JOURNAL OF MANAGEMENT AND ENGINEERING INTEGRATION

Scope: The Journal of Management and Engineering Integration (JMEI) is a double-blind refereed journal dedicated to exploring the nexus of management and engineering issues of the day. JMEI publishes two issues per year, one in Summer and another in Winter. The Journal's scope is to provide a forum where engineering and management professionals can share and exchange their ideas for the collaboration and integration of Management and Engineering research and publications. The journal will aim on targeting publications and research that emphasizes the integrative nature of business, management, computers and engineering within a global context.

TABLE OF CONTENTS

The Journal of Management and Engineering Integration
Vol. 1, No. 2, December, 2008

A Fast Solution to the Multi-Level Production Smoothing Problem

Mesut Yavuz
Shenandoah University
myavuz@su.edu

Abstract

This paper is concerned with a production smoothing problem arising in just-in-time manufacturing systems. Two widely studied goals in production smoothing address the consumption of parts and workloads of machines. This paper contributes to the literature by presenting a new optimization model that addresses both goals at four production levels. The paper shows how to adapt three different heuristics from the existing literature and proposes a new approach to obtain approximate solutions to this new problem. Performances of the methods are compared through an extensive computational study, and the proposed approach is found effective in solving the problem within practical computation times.

1. Introduction

In his seminal book on Toyota Production System (TPS), Monden [1] names production smoothing as the cornerstone of TPS. Triggered by Monden's book, production smoothing has been studied by numerous researchers who have developed alternative mathematical models and analytical methods to be used in a wide array or real-life scenarios (see surveys by Kubiak [2] and Yavuz and Akcali [3]).

The *Production Smoothing Problem* (PSP) focuses on reducing the variability inherent in a mixed-product just-in-time (JIT) manufacturing system. Two alternative approaches for the PSP are widely studied in the literature: namely, the *single-level* and the *multi-level* PSP. In the single-level PSP, one aims to minimize the variation of the production rates of the end-products. In the multi-level PSP, on the other hand, the aim is to reduce the variability in both production rates of the end-products and consumption rates of the sub-assemblies that are produced at the preceding stages of the manufacturing system. Note that, in both versions of the PSP, one is concerned with scheduling the final level of operations only, whereas the sub-levels are assumed to adjust themselves to the schedule of the final level via the pull system used on the shop floor. It is proven in the literature that the single-level PSP can be solved in polynomial time whereas the multi-level PSP is NP-Hard [2].

The production smoothing problem is typically modeled as a sequencing problem where the sequence length is equal to the total demand. Miltenburg [4] formulates the single-level PSP as an integer quadratic optimization model with an objective function of minimizing the total squared deviation from ideal production amounts, which are depicted in Figure 1. In the figure, the ideal cumulative production amount of product i is represented with a straight line, based on the assumption that demand (d_i) is uniformly distributed throughout the planning horizon (T). Actual cumulative production amount, on the other hand, increases only at the end of the stages devoted to that product.

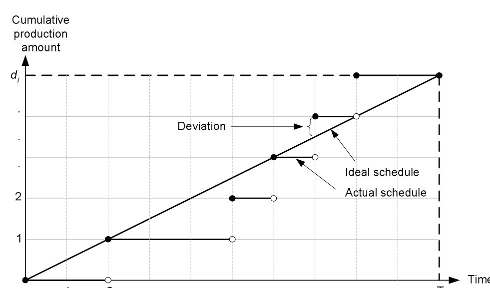


Figure 1. Measuring deviation

Early solution methods for Miltenburg's model include computationally expensive exact methods [4] [5] and constructive heuristics [4] [6] [7] [8] [9]. The most efficient exact solution of the problem is due to Kubiak and Sethi [10] [11], who show that the problem is polynomially solvable via a reduction to an assignment problem. This reduction will be discussed in detail later in the paper.

The multi-level PSP is formulated by Monden [1] by focusing at the second level only and measuring total squared deviation in sub-assembly consumption rates over the sub-assemblies and stages. Miltenburg and Sinnamon [12] generalize Monden's formulation of the problem by considering the deviation between the ideal and actual schedules at four levels. Their objective function incorporates deviations measured at all the four levels, with respect to their weights. Solution methods developed for the multi-level PSP include Kubiak *et al.*'s dynamic programming procedure [13], and several heuristic solution procedures by different authors [1] [12] [14] [15] [16] [17] [18] [19] [20] [21] [22]. Two widely recognized heuristics are Miltenburg and Sinnamon's [12] constructive (namely, one- and two-stage) heuristics. Also, Inman and Bulfin's [15] earliest due date (EDD) approach is an important contribution. These three methods will be further discussed later in the paper.

