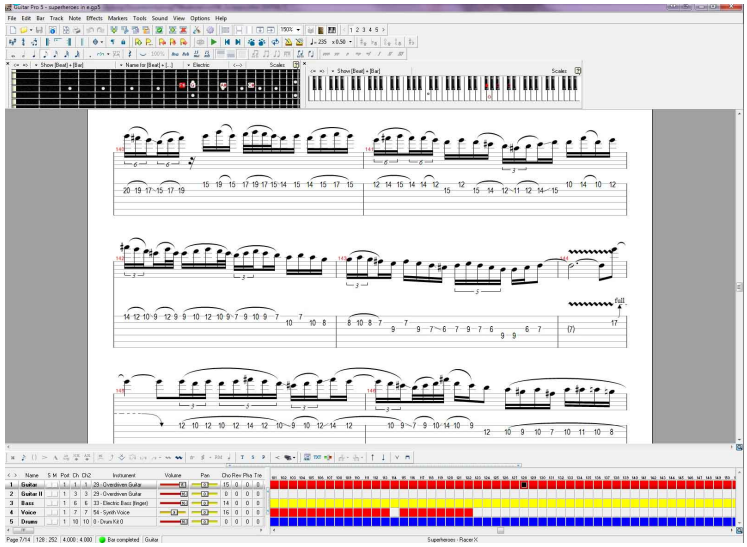


Fig. 2. Screenshot from Guitar-Pro, a theory based music interface[10].

Advanced Interfaces Systems which involve the previous interfaces and adds an array of different possibilities to manipulate the input from these more basic interfaces. By implementing the earlier mentioned interfaces and applying these methods to be used in one larger system, these software require the user to be an advanced user of similar applications. Not only does it require the previous knowledge of the theory based interfaces, but also implements aspects that are not traditionally connected to specific instruments, e.g. mixer boards, signal inputs and sampling.

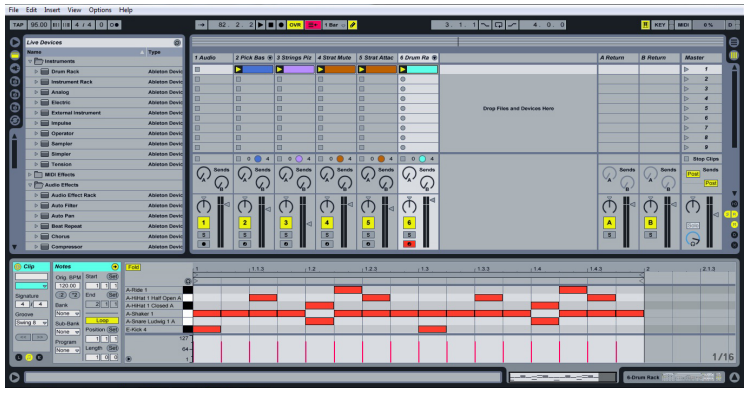


Fig. 3. A screenshot from Ableton Live, showing an advanced music interface[8].

Experimental Interfaces Systems that does not require any prerequisite knowledge of the system or the interface. Instead they allow the user to act in an unknown world and the result will *present itself* because of these actions. This type of contextual strategy is called being *thrown in* at the target and letting the user explore the world without knowing the end result [16].

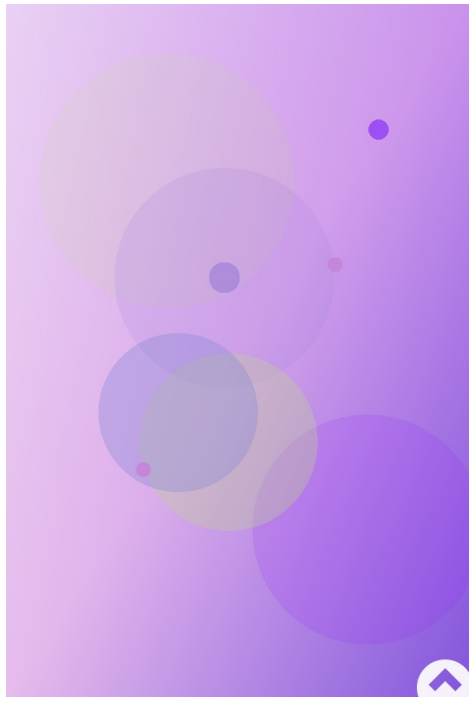


Fig. 4. Screenshot of start screen from the Bloom HD iPad application[9].

3.3 Analyzing Model Traits

By trying to define each attribute of the two design models within each of these divisions of interfaces the result of what traits are most occurring in every interface will become apparent. To provide a clearer understanding I have chosen to grade every interface after the design model traits on a scale from one to three where one is *not important*, two is *important* and three is *very important*, this is further discussed in the *further work* section of the article.

Mimicking Interfaces

– **Usability**

These interfaces have a high usability since the *relevance*, *efficiency* and *learnability* are very high. This is because the interface has the direct representation of an instrument that the user is expecting to be played. This removes the problem of teaching the user how to perform tasks in the system, since the symbol of an instrument is in the external, real, world.

– **Emotional Design**
While discussing the design aspects of mimicking software applications, the *reflectiveness* is rather low because of the same, previously stated removal of the symbol from the system. The user should not need to ponder the applications use any more than he or she would reflect on using an instrument. Similarly the *behavioral* and *reflective* aspects are not something that the design of the application should entail, but rather what the design of the real object should evoke in the user.

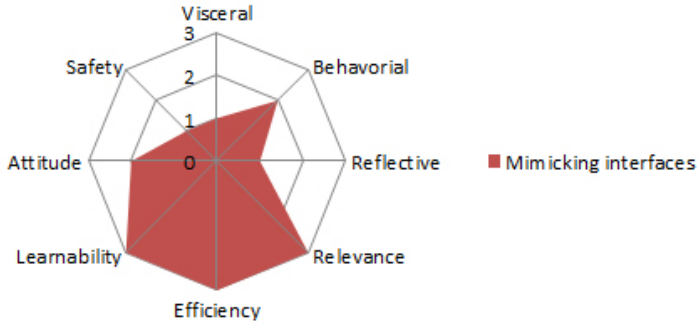


Fig. 5. Radar diagram of design traits in mimicking interfaces.

Theory-Based Interfaces

– **Usability**
In regards to creating music by notation and by music theory, the presentation of the information is of great importance. In this sense the theory based applications are inherently usable, since the only aspects that has to be acknowledged are the users possibilities to input the correct data and to let the application present the said data in a precise way. The systems *efficiency* and *relevance* is of great importance when designing these kinds of softwares, by the sheer idea of their implementation, that is its specified use, with few or no other opportunities for the user than using music notation.
These also use the representation of something that the user will have experienced in the real world, e.g. a music stand with a notepad, to let the actor comprehend the systems behavior, which leads to a high *learnability*. If the user has never been able to read or write notations or tablature the *learnability* will of course be very low, but this is not a part of what the application should teach the user, but rather an obligation on the user-end.

– **Emotional Design**
The design model traits are very similar to the *mimicking* applications, because this interface also mimics a domain that the user should have experienced before being able to fully use the system. The look and feel, i.e the *behavioral* aspect of Normans model, will be a direct depiction of what the user would encounter in the real world. This also removes a lot of the users *reflective* processing from the applications main use.
The *visceral* aspects will on the other hand differ quite a lot from mimicking interfaces as there are questions that the user will face that would not occur in the real word. For instance how do you write music notation on a screen instead of on a paper, and how do you regulate the pace? To get functionality like this to appear to the user, one must think about both *design* and *usability* and combine these to present a user-friendly functionality.

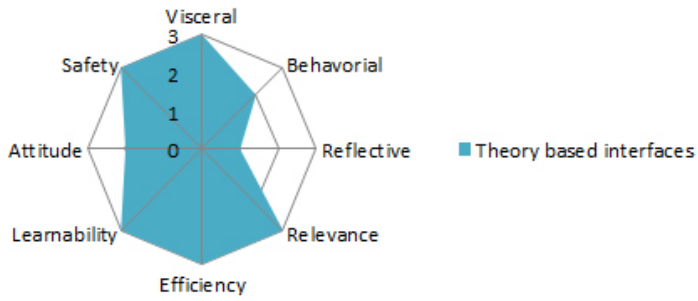


Fig. 6. Radar diagram of design traits in theory based interfaces.

Advanced Interfaces By enveloping both the *theory based* as well as the *mimicking* interfaces, these softwares must consist of very large parts of both models.

– **Usability**
The usability of a advanced system is quite complex because it implements large features of smaller systems. Therefore the *relevance* is very high, since it serves the needs of the user even though it is not specified for a specific use. The *efficiency* and *learnability* are also very important for an advanced interface, seeing how you could not implement such a vast array of functionality in an application if the user was not able to understand or remember how to be able to use it fully. A solution to this has been that the software companies provide external tutorials [13] so the user can learn the application without the intrusion of descriptory texts in the graphics. The *attitude* is not really addressed other than it want the user to be provided with a professional feel by the use of multiple systems in one large system. *Safety* is

also of grave importance as these softwares are intended to be used in large projects, and therefore the user should have the ability to correct errors that may arise during the work.

– **Emotional Design**
Presenting the user with such a catalog of functionality is difficult because you do not want the user to drown in the information presented. A lot of the advanced softwares therefore have graphics that are divided in different subgroups to let the users *visceral* processing determine what part of the interface is mostly suited for the preferred action at that point. The *behavioral* processing is that the system should be responsive and act accordingly to the exact input from the user. When presenting that much information at once, the interfaces are often implemented to show the flow of the music that has been created and by using symbols that would occur on real devices, e.g. mixer board knobs.
Reflectiveness should correlate to the *efficiency* aspect of the usability evaluation as the most important part of the system is to let an actor perform complex actions or sets of actions without any errors occurring.

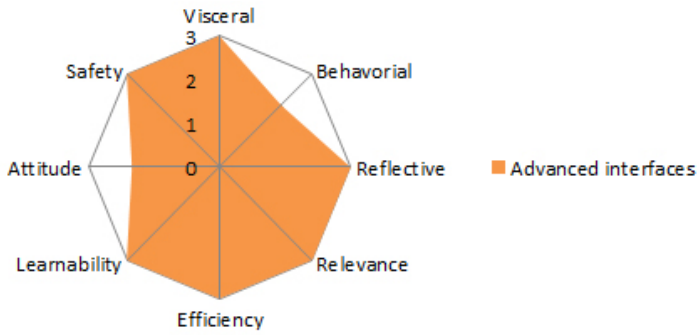


Fig. 7. Radar diagram of design traits in advanced interfaces.

Experimental Interfaces

– **Usability**
In experimental interfaces the usability is not of grave importance, but rather a step up functionality between the user and the application. The *safety* for instance, letting the user undo actions that are not wanted, may become obsolete. Maybe the undesired actions is something that the user wants, but does not know about? Both *relevance* and *efficiency* also become secondary because the preferred tasks may not even be defined and the users needs may not even exist as the system is something for the user to explore. What is important in usability aspects though, is the *attitude* attribute of the evaluation, the experimental applications want the user to act in the world.

They are designed to affect the *attitude* of the user, they want to evoke emotions rather than present the real world.

– **Emotional Design**
Design is the main focus of this type of software, it has no direct musical connection but seeks to instill some sort of emotional response from the user. The *visceral* processing should entice the user to explore the world more, it does not simply present the possibilities and then let the user act in the system. *Behavioral* experiences are often centered around originality and waiting for emergence in the system because of actions by the user. The interface also seeks to give the user a new experience, to present a new idea on how to create music and therefore has a very high *reflective* processing aspect.

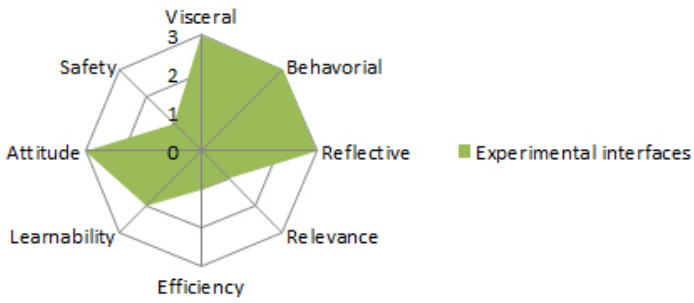


Fig. 8. Radar diagram of design traits in experimental interfaces.

4 Discussion

The results during this research shows a wide variety of interfaces and the different design model traits they possess. This leads to be able to define different softwares within the same domain into different subgroups, and be able to specify what it is that makes them differ from each other. One could use the results from this research in the start of implementing a music software and from the set grades of each attribute know what to focus on, based on what kind of interface that is to be created. As this was what was sought after in the research, the result is satisfactory, even though there are a lot more to do within the area.

4.1 Comparing Other New Medium Mappings

By examining the difference between the mapping of instruments to a 2D surface and the mapping of other 3D-to-2D mappings, you can make distinctions between the compared objects and why the result differ. If you take for example

a board game, e.g. chess, and look at how chess games are implemented within the same medium as a music creation software you will notice the lack of differences between the board game applications. This is because a chess game, or most other board games, has a predefined set of rules on how the user can manipulate the medium to reach a certain end point. The result is that which is important when playing chess, and the path to the goal is constrained by how the pieces can move on the board.

When implementing towards music, where the intended result is not always of the highest importance, you do not have to be restricted by specific rules to reach a goal. If someone would play an instrument for pure recreation, in a lot of cases the result will emerge during the playing. Instead these applications should provide the user with the set *possibilities* of stringing specific notes together rather than forming the application because of its constraints.

A similar mapping could be that of software made for creating graphics, where the basic idea is to provide a user with the possibility to paint on a canvas with specific brushes and colors. Similarly the result does not have to be specified beforehand but can emerge by the actions made by the user.

5 Further Work

The original goal of the paper was to propose models for implementing to a new medium. This is something that could be proposed in a future article. By constraining myself to a specific area of 3D-to-2D mappings a defined model for designing graphical interfaces for music software could be presented by dissecting further differences and by specifying each subdivision more thoroughly. One could in the future present an overall model to consider when implementing these sorts of interfaces.

The grading of the interfaces have been done as a step to further present the differences between the interfaces. Since this has been done solely by the researcher possessing deeper knowledge of these models it may fall victim to some subjectivity. With more time to research this area these traits could be more definitely defined by letting test groups grade these interfaces and systems.