Smoothing the usage of the sub-assemblies and the loading of resources are recognized as two important goals in JIT manufacturing. *Usage goal* concentrates on the production rates of end-products, as well as the consumption rates of subassemblies that go into the end-products. The usage goal is generally measured as a function of the deviations of *actual* (cumulative) production / consumption amounts from pre-specified *ideal* (cumulative) production / consumption amounts as discussed above. *Loading goal*, on the other hand, focuses on processing requirements and captures the deviation of actual workload levels on the production resources from the ideal workload levels. Note that the concepts of ideal and actual schedules as well as measuring the deviation are common for both goals. Models considering the loading goal in the single-level PSP use a weighted objective function of both the usage and loading goals [5] [23].

Miltenburg *et al.* [5] develop a dynamic programming (DP) procedure to solve the problem with the weighed objective function. Korkmazel and Meral [23] distinguish between the processing time requirements on different stations (on the assembly line of the final level), and penalize the difference between the actual production time that is spent for a product in a workstation and the ideal production time that should have been spent for that product in that workstation. With this formulation, the problem can be reduced to an assignment problem and solved efficiently, as will be further exploited in this paper.

A few papers in the literature address both the usage and loading goals in a multi-level production smoothing setting [19] [24] [25]. However, in the majority of these papers, the loading goal has a single-level structure, i.e., considers smoothing the workload of production resources at the final stage only. An exception is due to Aigbedo and Monden [19], who consider the sub-assembly level as well as the final assembly level, and formulate the loading goal to capture the variation in the station loads over those two levels. They propose a solution method that requires the user to prioritize usage and loading goals for the two levels, and finds a sequence that satisfies all four goals simultaneously, breaking ties according to the priorities provided by the user.

In this paper, we first build a new optimization model for the multi-level PSP with both the usage and loading goals measured at four levels in Section 2. In Section 3, we adapt some of the existing solution methods to our problem. In Section 4, we re-formulate the problem and introduce weights associated with each product. The weights capture both usage and loading variability incurred at both the final and sub levels. Our re-formulation has a single-level structure, and can be solved in polynomial time, which is also discussed in Section 4. In Section 5, we present an extensive computational study. We conclude the paper in Section 6 with discussing the results obtained in the computational study and point out possible future research directions.

2. A mathematical model

Consider a manufacturing system consisting of four levels. Smaller numbers represent downstream levels, i.e., level 1 is the final assembly level. Level l = 1,..,4 produces n_l different parts/products, e.g., n_1 is the number of end-products. Each unit of end-product i requires $b_{i,j,l}$ units of part j from level l (i = 1,..,n_1, j = 1,..,n_l, l = 2, 3, 4). Also, we define $b_{i,i,1}$ = 1 (i = 1,..,n_1) and $b_{i,j,1}$ = 0 (i = 1,..,n_1, j = 1,..,n_1, $i \neq j$). The number of critical production resources, i.e., machines, at each level is denoted by m_l, l = 1,..,4, and the processing time of part j of level l on machine h of that level is given with $p_{h,j,l}$ for each h = 1,..,m_l, j = 1,..,n_l, l = 1,..,4.

The demands for the end-products are given with $d_{i,1}$, i = 1,..,n_1, and they determine the demands for the parts at the sub-levels with $d_{j,l} = \sum_{i=1}^{n_1} d_{i,1} b_{i,j,l}$, j = 1,..,n_l, l = 2, 3, 4. Total demand at level l is denoted by $D_l = \sum_{j=1}^{n_l} d_{j,l}, l = 1,..,4.$ D_1 defines the sequence length and is used to calculate the demand rate of each part/product: $r_{j,l} = d_{j,l} / D_1$. The problem is to find a production sequence for the end-products, that is, a sequence of D_1 units long. Let s_k be the product scheduled in the kth stage of the sequence, k = 1,..,D_1. We denote the cumulative actual production/consumption of part/product j in the first k stages of the sequence with $X_{j,k,l}$ such that $X_{j,0,l}$ = 0, j = 1,..,n_l, l = 1,..,4, and $X_{j,k,l} - X_{j,k-1,l} = b_{s_k,j,l}, j = 1,..,n_j, k = 1,..,D_1, l = 1,..,4.$