References

1. Digital audio workstation, http://www.FLstudio.com
2. ipad keyboard, http://www.iphone-tips-and-advice.com/image-files/full-keyboard-ipad-vs-iphone.jpg
3. iphone and ipad applications, http://www.generativemusic.com
4. Music and podcast creating software, http://www.apple.com/ilife/garageband/
5. Music arranging, recording and editing software, http://www.steinberg.net/en/products/cubase/start.html
6. Music creating and producing software, http://www.ableton.com
7. Music video game, http://www.rockband.com/international
8. Screenshot of ableton live, http://www.flickr.com/photos/chromedecay/4482697887/lightbox/

9. Screenshot of bloom hd, http://pictures.recombu.com/news/M11772/1275383689_w670.jpg
10. Screenshot of guitar-pro 5, http://www.kyborgftw.com/picinst_Guitarpro5.jpg
11. Tablature editor, http://tuxguitar.herac.com.ar/
12. Tablature editor and multitrack editor, http://www.guitar-pro.com
13. Tutorials for using ableton live, http://www.ableton.com/tutorials
14. El-Bakry, H.M., Riad, A.M., Abu-Elsoud, M., Mohamed, S., Hassan, A.E., Kandel, M.S., Mastorakis, N.: Adaptive user interface for web applications (2010)
15. Hamman, M.: From symbol to symbiotioc: Representation, signification and the composition of music interaction. Journal of New Music Research, 28 (1999), No. 2, pp. 90-104 (1999)
16. Janlert, J.E.: A generic medium model for new media (2005)
17. Norman, D.: The Design of Everyday Things. Basic Books (1988)
18. Norman, D.: Emotional Design: Why We Love (or Hate) Everyday Things. Basic Books (2003)
19. Nunes, D.N.J.: Object modeling for user-centered development and user interface design: The wisdom approach (2001)
20. Paternó, F.: Model-Based Design and Evaluation of Interactive Applications. Springer (2000)
21. Thagard, P.: Mind: Introduction to Cognitive Science. A Bradford Book (2005)
22. Wittgenstein, L.: Tractatus Logico-Philosophicus (1921)

Enhancing XBMC by Using a Speech-Based User Interface

Marcus Lilja

Department of Computing Science
Umeå University, Sweden
id07mla@cs.umu.se

Abstract. This article investigates the current state of speech recognition and natural language processing to determine the maturity of the technology. Interaction today is increasingly natural, and natural user interfaces is one of the most important trends in technology today. The problems and advances in technology frames the discussion on whether or not it can be successfully implemented as a speech user interface for XBMC. One of the bigger issues is the environment with multiple simultaneous users. This factor and the results indicate that a speech user interface is hard to implement successfully in XBMC.

1 Introduction

One of the most important trends in digital technology today is the increased use of natural user interfaces for a wide variety of applications and products [1, 2]. Natural user interface (NUI) supports human abilities such as touch, vision, speech, handwriting, motion, and higher level processes such as cognition, creativity and exploration. Today most interaction with technology is through a command line interface (CLI) or a graphical user interface (GUI), which are artificial ways of communicating. The key assumption for NUI is that users should be allowed to interact with technology in the same way they are used to interacting with the real world in their daily lives [3]. Two examples of products that use a NUI are Microsoft Kinect and Apple Siri [2].

One example of a product that requires users to interact in an artificial way is the remote control for a TV. Since the introduction of the remote control in the 1950s [4], the idea of what a remote control is, has not really changed. One change is that an increased number of buttons have been added to the remote control due to the increased number of functions added to the TV, resulting in a more complex device. Despite the introduction of new buttons, studies have shown that channel up/down, volume up/down and the 0-9 digit buttons stand for 88% of all button presses [5]. When approximately 97% of all households in Sweden have a TV and as many as 86% have a computer with Internet connection [6], one could argue that TV is somewhat important, at least in Sweden.

Therefore XBMC Media Center (formerly Xbox Media Center) [7] can be used in conjunction with a regular TV to simulate a more complex device. XBMC

Bensch, Drewes, Gulliksson, Mejtoft (Eds.): Umeå's 15th Student Conference in Computing Science USCCS 2012, pp. 85–94, January 2012.

is a computer software that is used to consume different kinds of media such as movies, TV shows, music, pictures and Youtube. The idea of XBMC is simple enough, a GUI that is streamlined to be used on a TV. Interaction with XBMC is made through a regular remote control, which is similar to how one interact with a regular TV. However, because XBMC runs on a computer, the remote control could be a full-size keyboard as well.

This article focuses on how a speech user interface (SUI) can enhance the interaction with XBMC. The literature review shows that speech recognition and natural language processing (NLP) has improved due to advancement in their respective fields as well as computers increased processing power. However, the big issue with SUI today is that it has problems when multiple users speak at the same time, which has to be considered when implementing a SUI for XBMC. With this in mind the results of the article shows that XBMC is not ready for a SUI.

The following chapters contains a method briefly explaining how the material was gathered, a literature review of speech recognition and NLP, a short overview of XBMC's functionality, a discussion and a conclusion.

2 Method

The topic has been selected through a general review of NUIs, which has identified major ideas, issues and researchers. Most literature sources has been gathered with the help of Google, Google Scholar and academic databases like ACM Digital Library, Web of Science (ISI) 1945- and Academic Search Elite (EBSCO). All these sources has been carefully evaluated before being used as references in the literature review.

The literature sources in conjunction with the final topic selection has resulted in limiting the article to only discuss certain parts of NUIs such as speech recognition, NLP and multimodal interaction. In order to discuss the findings in the literature review, XBMC was added as a practical example. All usage examples of XBMC is derived from the author's own experiences.

3 Literature Review

The most important aspects of creating a successful SUI are speech recognition and NLP. Together, these two aspects allow the user to interact with a computer or device in a more natural manner. Although they can be used in conjunction with each other they are two separate fields of research. Speech recognition is mainly about interpreting the sound of the human voice to interpret what kinds of words are used. NLP can use the words interpreted from the speech recognition in order to determine what the words mean.

3.1 Speech Recognition and Natural Language Processing

Speech recognition as a research field has been around for a long time, and has seen substantial improvement over the last five decades. In the 1960s, sys-

tems could recognize small vocabularies ranging from 10 to 100 words. Twenty years later, using the hidden Markov model and the stochastic language model, speech recognition was improved to recognize vocabularies containing 1000 to an unlimited number of words. The last few years of research has resulted in the ability to implement very large vocabulary systems with full semantic models and text-to-speech synthesis systems [8].

There have been certain limitations regarding speech recognition [9]. The most noteworthy point in Shneiderman's article is that speech input and output both use short term memory for processing information. This was showed in a test where the users had to memorize mathematical symbols and afterwards issue a command. Schniederman compared users who issued the command with the mouse or with the voice. The users who used the voice command had more problems recalling the symbols than those who used the mouse for input. However, the same test showed that simple tasks were 30% faster when issued by voice than when the same command was made with the mouse.

Many SUI have single input and output sources where commands are issued by a speech interface and the responses are given through an auditory interface. A problem with this type of interaction is that users cannot process information in their own pace. If the user cannot recall the auditorial information given by the system and if this information is not repeated automatically the user will need to ask the system to repeat the information. This might be tedious if there is a lot of information to process. To minimize this kind of problem, multimodal interfaces have a better potential, where the auditory interface can be used in conjunction with a graphical representation [10].

Speech has proved to be more prone to produce errors in recognizing the correct command, relative to other input techniques [9, 11]. The reasons for the errors could be background noise, multiple people speaking or that people use words that the systems does not recognize. Lai and Yankelovich [11] discuss the problem of knowing a speech-based system's boundaries. When speech is used as input for a computer system, the user still relate to human-human interaction. This can make the user think that the system understands speech as good as a human does, and in turn tries to interact with commands that the system cannot handle.

A notable example of an application that uses speech recognition and NLP is Apple Siri. Originally Siri was a DARPA research project called CALO, which focused on creating an intelligent personal assistant. What makes Siri a notable example is that it combines speech recognition and NLP with a large knowledge base and contextual information. The information gathered and the interpretation of the question is then presented through an audiovisual interface, thus making it multimodal [2].

Even though speech recognition and NLP is often used in conjunction with each other, NLP is a separate research area. Watson [12] is an artificial intelligence computer system developed in a research project by IBM called DeepQA. Watson did not listen to speech, instead it processed information through written questions. In order to display the systems ability to process questions asked

in natural language, it participated in the quiz show Jeopardy. Watson played against two of the best human contestants and showed its potential by winning over both of them.

In the field of speech recognition most research efforts have been directed towards single user applications for speech recognition. However, one of the bigger challenges for speech recognition is the ability to distinguish between different users. If two users are using the same system simultaneous, as they would with a TV, it is hard to tell them apart [13].

3.2 Guidelines for Speech-Based User Interfaces

Almost all guidelines focus on the user when designing for speech-based user interfaces. Speech is learned implicitly at a very young age, not explicitly by education. Other interfaces needs to be learned in order to be able to complete tasks [10, 14]. With the user in focus, three important factors has to be considered in speech-based user interface design, state transparency, input control and error handling [14, 15].

3.2.1 State Transparency A speech-based user interface should always communicate the current state of the system [16]. This is generally done by providing the user with adequate feedback. If possible all feedback should be presented visually and auditory.

- **Provide feedback of system status in a correct way.** Visual feedback and auditory feedback should be combined to produce the best result, depending on the environment. The system should strive to provide information about sound levels, which makes it easy for the users to recognize whether or not the system is live. This area also includes making the user aware of when the system is unavailable for speech entry [16], for example, if search information takes a long time to process. Feedback is important but the user should never be overwhelmed with information, it needs to be crisp and clear in order to be easily processed [17].

- **At all times, the user should know what can be said.** Speech-based user interfaces need to be able to guide the user in a good way. This can be accomplished by phrasing questions like "Do you want to proceed?" or "Please say yes or no" [17].

3.2.2 Input Control When implementing an interface which is based on speech input it is better for the system be robust and always listen for input rather than having the user perform an explicit action to turn input on or off [17]. A user should not be forced to use a two-click protocol to interact with the system. If it is required to have some sort of filtering option to turn the input on or off, it is better to provide a push-to-talk function, because this reduces the cognitive load for the user.

3.2.3 Error Handling An area which consists of three different subgroups, error recovery, error detection and error correction. Error detection is a relative straightforward guideline which states that it should be a part of the system depending on whether or not automatic error recovery is part of the system.

Error recovery is needed in every system which deals with a speech-based user interface [16, 17]. There must always be some way to recover from a recognition error, this can be achieved by letting users undo actions or input.

Error correction lets the system handle incorrect input. This means that the system needs to have some sort of sense of what the users goal or intentions are in order to base its recovery on this [16]. It is always good if the system saves the input from the user so that it can be repeated to the user in the original form.

4 Overview of XBMC's Functionality

XBMC was released in 2003 as a media player and entertainment hub for digital media. Today the media player runs on all the mainstream platforms such as Windows, OS X and Linux. Even if it runs on most platforms available today, the graphical user interface has been created to run well on TVs. The software is completely free and open source, based on the GPL-license. The version used during this article was XBMC 10.1 "Dharma" [7].

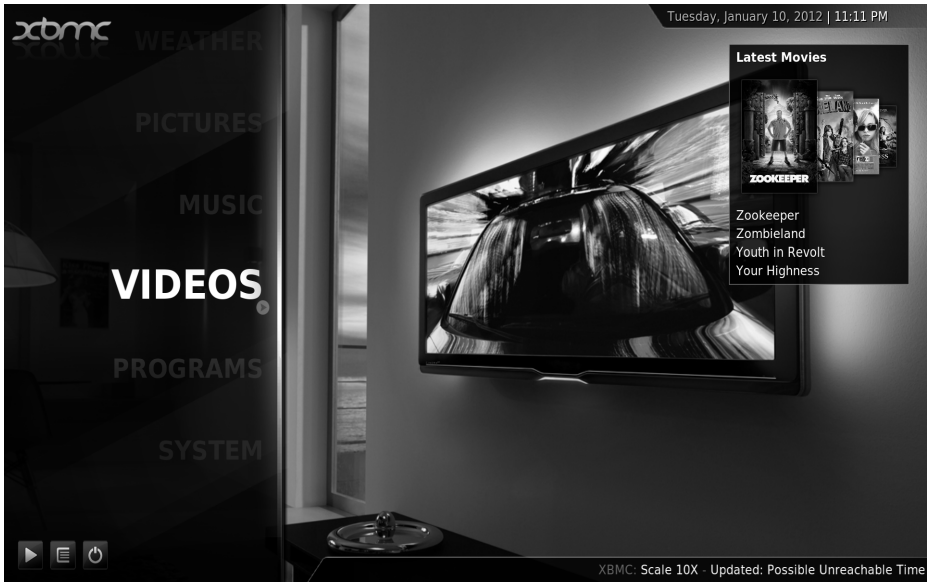


Fig. 1. The home screen in XBMC, which shows the main menu and recently added videos. The videos can be any video content, either movies or TV shows.

Fig. 2. The movie list in XBMC. It displays a visual navigation which is built up from information that XBMC gathers from Internet automatically.

The interface is very simple once the media sources has been configured. The main menu consists of categories for different media sources, movies, TV shows, music, pictures, programs and add-ons. XBMC automatically fetches information from the Internet, it can be ratings for a movie, different movie posters, fan art, links to trailers and general information about the movie [7].

XBMC can be controlled by different kinds of remote controls. The standard way of controlling it is by a regular keyboard. Other ways of interacting are also possible, for example using mobile phone applications, specific remote controls or a mouse. Some experiments on how to control XBMC with voice commands have already been produced [7].

Another feature in recent versions of XBMC is the add-on system. This enables XBMC to have features like SVT Play, Youtube, Vimeo and Gmail. It is also possible to watch live TV from regular channels if the hardware has access to a tuner card or if the internet provider provides IPTV [7].

5 Discussion

Judging by examples like Siri [2] and Watson [12], speech recognition and NLP as a technology has matured during the last few decades to be something to consider when designing new interfaces. Even if Siri is a good example of a SUI system that works well, it cannot be copied and work flawlessly on XBMC. This is because the environment for Siri and XBMC, is completely different. Siri

can be used anywhere and has only one user in mind, whereas XBMC is used primarily in the living room with multiple users. These two factors have to be taking in to consideration when implementing a SUI for XBMC.

5.1 The Problem with Multiple Users

Multiple users is probably one of the bigger challenges when implementing a SUI in XBMC. It is hard enough to filter speech from background noise and it is even harder to filter speech from speech [9]. This problem makes it hard to implement a SUI in XBMC. If the system cannot tell the users apart, it does not know who said what and what the commands are. If the SUI is adaptive and learns more about the users speech as the users use it, it needs to be told or know which user is currently submitting speech input.

To solve this problem the system can accept only one speaker at a time, use more sensors to decide who is speaking or demand that users manually let the system know who is speaking [18, 17]. The general guidelines recommends that the system is robust enough to allow continuous input, but if that is not possible the system can use a push-to-talk function [14]. This allows the system to know by who and when commands are issued.

One way to solve the problem with multiple simultaneous users, without forcing the users perform more actions, is to make the system more multimodal [15, 14]. If the systems has more microphones than one, it could easier determine who is speaking. But the problem is that the system needs to separate one voice from another, which is a problem with speech recognition today.

Another way of solving the problem would be to include cameras that can monitor the users. The cameras can spot who is looking at the TV, and make judgments about which user is speaking. Suppose two users are trying to use the system at the same time, even if the system can determine who is saying what, there is still a problem determining who the system should listen to. Increasing the number of sensors would increase the recognition accuracy [18], but still would not solve the problem with simultaneous multiple users.

5.2 The Living Room as an Environment for a SUI

The environment that XBMC is primarily being used in is the living room, which is a good environment for SUIs. This is due to the fact that there is a low risk of disturbing others and only minor background noise. Because XBMC already use a GUI it would be possible to follow the guidelines by presenting information both audibly and visual. Apple Siri is a good example of a device that allows the user to track the conversation, this shows how good the system understands the commands and how it interprets them. By allowing the user to both hear the information and track it visually on the screen, the cognitive load will be decreased by being less dependent on the short term memory.