The ideal production/consumption amount is obtained by $kr_{j,l}$, k = 1,..,D_1, j = 1,..,n_l, l = 1,..,4. Consequently, the total squared deviation measured for part j of level l is

$$U_{j,l} = \sum_{k=1}^{D_1} \left(X_{j,k,l} - kr_{j,l} \right)^2 . \tag{1}$$

Production of part j at level l requires $d_{j,l}p_{h,j,l}$ units of machine h's time. Therefore, ideal workload on machine h of level l is $\bar{p}_{h,l} = \sum_{j=1}^{n_l} d_{j,l} p_{h,j,l} / D_1$ units per stage. Actual workload is a function of the production sequence and is calculated with $Y_{h,k,l} = \sum_{j=1}^{n_l} X_{j,k,l} p_{h,j,l}$. As a result, the total squared deviation measured for machine h of level l is

$$L_{h,l} = \sum_{k=1}^{D_1} \left(Y_{h,k,l} - k\bar{p}_{h,l} \right)^2 . \tag{2}$$

Putting all together, we formulate the following model for the problem.

Minimize

$$Z = \sum_{l=1}^{4} \left[\sum_{j=1}^{n_l} w_{j,l}^U U_{j,l} + \sum_{h=1}^{m_l} w_{h,l}^L L_{h,l} \right] \tag{3}$$

Subject to:

$$\sum_{i=1}^{n_1} X_{i,k,1} = k, \; k = 1,..,D_1. \tag{4}$$

The objective function (3) is a weighted sum of the usage and loading goals, where the weights ($w_{j,l}^U$ and $w_{h,l}^L$ for the usage and loading goals, respectively) are part- and machine-specific at each level. The weights should be determined by the practitioner to reflect the relative importance of products, parts and manufacturing resources used. Constraint (4), along with the definition of the decision variable $X_{i,k,1}$, assures that exactly one end-product is scheduled at each stage.

3. Adaptation of existing solution methods

In a special case of the problem, we set all weights but $w_{j,2}^U$, j = 1,..,n_2 to zero, thereby obtaining the multi-level PSP with the usage goal only, which is an NP-Hard problem. Being its general case, so is our problem. Thus, computationally efficient heuristic methods are desirable to solve the problem. In this section, we adapt three well-known production smoothing heuristics.

3.1. One-stage heuristic

The one-stage heuristic is a greedy constructive heuristic we adapt from [12]. It starts with an empty sequence and proceeds by scheduling one product at each stage. At the first stage, all n_1 products are evaluated as candidates. The product with minimum contribution to the objective function is selected and scheduled at the first stage. If the demand for this product is equal to the total quantity scheduled so far, the product is eliminated from further consideration. This cycle is performed until all the products are scheduled to the D_1 stages, that is, a complete schedule is obtained. At any stage, ties can be broken arbitrarily.

3.2. Two-stage heuristic

The two-stage heuristic is also a greedy constructive heuristic we adapt from [12], and it is similar to the one-stage heuristic. That is, it starts with an empty sequence and proceeds by scheduling one product at each stage. The difference however is in the depth of calculations. More specifically, the two-stage heuristic considers the next two stages and selects the product which minimizes the total deviation measured in these two stages. (The selection of the product at the inner (second) stage is made using the one-stage heuristic.) Again, ties can be broken arbitrarily.

3.3. EDD

The EDD method is based on calculating a due-date for each copy of each product and sequencing the products in the non-decreasing order of their due-dates. In other words, it is a constructive heuristic that evaluates stages one-at-a-time and schedules the product with the earliest due date. The due date for the tth copy of product i is calculated with $\delta_{i,t} = (t - 0.5)D_1/d_{i,1}$, for all $t = 1,..,d_{i,1}$ and $i = 1,..,n_1$. With this heuristic, ties can occur frequently and they can be broken arbitrarily.