5.3 Functions to Consider when Implementing a SUI for XBMC

The aim for implementing a SUI in XBMC is to enhance the usability, user-experience and not having to use the remote control on a daily basis. In order to active this goal, the speech recognition needs to be robust. If the user needs the remote control to let the system know when input are about to be submitted, the system does not really benefit from a SUI. Therefore the system needs to be robust and always listening for input [14, 17].

The SUI should not be involved in managing settings for changing the theme, configuring media sources, choosing video rendering software or other complex functionality similar to this. A SUI should primarily be involved in making it easy to browse for media information, control playback and search for information. These kind of functions would be suitable because simple tasks with a small vocabulary are better suited for speech input [5].

5.3.1 Browsing for Media Information Navigating and browsing content in XBMC is simple, relevant choices are kept in the main menu for easy access. Information about movies or TV shows are kept in lists, which is alphabetically order. For different buttons are used to navigate, up, down, select and back. If this is supposed to be done with the help of a SUI, it cannot complicate things. In order to browse movies the system needs to respond to more than just a specific command like "browse movies", it should be able to find keywords in the questions to find out what the user wants to do [16]. As the guidelines state, the design of the system should revolve around the user, not the other way around.

If a movie list is navigated with the help of a remote control the user can choose the speed in which it is browsed. There can be no exception if it is browsed with a SUI. In a response to a user wanting to "Browse movies", the system could reply "Would you like me to scroll faster or slower? Please say slower or faster" [16, 17]. This would help the user knowing how to control the flow of the information.

This can also be applied to sorting functions, stop scrolling, resume scrolling, selecting movies, playing selected items and other functions. If the system is adaptive and learns about the users behavior like Siri [2], it would become more natural over time.

5.3.2 Control Playback As mentioned before, the system would perform better if the system is robust and can listen for input all the time. This can become a problem if the user use sentences with specific keywords when a movie is playing. To make this easier for the system all input for the system could start with "XBMC". It would make sense to address the system with its name when asking it to perform an action, because it is closer to a human-human interaction. Another way around this could be that the system asks for confirmation if it is unsure of the input submitted by the user, "Are you sure you want to stop the movie? Confirm with yes or no." [16, 17].

Movies, music or TV shows should be accessible without needing to resort to browsing for the content first. An example could be "XBMC, play music from The Beatles" and the system would reply by playing music that matches the question.

5.3.3 Search for Information Search is not available from the main menu, it can only be accessed when browsing a media list. This feature could be enhanced by allowing the users search for any content that XBMC can get a hold of. Vimeo and Youtube are to media sources implemented in XBMC through the add-on system. If a user would ask the system to search for The Beatles, XBMC could return a list containing information from all the available sources.

However, when searching the Internet for information the system could take awhile to process this. The guidelines here are clear, a user should be able to cancel the research and get the local information only. It is also important to give feedback to the user about how long it might take to process information, so that the user can make a better decision about canceling the search or not.

6 Conclusion

For an interface to be completely natural it needs to be multimodal in order to better mimic human-human interaction. When humans interact by speech, it is not only the speech that is communicated. The communication is based on more than just the sum of its parts. This needs to be taken into account when designers are implementing SUIs on existing, as well as new, software. An SUI can enhance the interaction with XBMC to be simpler and less tedious than it is today. However, because XBMC is used in an environment which is problematic, it is certainly hard to create a perfect implementation.

References

1. Gates, B.: The power of the natural user interface. http://www.thegatesnotes.com/Personal/The-Power-of-the-Natural-User-Interface (2011)
2. Hearst, M.A.: Towards "natural" interactions in search user interfaces. Technical report, UC Berkley (2011)
3. Group, N.: Natural user interface. http://wiki.nuigroup.com/Natural_User_Interface (2011)
4. Letourneau, K.: From lazy bones to redeye: A brief history of the tv remote. http://morecontrol.com/2009/08/lazy-bones-to-redeye-a-brief-history-of-the-tv-remote/ (2009)
5. Eriksson, R., Sjögren, F.: Enhancing the user experience with new interaction techniques for interactive television. Master's thesis, Umeå University (2007)
6. Nord, L.: Mapping digital media: Sweden. A report by the open society foundation (2011)
7. XBMC: http://xbmc.org/ (2011)

8. Juang, B.H., Rabiner, L.R.: Automatic speech recognition – a brief history of the technology development (2004)
9. Shneiderman, B.: The limits of speech recognition. Communications of the ACM **43**(9) (2000) 63–65
10. Cohen, M.H., Giangola, J.P., Balogh, J.: 1. In: Voice User Interface Design. Addison-Wesley (2004)
11. Lai, J., Yankelovich, N.: Designing speech user interfaces. CHI '99: CHI '99 extended abstracts on Human factors in computing systems (1999) 124–125
12. Ferrucci, D., Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A.A., Lally, A., Murdock, J.W., Nyberg, E., Prager, J., Schlaefer, N., Welty, C.: Building watson: An overview of the deepqa project. Association for the Advancement of Artificial Intelligence (2010)
13. Patterson, E.K., Gowdy, J.N.: An audio-visual approach to simultaneous-speaker speech recognition. Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on 6-10 April 2003 (2003)
14. Rudnicky, A.I.: Speech interface guidelines. Technical report, School of Computer Science, Carnegie Mellon University (1996)
15. Ogden, W.C., Bernick, P.: Using Natural Language Interfaces. In: Handbook of Human-Computer Interaction. Elsevier Science Publishers B.V. (North-Holland) (1996)
16. Kamm, C., Walker, M.A.: Design and evaluation of spoken dialog systems. Proceedings of 1997 IEEE Workshop on Automatic Speech Recognition and Understanding (1997)
17. Mané, A., Boyce, S., Karis, D., Yankelovich, N.: Designing the user interface for speech recognition in applications. SIGCHI Bulletin **28**(4) (1996)
18. Koutras, A., Dermatas, E., Kokkinakis, G.: Improving simultaneous speech recognition in real room environments using overdetermined blind source separation. In: INTERSPEECH'01. (2001) 1009–1012

Graphical Representation of the Search Process of a Computer Chess Engine

Jonatan Pettersson

Department of Computing Science
Umeå University, Sweden
c03jpn@cs.umu.se

Abstract. The search produced by a chess engine (a chess application written specifically to analyze chess positions) may contain billions of moves, while the author of the engine have little control of how it proceeds. In this paper several ways of graphically representing and visualizing a chess engine search are examined and improved on. While the visualizations may help to a certain degree it was found that the complexity of such searches are a severe hinderance to clearly seeing the full picture. In conclusion the techniques examined in this paper appear to be useful, but further refinement needs to be done for a fully satisfying representation of a chess engine search.

1 Introduction

A chess engine (and most two player board games) ordinarily uses an alpha-beta search to find the best move. The end result is a best move with a calculated score and a series of moves that both players are expected to follow. The output result is presented in Table 1:

Table 1. Example result of a chess engine search.

Best move	e5d4	Calculated best move in the position
Score	105	What side is in the lead, in centipawns
Depth	12	To what depth this position was searched
Principal variation	e5d4 h2h3 e4g3 a2a3	Line both players are expected to follow

This is the only information you get out of the engine at the end of the search (with the possible addition of a suggested series of moves at every depth). It is not possible to examine what moves are being cut by the search, at what depth, and why. In short, why the engine came to the conclusion that this is the best line. The above makes it very hard to debug semantic problems in an engine, it is simply impossible to know where the engine went wrong and why.

The best, and usually only, way an author of a chess engine has to determine if a change is good or bad is to play an extensive number of test games against different opponents and review the result.

Bensch, Drewes, Gulliksson, Mejtoft (Eds.): Umeå's 15th Student Conference in Computing Science USCCS 2012, pp. 95–109, January 2012.

This paper will examine three different techniques for representing the search. These techniques will aim to help a chess engine author to get an overview and a better understanding of how the search unfolds, move for move. They should also aid the author in analyzing and evaluating changes made, that is what impact a certain change has in terms of how the decisions the engine does differs.

2 Background – Computer Chess

This section covers the different notations and computer chess implementations that will be referenced throughout the paper.

2.1 Mediocre Chess - A Java chess engine

Throughout this paper the Java chess engine Mediocre [1] will be used for code reference and example generation. The reason Mediocre was chose is because it is written by the author of this paper and hence source code and examples can be used freely. The code is also extensively documented and fairly easy to follow compared to other open source alternatives.

Code examples will be shown in pseudocode as the it is the concepts that are important rather than the implementation.

2.2 Chess Game Representation

In this paper the standard algebraic notation [2] will be used for representing moves. In Figure 1 the start position of a chess game is shown, with `files` (vertically adjacent squares) marked a through h and `ranks` (horizontally adjacent squares) marked 1 through 8.

In table 2 the different pieces and their moves are listed. So for example a queen on the square h4 capturing a piece on d7 would be written `Qxd7` (the captured piece is not listed), where Q specifies it is a queen moving and d7 being the destination square (the square the pieces moved from is omitted in this type of notation).

FEN Notation FEN notation, or Forsythe-Edwards notation [3] is a way of representing a position on a chessboard in a compact, one line string. This will be used below all diagrams (figures of chess positions) in this paper to verify the exact position.

The start position of a chess game would in FEN notation look like

```
rnbqkbnr/pppppppp/8/8/8/8/PPPPPPPP/RNBQKBNR w KQkq - 0 1
```

First part represents where there are pieces, starting with the eigth rank of the board and a new rank for every slash (`/`) character. Lower case letters are black pieces and upper case white, with letters corresponding to pieces as in table 2. A number represent empty squares horizontally, so the number 8 means an empty

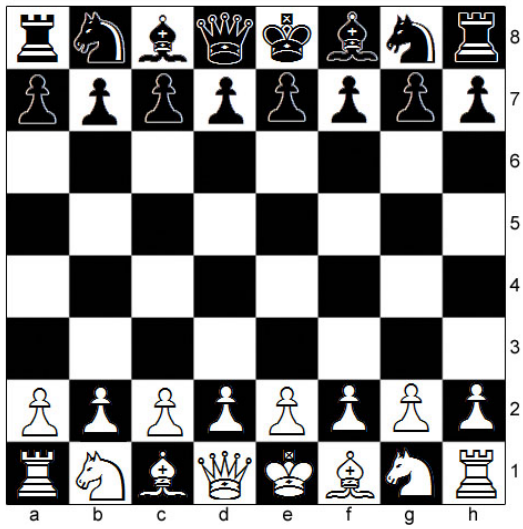


Fig. 1. Start position of on a chessboard

rank while 1p6 would be one empty square, then a black pawn and then six empty squares.

Second part indicates which side that has the move (`w` for white and `b` for black).

Third part indicates castling rights. `K` is for kingside castle (or short castle) and `Q` for queenside castle (long castle), uppercase for white and lowercase for black. A dash `-` means no side has the right to castle anymore.

Fourth part is the en passant square, that is the square behind a pawn that just moved two squares. A dash `-` if no en passant square.

Fifth part is the half move counter, that is number of moves (for black or white) since the last capture or pawn move.

Table 2. Move notations.

Piece/event	Notation	Example
King	K	Ke4
Queen	Q	Qe4
Rook	R	Re4
Knight	N	Ne4
Bishop	B	Be4
Pawn	-	e4
Capture	x	Qxe4
Short castle	0-0	0-0
Long castle	0-0-0	0-0-0
En passant	e.p.	dxe4 e.p.

Sixth part is the number of full moves (white and black moves) so far in the game.

2.3 Chess Engine Search

A chess engine search in general, and in this case Mediocre [1] in particular, usually consists of a search algorithm called `alpha-beta` at its base. Along with that a large number of techniques to reduce the number of moves search. For the purpose of this paper the techniques of Mediocre have been reduced to just two, `transposition tables` and `null moves`. These three concepts are briefly described in the following sections.

More specifically what is interesting for the purposes of this paper are cutoffs, that is when the search skips moves, which needs to be visualized to follow the search procedure.

Alpha-Beta Search Alpha-beta is a variation of the `Minimax` search technique[4]. Here follows a brief overview of its general idea.

Let us assume player A in a chess game has ten moves to choose from, and for each move player B has a certain number of moves to answer with. Player A wants to maximize the result and player B wants to minimize it.

To illustrate this in a simple example, replace the moves of player A with bags, and the moves of player B with dollar bills contained in the bags. Player A gets to choose a bag (the move of player A) and from that bag player B gets to choose the bill player A gets (the move of player B). As player A wants to maximize the score and player B minimize it, player A will choose the bag with most valuable minimum value bill.

In minimax search player A opens all bags and look at all bills, and then decides which bag to choose. That is every single move is examined in a position.

In alpha-beta search player A opens bag one and looks at all bills, let us assume this bag contains a five dollar, ten dollar and hundred dollar bill. As player B will minimize the score, player A will receive the five dollar bill. Now player A opens the second bag and looks at the first bill, which is a one dollar bill. Player A can now discard all other bills in this bag, as the best possible bill received is a one dollar bill. Even if the next bill turns out to be a thousand dollar bill, player B will minimize the result and choose the one dollar bill.

So in alpha-beta search, a large portion of all moves can be discarded without searching them, as the opponent would never allow them to happen (there are better options available already discovered).

What is interesting in terms of this paper is when player A can discard the rest of the bills in a bag, that skips the rest of the answers to his move. This is called a `beta cutoff` [4].

Transposition Table In chess it is possible to arrive to the same position with different sequences of moves. To avoid researching these positions a `transposition table` is used.

Recently searched positions are stored in the table, along with the best move found that time and its score. If the positions is reached on the board again the search stops and returns the move and score, hence avoiding searching it again [4, p. 171].

This will be referred to as a `transposition cutoff`, although it is actually not a cutoff (meaning moves are skipped), but rather a transposition to some position in the search tree already explored.

Null Move Null moves work under the assumption that having the ability to make two moves in a row is a significant advantage.

This is exploited in a chess engines by skipping a move, that is letting the opponent make two moves in a row. If the extra move does not enable the opponent to achieve a result better than one found in a previously searched branch, the opponent is too far behind and the rest of the moves in the branch are futile to search [5, p. 159].

This will be referred to as a `null move cutoff`.

3 Recording Data from Mediocre

To record data from a chess engine, in this case Mediocre [1], all moves will be recorded with traces in the code and marked with a rudimentary XML-markup [6].

There are also three possible events which are listed in table 3, as described in the previous section, which might skip moves, this also needs to be recorded with an XML-markup which is listed in table 3.

Table 3. Alpha-beta events and heuristics used in Mediocre[1]

Event	XML
Beta cutoff	`<betacutoff />`
Null move cutoff	`<nullmovecutoff />`
Transposition cutoff	`<transpose>`[id]`</transpose>`

As seen in table 3 the transposition cutoff will refer to the position it transposes into. Here follows a very short example on how this will be represented in XML markup[7]:

```
<search>
    <position>
        <id>1</id>
        <order>1</order>
        <fen>2k5/8/8/8/8/8/3P4/3K4 w - - 0 1 </fen>
        <move>Kd1</move>
        <white />
```

```
    </position>
    <position>
        <id>2</id>
        <order>2</order>
        <fen>2k5/8/8/8/8/8/3P4/3K4 w - - 0 1 </fen>
        <move>Kd1</move>
        <transpose>1</transpose>
        <black />
    </position>
</search>
```

Briefly explained in the following list.

1. `search` - Indicating a search of the engine
2. `position` - A reached position on the board
3. `id` - A unique id for the position for reference purposes
4. `order` - When the position was searched (a 1 indicates the position was the first to be searched)
5. `fen` - The position representation
6. `move` - Move played to reach this position
7. `white/black` - What side played the move
8. `transpose/nullmovecutoff/betacutoff` - If a cutoff occured `after` this move

4 Visualizing the Search of Mediocre

Visualizing the search of Mediocre will be done by adding prints to every significant event in the search and dumping it into a textfile which would then be parsed into the different visualizations covered in this section.

The searched moves are the obvious major piece of information, but also beta, null move and transposition cutoffs need to be noted.

4.1 Move Explorer

A way of organizing chess moves in a form of a tree that is browseable by the viewer. One example of this technique used is the chess database software ChessBase [8], where every move made by either side is given an exandable node and in that fashion exploring further moves is possible without overwhelming the user with information.

In ChessBase the technique is only used to explore opening moves, and also only relevant moves are listed. That gives a very limited set of moves to explore.

In terms of exploring an entire search tree of a chess program it would look like figure 2. Here every move can be explored by expanding or contractring the every move using the + or - to the left of the notation. In figure 2 a - means the user expanded that move, revealing all moves following it in the search, and +

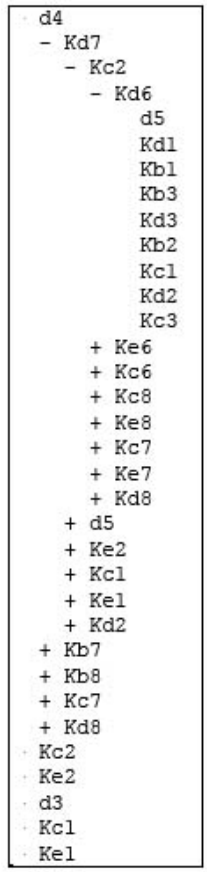


Fig. 2. Move explorer without enhancement

means it is unexpanded. A move with no sign in front of means the full depth has been reached and no moves are following it. The position in Figure 3 was used in in this search.

While this makes it possible for the user to explore all moves searched by the engine, it gives no information on wether the search was cut to a beta cutoff, null move pruning or any other techniques mentioned in previous sections. It also gives no relation to the code.

Improved Move Explorer To improve on the move explorer technique it will be color coded to distinguish between white and black moves, and also to mark cutoffs. Color coding is an effective technique, far more so than using shapes, but needs to be kept limited as to not distract the reader [9, ch. 10.1].

To encompass the transpositioning possibility, a node like representation will be used [9, ch. 10.4].

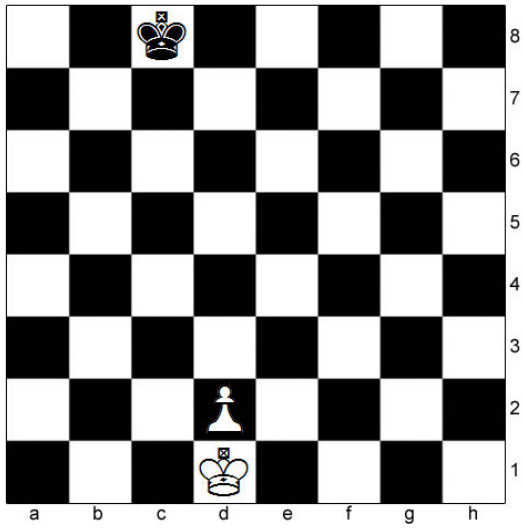


Fig. 3. FEN: 2k5/8/8/8/8/8/3P4/3K4 w - - 0 1

The result is in Figure 5, the position shown in Figure 4 was used here. The green lines shows transpositions to a different node, the red highlights indicate a null move cutoff happened on the move, and the blue a beta cutoff.

4.2 Combining ThemeView with a Chessboard

A very obvious way to graphically represent a search is to view it in combination with a chessboard. The ThemeView described in Search User Interfaces [9], which represents data as indendations on a plane, could very easily be adjusted and displayed on a chessboard, where the board would be the plane and the indentatation aligned along the squares. This could give a good notion of where moves are occuring in the search. A simplified way of doing this is shown in Figure 6.

The stacks are based on the position in Figure 3, with both first moves of both colors, that is white moves first then the response of black. Here it is clearly shown that white is preferring Kc2 in the position, while black most often responds Kd8.

In a more complex position this would not be as useful as many squares can be occupied by different pieces, reducing it to a simple statistic, while the actual moves are not known.

Extending ThemeView One way of solving the problem of several pieces moving to the same square is showing in Figure 7. Here every square gets one stack for each piece moving to the square, which gives both the combined sense

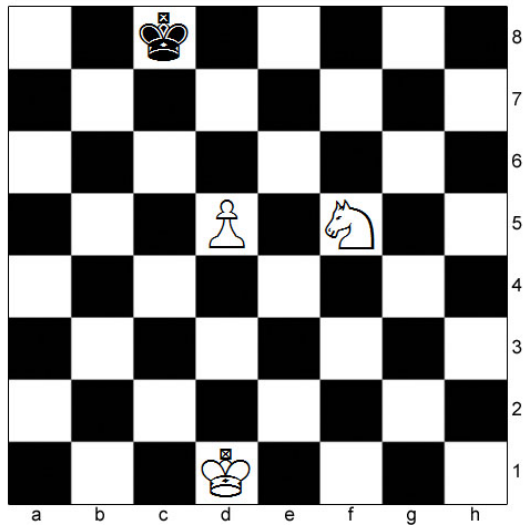


Fig. 4. FEN: 2k5/8/8/3P1N2/8/8/8/3K4 w - - 0 1

of popularity of the square in the search, as well as the more refined information of individual pieces moving.

The position used is the start postion, shown in Figure 1 and the grey staples are the possible knight moves, while the white are pawns. In this case the legend only defines grey as knight and white as pawn, but this could obviously be extended to more specific information identifying specific pieces.

In an application implementation of the extended theme view, the moving piece and the square it originated from could be highlighted when for example clicking a stack.

Here the number of moves are limited to one, any more would quickly deteriorate into unreadability. In Figure 8 an imaginary, and heavily simplified, search of a few moves is visualized. It quickly becomes obvious that any degree of viability is lost when deeper searches are considered. Possibly it could serve as a general feel for where action is taking place on the board, but as searches tend to spend quite some time on unfeasible moves [4, p. 171], the board would most certainly be heavily cluttered in the most straight forward positions.

However, as a one move statistic is could be used to compare different searches, especially those being altered by a change in static evaluation. Static evaluation evaluates the board at the end of the search to determine what side is ahead. If this evaluation would greatly award attacks on the king, this visualization would most certainly show an increased activity around the enemy king.

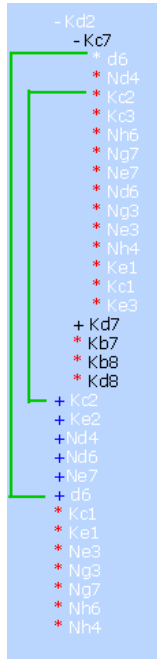


Fig. 5. Extended move explorer

4.3 Connection to the Source Code

To not just give a representation of the search but also help the programmer connect it to the code itself there needs to be a notion of where events occur in the code.

As described in *Software Visualization in the Large* [10], a good technique for this is the Line Representation. Line Representation is simply the code scaled down to give a view of the whole relevant section of code, with the possibility of zooming in on sections to see the actual text.

In Figure 9 the zoomed out version is shown, where relevant sections are highlighted. The highlights could include for example a part of the code covering null moves and another part covering transposition tables and so on. In Figure 10 the code is zoomed in on one of the sections.

The color coding in this example represents the different events, making it easy to connect them to the code.

For example when selecting a node in the Move Explorer representation, the code would immediately be zoomed in on where the event occured. Also with just moving over the nodes in the Move Explorer it would be possible to see where the cutoffs happen, and get a deeper understanding of how the code behaves.

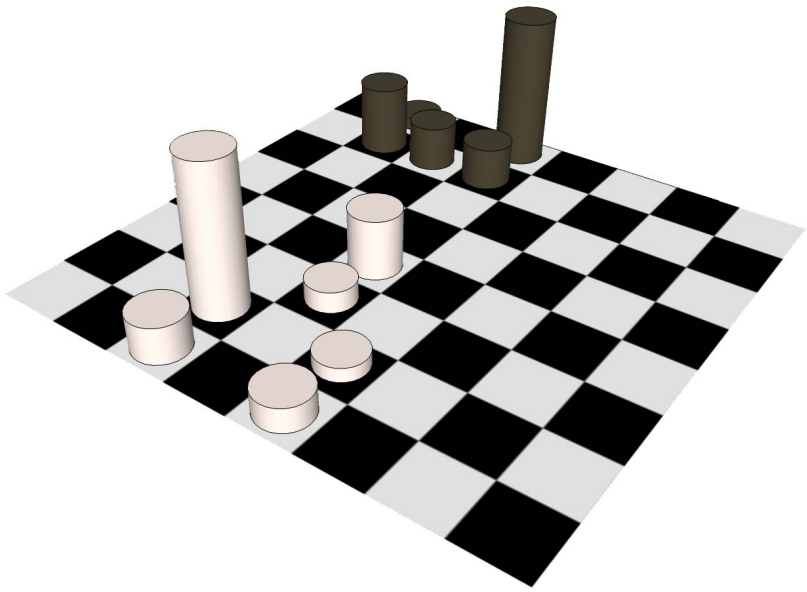


Fig. 6. ThemeView on a chessboard

5 Future Work

There are three main areas where future work would be especially interesting to be concentrated on.

5.1 Implementation

To really give the techniques discussed in this paper a real evaluation, they would need to be implemented in some form of tool, like a computer application. This way they can be freely explored and a more conclusive evaluation of how they work could be given.

This application would use the XML-representation suggested and hence be possible to use for any chess engine implementing the proper traces.

5.2 Evaluating the Techniques

To get a real idea of the usefulness of the proposed visualizations, given the subjective nature, a set of chess programmers should be interviewed and given the chance to discuss them.

A proper degree of certainty might require a large sample size, which would then be analyzed in terms of general agreements on usefulness and also incorporating any suggestions on implementation.

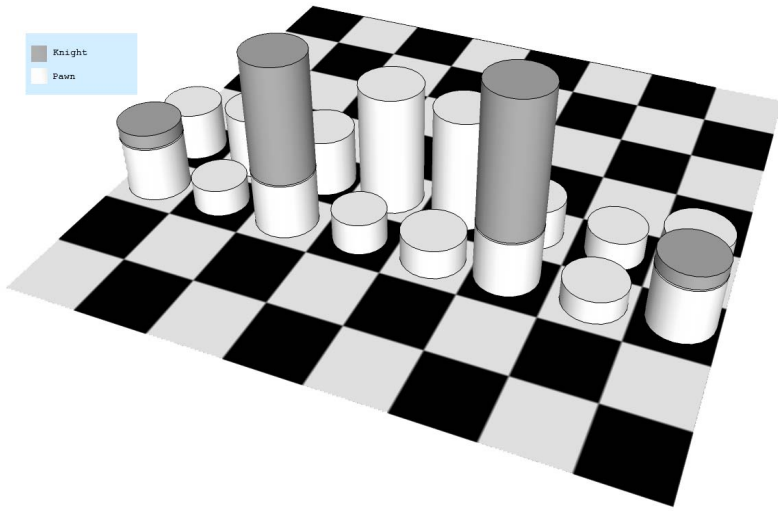


Fig. 7. Extended ThemeView with different pieces to same square

Fig. 8. Extended ThemeView on a larger scale

Fig. 9. Zoomed out Line Representation

```
        }
        System.out.println("<betacutoff />");
        return eval;
}
bestEval = eval;
if(eval > alpha) {
        eval_type = HASH_EXACT;
        bestMove = searchMoves[ply][i].move;
        alpha = eval;
}
```

Fig. 10. Zoomed in Line Representation

5.3 Broadening the Scope

As chess is a two player turn based game where an engine would use alpha-beta search to determine the best possible move. The explored techniques in this paper might be possible to extend any such game, a few examples might include Go, Checkers, Backgammon and more. While there are certainly adjustments to be made to fit differences in areas like search trees and heuristics, the general principles should still suffice.

6 Conclusion

The techniques examined in this paper appear very useful at first glance, especially in small scale with a limited number of colors used for highlighting. However the usefulness quickly decays with the increasing complexity of deeper searches. The Move Explorer approach is already at only five half-moves (one move for either white or black) quite large in its fully expanded state, and the expontential growth of the search tree size makes it very cumbersome to use.

That being said, the alternative of having only a list of moves, if that, is far worse. This seems to be a step in the right direction and certainly a useful tool if implemented correctly.

The same issues apply to the ThemeView visualization. In this paper it was limited to one or possibly two moves, any more than that makes it unreadable and hence useless. However, as a purely statistical tool, it is very useful for comparing different searches in terms of where the focus is generally located.

Finally the Line Representation of the source code is an excellent tool for a fast overview of the code and quick access to the relevant parts. Being able to connect the behaviour of the search with the corresponding piece of code appears extremely powerful. It could be argued that this holds for any application programming, but together with the Move Explorer and to some extent the ThemeView, the Line Representation is definitely something to look closely at for any chess programmer.

References

1. Pettersson, J.: Mediocre chess - java chess engine (2009) `https://sourceforge.net/projects/mediocrechess/`, version 0.34.
2. N.N.: Fide handbook `http://www.fide.com/fide/handbook?id=125&view=article`, accessed 2011-11-03.
3. N.N.: Portable game notation specification and implementation guide (1994) `http://www.chessclub.com/help/PGN-spec`, accessed 2011-11-03.
4. Russel, S., Norvig, P.: Artificial Intelligence A Modern Approach. (2003)
5. Marsland, T., Schaeffer, J.: Computers, Chess and Cognition. Springer-Verlag, New York Inc. (1990)
6. Ray, E.T.: Learning XML. O'Reilly Media (2003)
7. Walls, B.: Beautiful Mates: Applying Principles of Beauty to Computer Chess Heuristics. O'Reilly Media (2003)

8. ChessBase GmbH: Chessbase 11 chess database software (2011) https://www.chessbase.com/.
9. Hearst, M.A.: Search User Interfaces. Cambridge University press (2009) http://www.searchuserinterfaces.com/, accessed 2011-10-31.
10. Ball, T., Eick, S.G.: Software Visualization in the Large. (1996) http://research.microsoft.com/en-us/um/people/tball/papers/softvis.pdf, accessed 2008-01-22.

Evaluating Communicative Elements in Online Word Processing Programs

Katrin Svedberg

Department of Computing Science
UmeåUniversity, Sweden
katrin.svedberg@gmail.com

Abstract. Different graphical elements that help users to interact with an interface are viewed in this article. Based on the three interaction principles; visibility, affordance and feedback two word processing programs are evaluated heuristically. By using these principles a good human-computer interaction and a high level of usability can be achieved in an interface. The word processing programs that have been viewed use these concepts when they mediate to the users how their graphical elements and functions should be used.

1 Introduction

Word processing programs uses and also depends on graphical communication. This article focus on two such programs and how communicative elements and metaphors are used.

People are used to see graphical elements on a daily basis, maps in the subway, timetable and elements in our digital artifact, such as cameras, mobiles and computers. Do people understand symbols and metaphors as a result of trial and error or do systems and their elements tell us something even without us reflecting on it? The purpose of this article is to find out contributing factors that make us understand the graphic elements used in interfaces, especially in word processing programs. Metaphors are used in many systems and in many different ways, they can be found in both human-computer interactions as well as in human-human communication [1]. If a system has communicative abilities users are able to interact more efficiently.

Based on the design principles of visibility, affordance and feedback, communicative elements in Google Docs and Microsoft Word Web Apps start pages will be evaluated.

2 Method

Two word processing programs, Google Docs and Microsoft Word Web App, where chosen to be evaluated by an expert heuristic inspection. Based on the three interaction principles of visibility, affordance and feedback, these word processing programs usability and communicative elements are viewed. The principles chosen are important factors when it comes to creating systems with

Bensch, Drewes, Gulliksson, Mejtoft (Eds.): Umeå's 15th Student Conference in Computing Science USCCS 2012, pp. 111–120, January 2012.

high usability and systems with a natural human-computer interaction [2]. As metaphors are an important part in graphic design, and also a contributing factor to the communicative elements used in word processing programs, metaphors in Google docs and Microsoft Word Web App will also be viewed.

Google and Microsofts word processing programs where chosen due to the common factors that they both are online, free of charge and that they are made by established companies. The interface of these programs start page and the communicative elements presented there, is what is being reviewed. This limitation implicate that the depth and structure of the programs is disregarded, and only the elements from the first view of the programs are evaluated.

3 Usability

Usability can be seen from three key words [3]; *Effectiveness*, if the user finds what he or she is looking for. *Efficiency*, if it is hard to find the information. *Satisfaction*, in what degree the user gets a positive feeling from using the product. By applying these concepts to an interface design, it is more likely that the user gets a positive feeling when using the system. But if the users do not find, or are having problem finding the tools that they need in the graphical user interface, they will not get a positive feeling and are likely to dislike the system and therefore leave [4]. And, as Nielsen [4] points out, leaving is the first line of defense when users encounter a difficulty. A system with a good user interface design makes the interaction between the user and the computer easy, natural, and it also makes the user comfortable using the system [2]. Usability is an important part for any system if the users are to get a good experience.

Some important aspects, when it comes to user interface design, are the principle of visibility, affordance and feedback [2], [5]. These principles help to ensure a good human-computer interaction [2], they are also a contributing part in creating usability.

- **Visibility.** The principle of visibility states[2]; It should be obvious what a control is used for. In order to create visibility, the system also needs to keep the user informed over the systems current state and give correct feedback at the right time [2]. Visibility is in many cases depended on a systems use of feedback, and the result of an action should be apparent [5]. Visibility is an important part of word processing programs and their use of metaphors; it should be obvious what a symbolic metaphor is used for.
- **Affordance.** The principle of affordance states [2]; It should be obvious how a control is used. When affordance is used in the right way, the user should know what to do just by looking at the different graphical elements [5]. An example for how user interfaces show affordance is when the user hovers a symbol with the computer mouse, the symbol then indicates that it is clickable, the background color might change or some other type of indication. By doing this, the system tells the user how they can interact with a symbolic metaphor in a user interface.

– **Feedback.** The feedback that the user gets from the system is a contributing factor to how the system is perceived, and as the principle of feedback states [2]; it should be obvious when a control has been used. The interaction that take place within an interface can be seen from Donald A. Normans The Action Cycle [5], (Fig. 1).

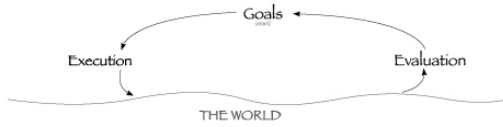


Fig. 1. Donald A. Normans The action cycle.

The user starts with a goal, the user then execute an action towards the interface, hoping to achieve the goal, and finally evaluates the results and if the goals that were set up in the beginning has been achieved. If a user use a button with a symbolic metaphor the user can get double feedback, if the system both execute the function and also displays the button that were used in a way that shows that it has been used or are in use.

3.1 Designing Usability in Graphical Systems

When a designer creates a graphic system there are a lot of things to take into consideration [6], for instance how the system should appear and be perceived, what the main target consumers expect in the new product, and the quality of the final product. Norman [5] points out that the fundamental principles of designing for people are to provide a good conceptual model and making things visible. A good conceptual model should allow the user to predict the effect of an action.

When developing a system with good communication ability it is essential to define the goals and means of the interaction as early in the development as possible [7]. By specifying how the system will be used and by narrowing down the target group the task becomes easier and in order to create meaningful interactive design the users needs of the service must be in focus [2], [7]. It is also important for a designer to be able to present the essential parts of the system without risking the loss of other functions [2]. Users that can attend their goals with minimum frustration are more comfortable with the interaction and satisfied users are shown to be more productive [2].

4 Communicative Metaphors

Interaction is based on communication, whether the interaction takes place between human-human or between human-computer. In order for the communication to work, both parties are expected to know the basics of communication, and

human-human as well as human-computer uses metaphorical references when they are communicating [1]. By understanding the human perceptive and cognitive skills it becomes easier to create a system that enables humans to make decisions [8].

Many of the early computer systems required the user to adapt to the system but this concept is not accepted today. Now it is the other way around, the system is expected to adapt to the user, but many users still find it hard to learn and interact with computer systems [9]. By creating a system that the user finds pleasurable, it is more likely for the system to be accepted [2].

4.1 Metaphors Value in Human-Computer Interaction

A concept often used in interface development is metaphors, it helps users remember and understand different concepts [1]. Metaphors can be used to make a user familiar with an unfamiliar subject as well as be used to organizing data [7]. Humans process traditional information in a way that transforms external events into internal symbolic representations and for human-computer interaction the interface is "outside" of cognition and is only brought inside through symbolic transduction [6]. Therefore it is important, and at the same time hard, to create metaphors that are perceived in the same way for different users. Between tracking reality and allowing symbolic freedom of thought, designers need to find a balance that supports the users thinking [10].

If a symbolic metaphor fails to convey its purpose it can lead to a problematic state that enquire the user to create its own metaphor in order to understand and interact efficiently with the system [11], [7]. As one important psychological principle states [2], it is easier to recognize something than to recall it. By using this concept in a user interface design the users can recognize what they want instead of having to recall it [2].

User interfaces are artificial environments that users can understand due to our experiences from the physical world, and all graphical elements generated by a computer can be perceived as metaphorical [2].

4.2 Metaphors Usability

When it comes to creating metaphors with high usability the three key words mentioned earlier are important to take in consideration. As the first word, *Effectiveness*, points out, in order for the user to find what he or she is looking for, a symbolic metaphor needs to be designed in a way that cannot be misinterpreted. The second word, *Efficiency*, whether a function is hard to find, can be applied to the way metaphors are structured in an interface. By matching together symbolic metaphors the user is more likely to find the function that he or she is searching for. As for the third word, *Satisfaction*, whether the user gets a positive feeling from using the system, is a way of looking if the user likes the metaphors and understand their representations.

Users are able to interact more effective with an interface if the design helps the user to focus on the right thing at the right time [2], therefore good metaphor design is an important part when it comes to creating usability in a system.

5 Word Processing Programs

Word processing is a program that can create printable materials, a program where the user can create, edit, format and compose material such as texts and pictures. The name, *word processing*, was created in the late sixties by IBM and this type of program was one of the first programs in the personal computer.

Google Docs and Microsoft Word Web Apps are both free, online word processing programs, Google Docs was made available in 2007 and Microsoft Word Web Apps in 2008. By using online word processing programs users can reach their documents from all over the world if they just have access to a web browser. Both of these programs allows the user to create different types of documents such as presentations and spreadsheets but this article focus on how communicative elements are used in their word processing programs. When evaluating the communicative elements in Google Docs and Microsoft Word Web App, the three principles, visibility, affordance and feedback will be in focus. In Fig. 2 is an overview of Google Docs and in Fig. 6 Microsoft Word Web App.

5.1 Evaluating Google Docs

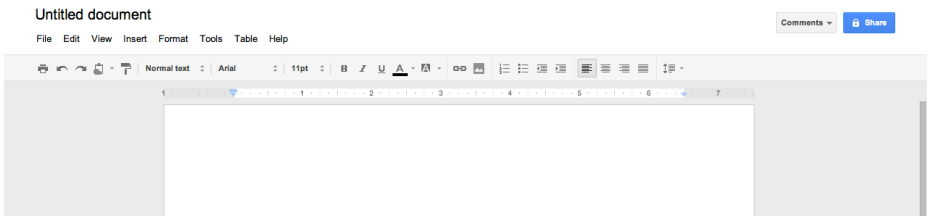


Fig. 2. Overview of the start page in Google Docs

– **Visibility.** In Google docs, when a user hovers over a visual metaphor with the computer mouse, a text is shown clarifying the metaphors function. This is called metaphorical language [2] . The explanation is centered underneath the symbol and it also shows a hotkey for that specific function. The use of hotkeys is a wanted function for users that uses the system regularly [2]. By explaining the different symbolic metaphors, the visibility increases in the system and fewer mistakes occur. Another visibility aspect that Google docs uses is the use of vertical lines to create alignment in their graphic, matching together functions that are similar or in the same genre (figure

3). This creates a high level of efficiency, which leads to better usability in the system. The users are more likely to find what is being search for when similar functions are grouped.

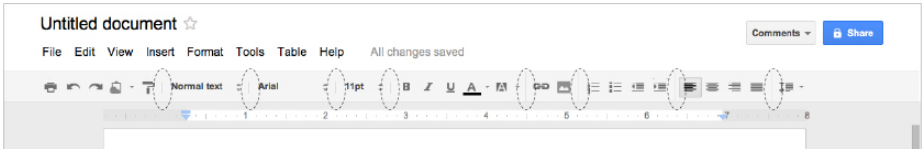


Fig. 3. Vertical lines in Google docs creating alignments

– **Affordance.** When hovering a symbolic metaphor, Google docs indicate the symbols affordance by making the symbol appear in a slightly different way. By doing this, the user understands that the symbolic metaphor is clickable.

  When a symbolic metaphor refers to a physical object, such as the symbol for printing (farthest to the left in google docs menu bar), the symbol itself has a high level of affordance. By creating metaphors from physical objects the users can create a god conceptual model that helps them to predict the effect of an action.

– **Feedback.** When a symbolic metaphor is clicked the user gets immediate feedback to what function they have in use. Google docs illustrate that a function is in use by making the symbol look like it is pushed down, as if it was a physical button(figure 4).

Fig. 4. Feedback when a function is in use in Google docs.

An important function in word processing program is the save function and in Google Docs it is done automatic. Google regularly saves the document and inform the user of the current saving state, illustrating it with text (figure 5). The placing of this function is in the top of the page and near the center, the text is grey and melts in to the background. This placement provide the saving informatin to the user without distracting.

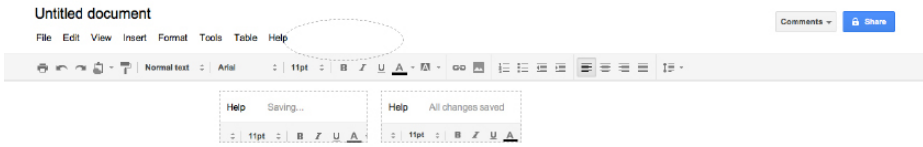


Fig. 5. The different saving states in Google Docs.

5.2 Evaluating Microsoft Word Web App

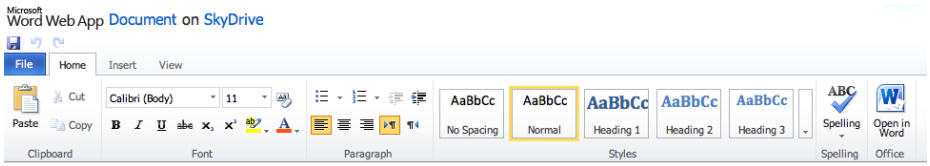


Fig. 6. The startpage in Microsoft Word Web App.

– **Visibility.** Microsoft Word Web App also uses metaphorical language, when hovering a symbolic metaphor. Their explanation is slightly right aligned and appears with a delay. The delay that they have create a high level of usability due the fact that users find it annoying when a system constantly explains it self [2]. With this delay, users that know the system can interact without being disturbed by explanations.
Microsoft use vertical lines in their graphic (figure 6) in the same way as Google Docs. But Microsoft clarifies it even more, describing what kind of alignment it is by printing it out underneath the different sections, such as Font, Paragraph and Style. This creates a higher level of visibility and also a higher level of efficiency, making it more obvious what the different symbolic metaphors are used for.

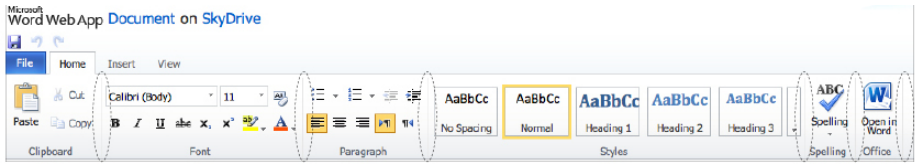


Fig. 7. Alignments in Microsoft Word Web App.

- **Affordance.** Microsoft Word Web App illustrates the symbolic metaphors affordance in a similar way as Google Docs, indicating that the symbolic metaphors are clickable by making them appear in a different way when the user hovers the symbol. Both Google and Microsoft achieve satisfaction from the users by making the different symbols indicate their affordance, the users easily understand how to interact with the different symbols.
- **Feedback.** When a symbolic metaphor is in use Microsoft illustrate this by making the symbols background color more prominent (figure 7). The user thereby gets clear feedback to the operations they have performed.

Fig. 8. How Microsoft illustrate a symbol in use.

When it comes to the function of saving, Microsoft Word Web App has a save button, it is placed in the top left corner and it has a floppy disk as a symbolic metaphor. When the save button has been clicked the screen fades out and a saving symbol appears in the center of the page, indicating that the document is being saved (figure 9).

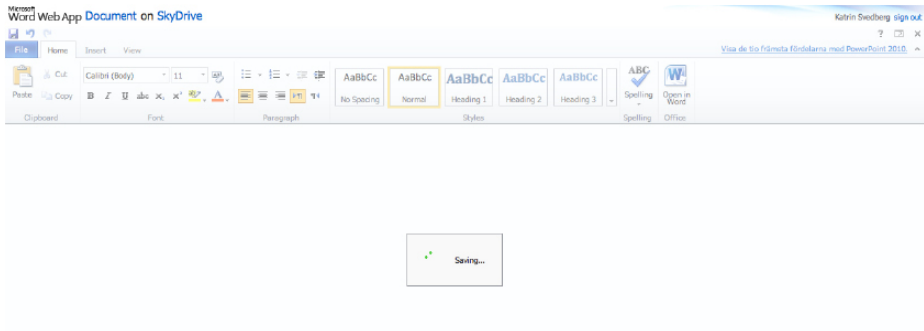


Fig. 9. The saving process in Microsoft Word Web App.

The way that the user gets feedback from saving cannot be misinterpreted, due to the fact that the entire screen is changed and the user cannot proceed his or her work until the saving process is completed. Microsofts saving functions has a high level of effectiveness as the saving process cannot be misperceived. The user gets visible feedback to the preformed operation and high usability is achieved.

6 Discussion

Both programs use the principle of visibility by giving the user explanations when the computer mouse hovers over a symbolic metaphor. But is this perceived in a positive way for the user, or is it a function that the user should be able to turn off once they have learned the system? Microsoft has a delay before their explanations are shown which gives the user an interface with fewer disturbances. But why is not the symbolic metaphors visibility high enough for them to be able indicate their functions without explanations? Some of the aspect that affect, are that symbolic metaphors are perceived in different ways depending on the users experience, cultural aspects, and also the users age.

Both Google Docs and Microsoft Word Web App use vertical lines to create alignment in their graphic and matching together functions that are similar or in the same genre. By doing this, they create a better visibility for the system and the layout becomes more structured, which makes it easier for the user to perceive the interface [2]. This leads to the fact that the user finds that the system has a higher level of usability.

When programs are updated it is common that some symbolic metaphors are redesigned in order to improve the designs communicative ability, but the task of redesigning established metaphors is hard. For example, the floppy disk, used by Microsoft (and many other word processing programs as well), is a symbolic metaphor that is no longer a symbol that a user can look at and immediately understand what will happen if it is used, the floppy disk is rather a symbol that is associated with a function. An example of this was when a 70 year old woman was learning to use Microsoft's word processing program, and when she was asked to save her document she said; for that I click the television button. When a symbolic metaphor loses its metaphorical function it may affect the entire systems' usability.

Affordance is used in a similar way in both Google Docs and Microsoft Word Web App, when the computer mouse is hovering a symbolic metaphor. Only then is it illustrated how the user can interact with the symbol. Just by looking at the systems it is not clear which parts that are clickable.

7 Conclusion

The three principles, visibility, affordance and feedback are used by both systems that have been reviewed in this article. By using these principles in user interface design, a good human-computer interaction can be achieved, and thereby a high level of usability. Google Docs and Microsoft Word Web App use metaphorical concepts that communicate to the user through the use of symbolic metaphors.

The use of metaphors in an interface can make the human and the computer understand each other better. By designing a system with good communicative abilities the user is able to interact more effortless with the system. When creating the different communicative parts in an interface the system as a whole must be in focus and at the same time, the objects should be able to stand for themselves.

References

1. Marcus, A.: Human communications issued in advanced uis. Communications of the ACM **36**(4) (1993) 100–109
2. Stone, D., Jarrett, C., Woodroffe, M., Minocha, S.: User interface design and evaluation. Elsevier, San Francisco (2005)
3. 9241-11, I.
4. http://www.useit.com/alertbox/20030825.html
5. Norman, D.A.: The design of everyday things. Basic Book, New York (2002)
6. Philip Kotler and G. Alexander Rath: Design: a powerful but neglected strategic tool. The journal of business strategy (2007) 16–21
7. Shedroff, N.: Information interaction design, a unified field theory of design. Technical Report 94107-1814, Vivid studios, San Francisco, USA (1994)
8. Maneesh Agrawala and Wilmot Li and Floraine Berthouzoz: Design principles for visual communication - how to identify, instantiate, and evaluate domain-specific design principles for creating more effective visualizations. Communications of the acm **54**(4) (2011) 60–69
9. John D. Gould and Clayton Lewis: Designing for usability: Key principles and what designers think. Communications of the ACM **28**(3) (1985) 300–311
10. Janlert, L.E.: The evasive interface: the changing concept of interface and the varying role of symbols in human-computer interaction. Technical Report LNCS 4550, Dept. of Computing Science, Umeå University, Umeå, Sweden (2007)
11. Hutchin, E.: Metaphors for interface design. Technical Report ICS Report 8703, Institute for Cognitive Science, University of California, San Diego La Jolla, California (1987)

How Do You Create a Worldwide Obsession?

Sandra Valett

Department of Computing Science
Umeå University, Sweden
id07svt@cs.umu.se

Abstract. Why is the world obsessed with playing Angry Birds? Why do we spend so much time trying to hit green pigs with birds? This article looks closer at different aspects of game design and the use of addictive elements. The aspects are then used to evaluate four successful games to try to find what actually makes them successful.

1 Introduction

Why do the world keep throwing birds at green pigs? Almost everyone knows it is because the pigs have kidnapped their eggs, but why do we know this and why do we keep playing? Is it the heartfelt backstory that makes us play, is it the taunting of the laughing pigs that we fail to hit or is it the cunning skill of its game designers?

The idea of an interactive games goes back as far as 1947 when Thomas T. Goldsmith Jr and Estle Ray Mann filed a request for a United States patent on what they called a "cathode ray tube amusement device"[1]. In 1952 A.S. Douglas created the first graphical computer game, OXO, a virtual Tic-Tac-Toe game as a part of his PhD degree [2]. In 1958 William Higinbotham created "Tennis For Two" for the Brookhavens National Laboratory's visitor day and in 1962 MIT-students Martin Graetz, Steve Russell and Wayne Wiitanen created the game "SpaceWar!", which often is referred to as the very first video game. The first ever commercial home video game console, the Odyssey, was released by Magnavox in 1972[2] and this was the real start for video games as a popular form of entertainment which today, some might say, is a part of the modern day culture.

But what makes a successful game? What is it about a game that can create a world obsession? First of all the player will have to want to play it, and to achieve this, according to Richard Rouse [3], a game have to fill at least one of the eight basic wants of the player. Playing games has always been about the entertainment, about having fun, but what makes a game fun? Nicole Lazzaro's research shows that gaming has four keys to fun [4] and that a successful game should include at least three. But the player will not only have to want to play the game in order to make it successful, it will need to keep the player interested enough to keep playing the game. In order to do this almost every game includes addictive element [5]. The addictive elements are not always obvious to the player and can be something as simple as including a high-score. A high-score works as

Bensch, Drewes, Gulliksson, Mejtoft (Eds.): Umeå's 15th Student Conference in Computing Science USCCS 2012, pp. 121–131, January 2012.

a constant push to make the player to want to do better and allows players to compare their skills and create an competitive feeling in the game. By evaluating these different parts in four successful games, Angry Birds, Farmville, StarCraft and Tetris, the aspects and needs they fill and which addictive elements that has been incorporated, this article will result in a formula of four important questions to guide the game design to a future success.

2 Game Design

An important part of designing a successful game is to give the player what they want, but this requires a knowledge of what the player want, something the players might not even know themselves. So instead many game designers look at the most popular games at the moment and thinks that this is what the players want. This however often result in a game that is basically a clone of last years game, which will never be a huge hit [3]. Instead successful game designers look at their own favorite games and asks themselves what part it is that makes the game so popular, and how they can use this knowledge to create something new.

Famous video game designer Shigeru Miyamoto, manager of the Nintendo Entertainment Analysis and Development, once said: *"But when I come to think more on it, the biggest reason it has become that popular is Mr. Tajiri, the main developer and creator of Pokemon, didn't start this project with a business sense. In other words, he was not intending to make something that would become very popular. He just wanted to make something he wanted to play. There was no business sense included, only his love involved in the creation. Somehow, what he wanted to create for himself was appreciated by others in this country, and is shared by people in other countries. And that's the point: not to make something sell, something very popular, but to love something, and make something that we creators can love. It's the very core feeling we should have in making games."*[3] This chapter will look closer on the reasons we play games and what it is that makes a game fun.

2.1 Why Do We Play?

When designing a video game an important question is: Why will someone play this game? What do we play games at all? According to Richard Rouse there is eight different reasons to play games [3].

The Challenge People have always enjoyed to be challenged, we challenge our minds with physical puzzles such as Rubic's Cube and our bodies with different sports. For single-players it is often the challenge that the game provides that is the biggest reason to play. Challenges is learning opportunities and when overcome they give the player a knowledge to be used in other situations, in games this can be how to navigate in the environment or how to use tools. But games can also give knowledge to be used in real life such as problem solving methods, improved spatial skills etc.

To Socialize Video games is often seen as an antisocial activity which might be true for computer-based single-player games but hardly for others, many players play just for the social aspect of sharing the experience with friends and family. Just look at the ever so popular event LAN-fests, where players gather somewhere to play games together or against each other. Video games generally tend to become more and more focused on socializing with consoles like Nintendo Wii and X-box Kinect becoming popular entertainment for the whole family.

The Dynamic Solitary Experience Just like in real life players do not always want to socialize but rather to be alone. Games like these gives players an opportunity to interact with something while still being in total control and can start or stop playing whenever they want.

The Bragging Rights Ever since the first arcade machines players have fought to have their name on the high-score-list for everyone to see. The respect a good score, a high ranking or a victory brings is what many players play for, it can give them a feeling of self-satisfaction to know that they do something well, maybe even better then most people.

The Emotional Experience The fact that games can give a such a big experience, with a lot of personal involvement, gives them a ability to evoke more emotions then most media. This is something that should be taken advantage off. People often use entertainment for some form of emotional payoff, in a video games this can be satisfaction of completing a challenge, adrenaline rush in an action filled game or frustration of complicated task.

To Explore Level-based games use the players desire to explore new levels and environments as a motivating force to keep them playing. The feeling of curiosity of coming to a new level and the want to just look around and explore the environment and its characters can sometimes be a bigger motivation to keep playing then actually completing a task. Some games take advantage of this and gives the player an experience to explore worlds on their own terms.

To Fantasize Many players look to video games to escape from their ordinary everyday life to somewhere exotic, glamourous or dangerous. The difference from other forms of escapism, such as novels and movies, games actually lets the player become a part of the other world instead of just following the life of its characters. A game can give the player a chance to experience great parts of history or visit exotic countries or even other planets. Games can also provide a safe environment where the player can behave in a way that socially unacceptable, like robbing banks or outrunning the police.

To Interact The thing that separates gaming from any other form of entertainment is that games provide an interactive environment. The human being have a need for interaction that we can not get from a movie, a book or a piece of music. A video game however can allow the player to control the environment to be involved in what happens and that is something they can not get elsewhere.

2.2 What Makes a Game Fun?

The main goal of a game is to provide entertainment, in other words; it should be fun. Nicole Lazzaro, president and founder of XEO Design, says that their research have shown that there are four keys for fun in gaming [4]. The first kind they call Hard fun, and it is the experience of being challenged in some way, in a complicated puzzle or a strong enemy, and the satisfaction of completing it. It is important that the challenge is not too easy but allows the player to get frustrated and annoyed with the game to create the satisfying feeling of joy and pride when victorious.

The second type of fun is what Lazzaro calls Easy fun [4], which is when a game lets the player explore the game on their own. The player's curiosity will make them try new actions and see what happens if they, for example, drive a racetrack backwards or removes the ladder from an occupied pool. The feedback from their actions will give the player a feeling of wonder and surprise and maybe even curiosity to try new actions. The third type of fun, Serious fun, is all about the reward of the game. How you feel before playing the game, during and after. A player can for example sooth their anger by playing an violent game, or maybe loose weight in a fun way with games such as Dance Dance Revolution. Games with the focus on teaching such as a flight simulator where the player learns the controls of an airplane is also part of the Serious fun category and gives the player a feeling of excitement and relaxation.

Lazzaro's forth and last kind of fun is People fun [4]. People fun is the social part of the game, the part that brings people together and creates an social bond. This can be achieved by letting players interact and share with each other, if the player can for example give a powerpack to their ally in need it gives the player a feeling of generosity and the ally a feeling of gratitude. The game allows the players to create a sense of human kindness and give them a feeling of pride when seeing a friend succeed or a feeling of pleasure at an enemies' misfortune. Lazzaro's research also showed that the best-selling games have at least three out of the four keys to fun and the most successful games provide three out of four within twenty minutes of gameplay.

3 Addictive Elements

Great game design does not stop at attracting players to the game but to keep them playing it. The game must have a way of bringing the player back to the game time after time, and to do this game designers cunningly incorporate addictive elements into their game.

3.1 The Game

According to CRC Health Group there are five main "hooks" that can be used to make a game addictive [5].

High-Score The high-score is one of the oldest and most recognizable way to keep a player playing. The high-score gives the player a rating in a way that motivates him to try harder, be better.

Beating the Game Almost every game gives the player a desire to beat the game, to finish the story or to win. The game often gives the player an incentive every now and then to enhance their desire by for example providing a new level. It is the desire for completion that keeps the player playing until they have fulfilled it.

Roleplaying Roleplaying games gives the player an emotional connection with his character as they embark on an adventure together which gives the player a need to continue playing

Discovery Discovery is another hook of roleplaying games where a lot of time is spent exploring different environments. The thrill of exploration creates the need and curiosity to see and experience everything the game have to offer, which makes the game hard to give up.

Relationships The main reason to keep playing online games is the social part of the game and the relationships to other players. Games brings together players with a common interests and gaming often become a way to spend time with friends and meet new people.

3.2 The Interaction

It is not just the gameplay that is addictive but also how it is played, so when you design a game it is important to think about the interaction. Most popular games have a small learning curve, in other words they are easy to learn to play [6]. Players do not want to learn how to play before they can play it, instead most games teach the player how to play as they go along. Many of the most addictive games are the ones that are mentally effortless like Tetris which, after learning the basics, you can basically play without thinking. Another aspect of a game that can be addictive is that it is not time consuming, by being quick to start up and let the player start playing, games can lure the player in for a "quick game", this is often used in, so called, casual games like Bejeweled.

4 Evaluation

Using the aspects and parts of game design and addictive elements that Rouse [3], Lazzaro [4], CRC Health group [5] and Kekeljvic [6] has given we look closer at four successful games. By looking at the game-design and gameplay the games will be evaluated to see if there is any relation between the game's game-design and its success and how well the different strategies work.

4.1 Angry Birds

Angry Birds is a casual puzzle game originally intended for mobile platforms. The game is based on a story of green pigs stealing the birds eggs, which of course makes the birds very angry and bark out on a mission retrieve their eggs. In each level the player are to shoot birds with a slingshot and try to kill all the green pigs and at the same time collect points by destroying as much as possible. The different types of birds have different qualities, one can speed through wooden boxes another can drop eggs etc. To continue to the next level the player have to kill all pigs in the current level.

Angry Birds and its two sequels have together been downloaded over 500 million times cross-platform [7]. But Angry Birds is no longer just a game but an franchise with games, dolls, boardgames, toys and clothes and there is even rumors about a full length movie. So what makes it so successful? The game is incredible easy to start playing and you are thrown into the story as soon as you start the application. The game becomes more and more advanced as you pass levels and even though it does not have any real kind of high-score it motivates the player with using stars. Each level has three stars, if you kill all the pigs in a level you can move on to the next but you have to kill the pigs and reach a certain amount of points to get all three stars. This makes creates a need not only to pass each level but pass each level with three stars. The later levels is also so complicated that it often creates a competition among friends and family of who have completed the most levels and who has all the stars. If that were not enough the game designer has added sound to the levels so that if the player fail to kill all the pigs they mockingly snorts and laughs at him, this to even further irritate the player into keep playing.

4.2 FarmVille

FarmVille is a casual social network game that for example can be played on Facebook. In FarmVille the player owns a simulated farm and can grow and harvest crops and raise farm animals. There is no clear goal in the game other then creating your very own dream farm by collecting farmcoins by completing tasks or selling crops. The farmcoins can be used to by seeds, animals, decorations etc. to grow your farm. But FarmVille is also a social game where the player can invite their friends to be their neighbors in the game, they can also give each other useful items in form of gifts and join together to complete tasks.

FarmVille was launched in 2009 and quickly became the game to play on Facebook and players spent massive amount of time tending their farms. Players play FarmVille because it lets them both fantasize and socialize but it also allows them to brag about their accomplishments to not only the other players but to all their friends on the social community. The game clearly have the key of Easy fun [4] with the work on the farm but also the key of people fun where the player gets the emotional experience of helping each other and playing together with friends. The combination of the joyous easy-going gameplay and the social collaboration between friends is what made FarmVille both enjoying and addictive.

4.3 StarCraft

StarCraft is an science fiction strategy game played in realtime and is one of the best-selling computer games of all time. The original game was released in 1998 and a long awaited sequel, StarCraft II, was released twelve years later. StarCraft takes place in a distant part of the Milky Way galaxy in the 26th century, and lets the player choose to play as one of the three races called Terrans, Zerg and Protoss, each with their own skills and abilities. The game is mainly a multiplayer game but it is also possible to play against a built-in virtual opponent. A game is played on a map where each player gets a simple base to start on, the purpose is then to strategically destroying the other players's armies while building up your own using the shattered resources of the map.

The reason StarCraft has become so popular is its well created gameplay, compelling story and the massive challenge it offer its players. The game has a high status among gamers which creates a desire to master it and receive the bragging rights it brings. StarCraft does not offer a high-score list but motivates the player by publishing rating lists, much like the ones you might find in tennis or golf. Despite StarCraft's hard learning curve, complexity and time-consuming gameplay it still appeals to the hard-core players, that strive on the challenge and adrenaline rush of the exciting, action filled game.

4.4 Tetris

One of the most popular games of all time is Tetris, which have been around for over 25 years and is still being played. Tetris is a simple game where blocks fall down from the top of the screen, while it falls the player can move and rotate it to fit into the other block at the bottom of the screen. Tetris is a great example of a casual game with its small learning curve, the mental effortlessness of its interaction and the fact that you can not complete it. Tetris offers a high-score to motivate the player to become better and also compare himself to other players. In a games Tetris is also designed to constantly remind the player of his earlier mistakes which creates a frustrating need to undo them which keep the player playing.

4.5 Summary

The Eight Wants of the Player As seen in Table 1 below, Angry Bird fills four of Rouse's eight wants [3]; The Challenge (to complete the levels and receive three stars), The Bragging rights (compare your points and progress to others), The Dynamic solitary experience (the player is in complete control when deciding when and where to play) and the want To Explore (finding out what happens in the next level?). Farmville fill five wants; To Socialize (the social aspect of the game such as inviting friends, sharing crop and working together in challenges), The Dynamic Solitary Experience (the player decides when to play and what to do), The Bragging Rights (the player can show his farm and his stats on Facebook), The Emotional Experience (Sharing things with friends and helping others) and To Interact (to interact with the farm and its animals but also with the other players in the game.).

Being the complex game it is StarCraft fills six of the eight wants; The Challenge (to learn how to play, to master it and to play games against other players), To Socialize (play with or against friends and meet new people), the Bragging Rights (the game's high status means skills are high value), The Emotional Experience (the game's stressful action filled gameplay), To Explore (to find new tactics and new resource to win and build armies) and to Interact (to interact with the environment to build armies and win battles). Tetris is filling three of Rouse's eight wants; The Challenge (the puzzling gameplay), the Dynamic Solitary Experience (The player has the control of when and where to play) and the Bragging Rights (the respect of being the best).

As conclusion Table 1 shows that all four games fulfills the players need for bragging rights, which seem to be an important reason to play. The bragging rights can be your name on the highscore list, completing a game or winning against other players and can in some way be found in most games. In Angry Birds it is completing a level with three stars, in Farmville it is showing off your farm to your friends, for example on Facebook, in Tetris it is the highscore and in StarCraft it is the win against other players and the ranking among players. The table below also show us that all four games fulfill several of the eight reasons to play which means that the game attracts players with different reasons to play and therefore attracts more players which can be a reason to their success.

The four keys to fun As shown in Table 2 Angry BIrds has at least two of the four keys to fun [4]; Hard fun (the challenging puzzle-levels) and Easy fun (the storybook chapters between levels and the option to use the birds qualities in new ways). Farmville also have two of the four keys to fun; Easy fun (the easy relaxed gameplay) and People fun (the social part of the game, collaborating with friends etc.). StarCraft have three of the four keys; Hard fun (the complicated game play and challenging battles), Serious fun (the serious of playing against great players and competing in tournaments with real cash prices) and People fun (playing against other players and being part of the StarCraft community). Tetris only has one of Lazzaro's four keys to fun namely Hard fun(the puzzle of fitting the falling pieces).

Table 1. The eight needs of the player [3]

	Angry Birds	Farmville	StarCraft	Tetris
The Challenge	X		X	X
To Socialize		X	X	
The Dynamic Solitary Experience	X	X		X
The Bragging Rights	X	X	X	X
The Emotional Experience		X	X	
To Explore	X		X	
To Fantasize				
To Interact		X	X	

As a conclusion it can be said that Lazzaro's four keys to fun [4] might not be the most important part of a successful game since Tetris only fulfill one out of the four keys but has still become one of the most successfull games and has managed maintain its popularity for over 25 years. Neither Angry Birds or Farmville fulfills the recommended three out of the four keys, only StarCraft with its complex gameplay fulfills three out of the four keys. Since all four games has had major success, Lazzardo's four keys might not be the most important key to making a successful game.

Table 2. The four keys of fun [4]

	Angry Birds	Farmville	StarCraft	Tetris
Hard Fun	X		X	X
Easy Fun	X	X		
Serious Fun			X	
People Fun		X	X	

Addictive elements Table 3 shows that Angry Birds has three clear addictive elements [5]; High-score (both most points on level and three stars), Beating the game (completing the level and the entire game) and Roleplaying (the backstory and adventure of getting the birds eggs back). Farmville also has three addictive elements; Roleplaying (taking on the role of a farmer, taking care of a farm and its animals), Discovery (Finding new crops, animals and decorations for the farm) and Relationships (social aspect of the game, spending time together with other players by sharing crops and helping each other). StarCraft also has three out of the five addictive elements; High-Score (rating of players around the world), Discovery (finding new resources and coming up with new strategies) and Relationships (the social part of playing against other people and being

part of the StarCraft community). Tetris however only have two clear addictive elements; the High-score and the way it creates a need for the player to Beat The Game.

As a conclusion it can be said that all four games include at least two addictive elements. Angry Birds, Farmville and StarCraft all have three out of the four addictive elements which can indicate that including addictive elements in a game can be important for its success. Since addictive elements keep the players interested and more likely to continue playing and talk about it to their friends, which will increase the games popularity and therefore its success.

Table 3. Addictive elements [5]

	Angry Birds	Farmville	StarCraft	Tetris
High-score	X		X	X
Beating the game	X			X
Roleplaying	X	X		
Discovery		X	X	
Relationships		X	X	

5 Conclusion

The evaluation of the four games using the aspects of Rouse [3], Lazzaro [4], CRC Health group [5] and Kekeljvic [6] can be concluded as it takes a great deal of knowledge to create a successful game, knowledge not only of how to create a game but knowledge of how you should structure the design and interaction of the game. The evaluation showed that the Lazzaro's four keys to fun [4] was not that important to a game's success. Instead should a game-designer try to design a game to appeal to different types of players by fulfill several of Rouse's eight reasons to play [3]. This will give the game a bigger audience and a larger opportunity of success. A game should also include at least one type of addictive elements since that will keep the players interested in the game and help attract new players to the game which is vital for a successful game. The evaluation can be summarized into four questions to be used when creating a game:

Who will Play the Game? The first part of designing a game should be figuring out who you are designing for. Who is your target group? Is it Hard-core players or casual players? What other games do they enjoy? Casual players prefer a game that is easy to learn whereas more hard-core players are more appealed to a complex gameplay and exciting challenge then a small learning curve.

Why will They Play the Game? Think about the wants Rouse described [3]. Do the player want the challenge of your game, explore the world you have created or maybe the emotional experience the game provides? Also try to if you can incorporate the four keys of fun in your game [4].

Where will They Play the Game? Where and when the player plays the game has a big impact on how to design the interaction of the game. If your game played while waiting for the bus or during the TV commercials you should design the game so its quick to start and quick to finish or at least to save your progress. If your game is played at a home computer it is allowed to take longer to start and quit but has higher demands on the complexity and quality of the game.

Why will They Keep Playing? How to keep the game interesting is an important part of designing a game that is frequently overlooked. To keep the player playing the game is important to be successful and you should therefore think of how you can incorporate addictive elements such as high-scores or discovery.

So if you are planing to create the next world obsession, the new Angry Birds perhaps, think about the questions above and how you are going to make your game successful, or maybe ever better, do as Shigeru Miyamota suggested, create something you would like to play and hope the rest of the world does too.

References

1. Cohen, D.: Cathode-ray tube amusement device - the first electronic game. About.com
2. Bellis, M.: Computer and video game history. About.com
3. Rouse, R.: Game design: theory and practice. Jones and Bartlett Publishers, Sudbury, MA (2005)
4. Lazzaro, N.: What makes a game fun? Videointerview and transcript by Austin Allan (02 2010)
5. http:www.video-game addiction.org: What makes a video game addictive?
6. Kekeljevic, I.: How to make addictive game. http://apps-on-android.com
7. Nerney, C.: Angry birds downloads pass half billion mark

Interactive Algorithmic Composition Using Markov Chains and Pentatonic Scales

Fredrik Vestermark

Department of Computing Science
Umeå University, Sweden
dv07fvk@cs.umu.se

Abstract. A major pentatonic scale is a musical scale where the notes can be freely combined without creating unpleasant harmonies. A Markov chain is a statistical process that is sometimes used to generate music. This paper describes a potential way to interactively generate music for games in real-time by combining pentatonic scales and multiple Markov chains. The reader gets a short introduction to pentatonic scales and Markov chains as well as how they can be used in music generation.

1 Introduction

In many computer games that use music, different tracks are played depending on events. For example, the music may change when the player enters a different room or if time is running out. Some computer games use dynamic music where the compositions are merged together more or less seamlessly. For this purpose, our approach will be to use the so-called pentatonic scale.

A pentatonic scale is a type of musical scale that is based on five pitches per octave, and has been used in many music cultures [1]. One reason might be that most of these scales only consist of pitches that can be freely combined without creating unpleasant harmonies. Because of this, they are probably suitable for creating harmonies and making seamless transitions between different melodies using the same scale.

One way to algorithmically compose music is to use Markov chains [2]. The idea of summing up a musical style with a Markov chain dates at least back to the end of the 1940's, and in 1957 it was used to compose music. Each state in the chain may represent a pitch, duration, note, pause, chords or even phrases, and the transition weights represent the probability to play the corresponding state next. Sometimes these chains are created by analyzing existing pieces, which result in generated sequences that are similar to the ones in the other piece [3].

In this paper, we explore how Markov chains and pentatonic scales can be combined to interactively compose simple computer music. The intention is to present a basis for how music in games can evolve in real-time depending on the course of event in a game. We address questions such as: Is the pentatonic scale actually suitable? Can multiple chains play in parallel? How is it possible to create transitions between music created by two or more different chains?

Bensch, Drewes, Gulliksson, Mejtoft (Eds.): Umeå's 15th Student Conference in Computing Science USCCS 2012, pp. 133–143, January 2012.

2 Background

Even though many papers have been written on the subject of algorithmic composition with Markov chains [3, 2, 4], to our knowledge only little scientific research has been done on using specifically pentatonic scales. However, pentatonic scales are well known and the combination has been used in practise by Chua [5], Cabrera et. al. [6] among others. This section provides a brief introduction to pentatonic scales and Markov Chains.

2.1 Pentatonic Scales

Pentatonic scales consist of five notes per octave. In the most common ones, it is possible to combine any note with any other without creating unpleasant harmonies. Notes that can be combined pleasantly are considered *consonant*. The opposite to consonance is *dissonance* [7]. Henceforth, all scales are assumed to be major C scales.

The heptatonic scale consists of the seven notes *C, D, E, F, G, A, B*, while the pentatonic uses the subset *C, D, E, G, A* [7]. We will use *C'* to denote the note one octave higher than *C*. Furthermore, the five notes in the pentatonic scale create a C6add9 chord [7], so playing a melody with them is equivalent to playing a melody with the notes of the C6add9 chord.

2.2 Markov Chains

A Markov chain represents a set of states, and for each of those states the probability for transitioning to other states [4]. Probabilities are values in the interval $[0, 1]$. A graphical representation is shown in Fig. 1.

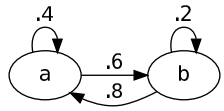


Fig. 1: A simple example of a Markov chain.

Markov Chains are often represented as transition matrices, which contain transition probabilities [4]. The transition matrix in Table 1 represents the same Markov chain as Fig. 1. A slightly more advanced transition table is shown in Table 2, where the probabilities are generated by a script that takes a sequence of pitches as input.

Short Term Memory. A Markov chain has no memory except for knowing its current state. It is however possible to simulate a short-term memory by creating permutations of the states [3]. The number of remembered previously

Table 1: A simple example of a transition matrix.

source	destination	
	a	**b**
a	.4	.6
b	.8	.2

Table 2: A transition matrix where each state represents the pitch. It is approximately derived from the first three lines of the verse in the song *Don't Worry, Be Happy* by Bobby McFerrin.

source	**destination**					
	C	**D**	**E**	**G**	**A**	**C'**
C	.25	.375	.375	0	0	0
D	.333	.167	.5	0	0	0
E	.428	.286	.143	.143	0	0
G	1	0	0	0	0	0
A	.167	.167	.167	.167	.166	.166
C'	.167	.167	.167	.167	.166	.166

visited states, including the current one, is called the *order* [4]. A second-order Markov chain is visualized in Fig. 2, where each node represents the previous state and the current one. By definition, it is only possible to have a transition from a state s_{ij} to s_{jk}.

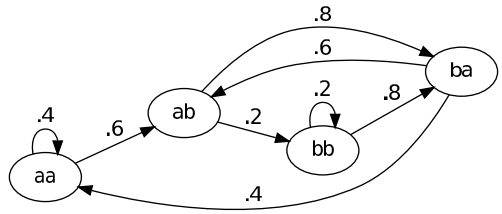


Fig. 2: An example of a second-order Markov chain, where each node represents the previous state and the current one.

A slightly different approach mentioned by Mc Alpine et. al. [2] is to use a multidimensional matrix which contains several different chains. Which chain to use is determined from the history of the few last visited states rather than only the last one, depending on the order.

3 Pentatonic Markov Chains

We define a Pentatonic Markov Chain (PMC) as a Markov chain that represents transitions between pitches in a pentatonic scale. It may contain pitches from multiple octaves, but for readability of this paper focuses on a single one. First order PMCs are used extensively in this paper, but the mentioned techniques can be applied to higher order PMCs with no or minor changes.

3.1 Generating Chains

To our knowledge, Markov Chains can be created manually, randomly, by generating them from existing pieces, or a combination of these techniques.

To generate a PMC from a given piece, first create a table in which the number of transitions from one pitch to an other are entered. Possibly include the transition from the last pitch to the first one. See Table 3a for an example.

Lastly, normalize the rows in the matrix by dividing each element in a row with the sum of the entire row. The resulting transition table is shown in Table 3b.

Table 3: Creating of a transition table for a merry tune with the pitch sequence *C, D, E, G, A, G, A, G, C', A, G, E, A, C', G, E.*

(a) Transition counts.

from	**to**					
	C	**D**	**E**	**G**	**A**	**C'**
C	0	1	0	0	0	0
D	0	0	1	0	0	0
E	1	0	0	1	1	0
G	0	0	2	0	2	1
A	0	0	0	3	0	1
C'	0	0	0	1	1	0

(b) Transition table created by normalizing the rows in Table 3a.

source	**destination**					
	C	**D**	**E**	**G**	**A**	**C'**
C	0	1	0	0	0	0
D	0	0	1	0	0	0
E	.333	0	0	.333	0.334	0
G	0	0	.4	0	.4	.2
A	0	0	0	.75	0	.25
C'	0	0	0	.5	.5	0

However, not every piece uses every pitch. When generating a PMC from such a piece, some rows will turn out to be all zeros after the first step, causing division by zero in the second step. The simplest solution is probably to give zero rows equally distributed transition probabilities. An example transition table using this solution is shown in Table 4. Other considerable solutions are purposed in Sect. 4.2 under *Generating Chains*.

Example Chains. The previously shown PMCs in Table 3b and 4 are used at later stage in this paper. For some examples, the more dissimilar PMCs in

Table 4: Transition table for a sorrowful tune. The transition table was generated from the sequence *A, A, E, A, A, E, C, C, E, C, E, A, C', C, A, E*. Since *D* and *G* are not present in the given sequence, their probabilities have been evenly distributed.

source	**destination**					
	C	**D**	**E**	**G**	**A**	**C'**
C	.25	0	.5	0	.25	0
D	.167	.167	.167	.167	.166	.166
E	.4	0	0	0	.6	0
G	.167	.167	.167	.167	.166	.166
A	0	0	.5	0	.333	.167
C'	1	0	0	0	0	0

Table 5a and 5b are used. These are deterministic, since every probability is 0 or 1.

Table 5: Transition tables for repeatedly playing the scale in increasing and decreasing order respectively.

(a) Increasing order.

source	**destination**					
	C	**D**	**E**	**G**	**A**	**C'**
C	0	1	0	0	0	0
D	0	0	1	0	0	0
E	0	0	0	1	0	0
G	0	0	0	0	1	0
A	0	0	0	0	0	1
C'	1	0	0	0	0	0

(b) Decreasing order.

source	**destination**					
	C	**D**	**E**	**G**	**A**	**C'**
C	0	0	0	0	0	1
D	1	0	0	0	0	0
E	0	1	0	0	0	0
G	0	0	1	0	0	0
A	0	0	0	1	0	0
C'	0	0	0	0	1	0

Multiple Octaves. When creating a PMC from an existing piece, there are a few ways to handle pitches in the piece that are not within the range of the intended PMC. Chua suggests the following solutions to a very similar task [5]:

- Discard invalid pitches.
- Transpose invalid pitches to the considered octave.
- Use a wider range of pitches.

These solutions are also applicable when generating a PMC.

3.2 Multiple Chains in Parallel

Due to the benefits with pentatonic scales described in Sect. 2.1, it is possible to add new PMCs without producing disharmonies. However, it might be hard for the listener to distinguish between chains, which is shown in Fig. 3. This is especially true when the PMCs have the same kind of tone. Figure 4 shows that assigning different octaves to the chains makes it easier to distinguish between the chains. The examples uses two copies of the merry PMC in Table 3b.

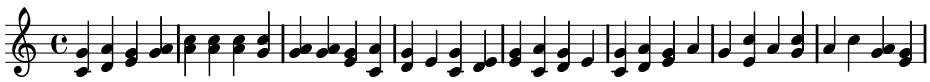


Fig. 3: A piece generated by two identical PMCs.

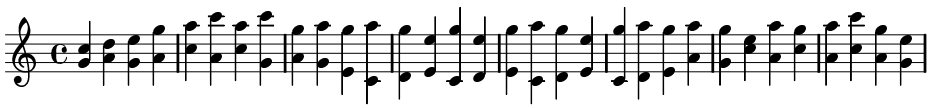


Fig. 4: The same piece as in Fig. 3, with a unique octave for each chain.

Since the previously produced notes have the same duration, the generated music will most probably sound monotonous even though the pitch combinations and sequences might be appealing in a different context. Cabrera shows that PMCs can be assigned individual durations [6], which reduces monotonicity. Figure 5 shows an example of using three copies of the PMC in Table 3b, and assigning different durations and octaves to each of them.

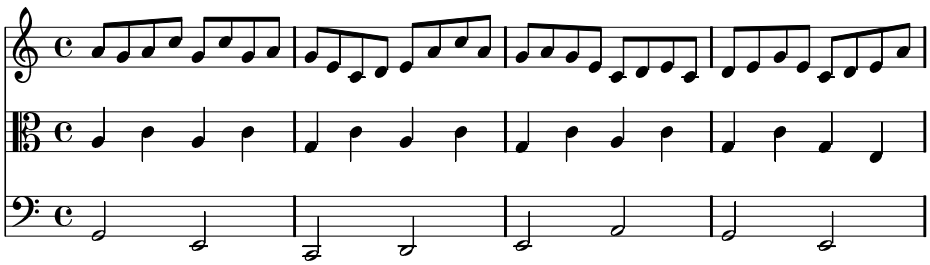


Fig. 5: A piece generated by three identical PMCs, assigned different durations and octaves.

3.3 Transitions Between Chains

Two techniques on how to make transitions between chains are described. They both require that the chains are of the same size.

Hot Swapping. The simplest transition technique is probably to replace one chain with another in a single operation. This gives a sharp transition, especially if the PMCs have very different probabilities. An example is shown in Fig. 6.

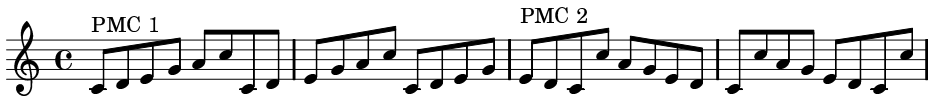


Fig. 6: Hot swapping from the PMC in Table 5a to the one in Table 5b after half the time.

Interpolation. To create a smooth transition from one transition table to an other, it is possible to use interpolation (Roads [8], p. 874).

Let

$$A_{n,n} = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{bmatrix} \text{ and } B_{n,n} = \begin{bmatrix} b_{1,1} & \cdots & b_{1,n} \\ \vdots & \ddots & \vdots \\ b_{n,1} & \cdots & b_{n,n} \end{bmatrix}$$

be two transition matrices where each row sums to 1. If each element in A is multiplied by $(1 - i)$ and each element in B is multiplied by i, where $i \in [0, 1]$, then each row in the newly created $C = (1 - i)A + i(B)$ must also sum up to 1. Therefore, C is also a valid transition matrix. An interpolation is done by increasing or decreasing i over time.

The two runs in Fig. 7 and 8 can be considered as interpolated versions of Fig. 6, since they show interpolations from the PMC in Table 5a to the one in Table 5b. Fig. 7 was selected as a random sample, and both ends have a similar pattern to what the non-interpolated chains would generate. However, the generation shown in Fig. 8 have less probable ends, and was selected as an example of a less successful result.

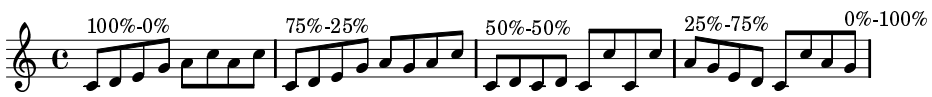


Fig. 7: Result of an interpolation between two chains. The ratio between the chains is constantly changing, shown above the score at a few moments.

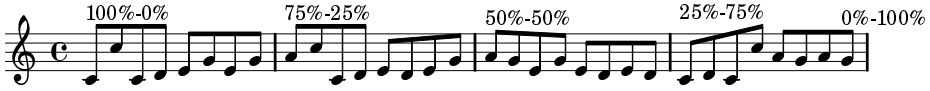


Fig. 8: Less successful result of an interpolation between two chains. The ratio between the chains is constantly changing, shown above the score at a few moments.

Figure 9 shows the result of an interpolation from the sorrowful PMC in Table 4 to the merry PMC in Table 3b. The PMC has been assigned three different durations and octaves. These PMCs differ less than the ones in Table 5, which makes it harder to decide which transitions originated from which chain.

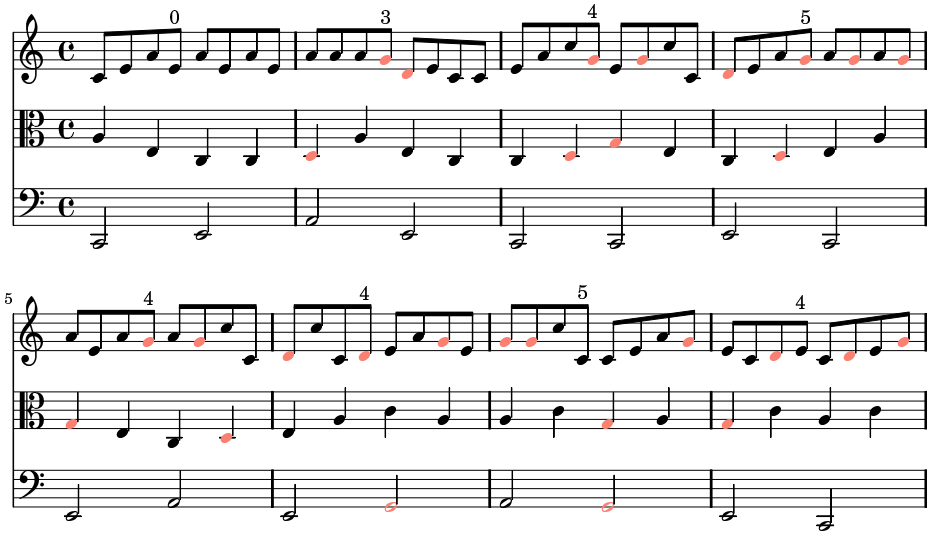


Fig. 9: An example result of an interpolation from the PMC in Table 4 to the one in Table 3b. The light red (or gray in black and white printouts) notes would not have been selected by solely the first PMC (D and G), and the number of red notes within each bar is shown above the bars.

The result might not be as clear as expected. However, the PMC in Table 3b would never generate any combination of D's or G's in a row, since these probabilities are 0. Similarly, the PMC in Table 4 would never make a transition to a D or G without playing one of them in the previous step. Because of this, subsequent red notes must be a combined result of the PMCs and the approach used to remove zero rows.

3.4 Suggested Use of PMCs in Games

Even though no game was implemented, a few suggestions on how PMCs could be used for generating music in games are described below.

Assigning PMCs to Objects. PMCs can be to objects, and by interpolating the PMCs depending on the distance between the objects and the player. The generated music would then correspond to the players surrounding. The interpolation could as well be based on *functions* of the distances, which enables different intenseness of the PMCs.

Assigning PCMs to Properties. An other approach worth considering is to assign PMCs to properties. Some possible properties would be the remaining time, life, or even the brightness or quantity of a certain color on the screen.

4 Discussion and Conclusions

4.1 Pentatonic Scale

During the work, a further seven PMCs were created and combined in different ways. The reason why the pentatonic scale was chosen in the first place was its lack of dissonant notes, and as long as multiple PMCs were used and assigned different octaves and duration, we found the result pleasing. Despite that, due to the limitations of the pentatonic scale, it was not always possible to express the kind of music we wanted as a PMC.

For comparison a few similar tests were done for the heptatonic scale. Under the same circumstances, and as long as transitions to the 'new' dissonant notes were used sparingly, we considered these results as pleasing as well. However, if two chains were assigned the same duration and dissonant notes were generated, our experience was that it sounded notably more unpleasant than if they had different duration. It is important to emphasize that only a few tests were performed with this scale, and the results may not should be considered as facts. Additionally, Berg [7] writes that "Except for one note, every note of the major scale sounds reasonably consonant over tonic harmony, because these pitches can be explained as chord tones or extensions". The particular note is the fourth in a major heptatonic scale, which in the major C scale is *F*.

Conclusion. The heptatonic scale can by definition express the same things as the pentatonic, and more. Instead of limiting a system to the pentatonic scale, one should consider using the heptatonic scale. However, it is favorable to use transitions to dissonant notes sparingly if possible, especially transitions to the fourth one in the scale.

4.2 Using Markov Chains

To be able to show that Markov chains can be suitable for creating interactive game music as described in this paper, a real use case must be implemented and evaluated by players. However, none of the generated results have indicated that soundly created Markov chains (with no zero rows and only few transitions to dissonant tones) is a bad idea. PMCs does however not have a long-term memory, which makes it hard to reuse larger patterns – something that is common in classical music. As mentioned in Sect. 5, it might be possible to create longer repeating patterns by introducing a deterministic PMC.

Generating Chains. As indicated in Sect. 3.1 and in Sect. 3.3 under *Interpolation* about Fig. 9, the solution to evenly distribute the transition probabilities in zero rows is an easy, but turned out not to be a suitable solution.

To create meaningful transition probabilities for zero-rows, the average probabilities could be used instead. These probabilities can be computed by summing the columns into a vector, and then normalizing it.

If a particular behavior is wanted, it might however make most sense to edit the transition values by hand.

Interpolation of Markov Chains. The tests performed on interpolation between Markov chains turned out well, and the music generally made a smooth transition.

Since the theoretical size of the transition table increases quadratically with the pitch range and exponentially with the order, these properties have a sharp limit where real-time interpolation is no longer possible. Exactly where this limit goes depends on the implementation.

Existing use of interpolated Markov chains for interactive algorithmic composition was not found until the very end of the project. The existence of prior use and that it is mentioned in at least one book for computer music may be seen as yet an indicator that interpolation is an applicable approach for interactive game music.

5 Future Work

There are several interesting topics for future work:

- Heptatonic scale. This paper has focused on the major pentatonic scale. Eventually it turned out that also the larger heptatonic major scale may be suitable. More research has to be done before drawing any strong conclusions.
- Rhythm. The only 'rhythm' in the generated music was the one introduced by assigning PMCs different duration. McAlpine states that "Markov processes are exceptionally suited for rhythm selection [2], so introducing Markov chains for rhythm could improve the possible musical difference.
- Deterministic Markov chain. Markov chains sometimes generate improbable music. What is the result if the Markov chains always select the most probable transition? Can this generate larger repeating patterns?

References

1. Iusfin, A.: Some problems of analyzing anhemitonic structures. Studia Musicologica Academiae Scientiarum Hungaricae **13**(1) (1971) 123–136
2. McAlpine, K., Miranda, E., Hoggar, S.: Making music with algorithms: A case-study system. Computer Music Journal **23**(2) (Summer 1999) 19–30
3. Ames, C.: The markov process as a compositional model: A survey and tutorial. Leonardo **22**(2) (1989) 175–187
4. Aschauer, D.: Algorithmic composition. Master's thesis, Vienna University of Technology (2008)
5. Chua, Y.S.: Composition based on pentatonic scales: A computer-aided approach. Computer **24**(7) (July 1991) 67–71
6. Cabrera, A.: Using python inside csound: An introduction to the python opcodes. Csound Journal (6) (1997)
7. Berg, S.: Alfred's Essentials of Jazz Theory: A Complete Self-Study Course for All Musicians. Alfred Music Publishing (2007)
8. Roads, C.: The computer music tutorial. MIT Press (1996)

Author Index