4. Re-formulation of the problem

Heuristic solution methods can simply be divided into two categories based on whether they approximate the optimal solution of the problem or the problem itself. The ones in the first category aim at approximating the exact optimal solution of the problem. Among such methods are simple constructive heuristics and more sophisticated local and meta-heuristic search procedures. The one- and two-stage heuristics discussed above fall into this category. However, in some cases, it is more beneficial to modify the problem formulation to obtain a polynomially solvable problem. Then, the exact optimal solution of the new problem is obtained efficiently, which can also be a near-optimal solution to the original problem. The EDD heuristic discussed in the previous section belongs to this class. In this section, we first re-formulate our problem and then show how it can be solved in polynomial time.

Steiner and Yeomans [26] address the multi-level PSP. In their work, the sub-levels are assumed to be consisting of dedicated resources to feed the end-products. More specifically, the sub-assemblies and components that go into an end-product are produced for that end-product only. In this case, the variability in the consumption rate of a component depends only on the variability of the production rate of the end-product it goes into. This significantly simplifies the problem by reducing it to a single-level PSP. We observe a similar approach in solving the single-level PSP with the loading goal. Miltenburg *et al.* [5] define a processing time for each end-product, measure workload variability for each end-product separately, and obtain the total workload variability by adding all together. Korkmazel and Meral [23] define multiple stations on the final assembly line such that each product may have a different processing time in each station. They measure the workload variability using the actual processing time devoted to a product in a station and the ideal thereof. Then, they obtain total workload variability by summing over the stations and end-products. A close observation reveals that a common approach in these three papers is the decomposition of total consumption/workload into pieces of consumption/workload created by each end-product. This simplifying approach in formulation makes the problem solvable in polynomial time.

Kubiak and Sethi [10] [11] note that, when the objective function is formulated in the form $\sum_{k=1}^{D_1}\sum_{i=1}^{n_1} F_i(\cdot)$ where $F_i(\cdot)$ is a unimodal convex function that has a minimum of $F_i(0) = 0$, for each individual unit (i.e., copy) of an end-product, it is possible to define an ideal position in the sequence. It is also possible to define a cost function that increases if a particular copy of an item deviates from its ideal position. Kubiak and Sethi's cost definition leads to a reformulation of Miltenburg's model as an assignment problem with D_1 elements, and, hence is solvable in $O(D_1^3)$ time. In order to reduce our objective function to the solvable structure, we first split it into four as follows.

$$Z_1 = \sum_{i=1}^{n_1} w_{i,1}^U U_{i,1} \quad (5)$$

$$Z_2 = \sum_{l=2}^{4} \sum_{j=1}^{n_l} w_{j,l}^U U_{j,l} \quad (6)$$

$$Z_3 = \sum_{h=1}^{m_1} w_{h,1}^L L_{h,1} \quad (7)$$

$$Z_4 = \sum_{l=2}^{4} \sum_{h=1}^{m_l} w_{h,l}^L L_{h,l} \quad (8)$$

Here, Z_1 is the production rate variation component, Z_2 the consumption rates at the sub-levels, Z_3 workload variation at the final-level and Z_4 workload variation at the sub-levels. Note that, Z_1 is the only one that has the desired $F_i(\cdot)$ structure. All four components of the objective function are re-combined as follows.

$$\text{Minimize} \quad Z' = \sum_{i=1}^{n_1} W_i U_{i,1} \quad (9)$$

Where

$$W_i = w_{i,1}^U + \sum_{l=2}^{4}\sum_{j=1}^{n_l} w_{j,l}^U \left| b_{i,j,l} - r_{j,l} \right| + \sum_{h=1}^{m_1} w_{h,1}^L \left| p_{h,j,1} - \bar{p}_{h,1} \right| + \sum_{l=2}^{4}\sum_{h=1}^{m_l} w_{h,l}^L \left| \left(\sum_{j=1}^{n_l} b_{i,j,l} p_{h,j,l} \right) - \bar{p}_{h,l} \right| \quad (10)$$

Obviously, $U_{i,1}$ has the desired $F_i(\cdot)$ structure and multiplying it with a positive constant (W_i) does not change it. Therefore, the new objective function can be optimized in polynomial time. The weight W_i is calculated in such a way that takes all related variability measures into account. $w_{i,1}^U$ is the weight associated with the production rate variation of product i, and it is preserved in W_i. For all the parts at the sub-levels, the absolute difference between $b_{i,j,l}$ and $r_{j,l}$ is the deviation of product i's consumption of part j from the average consumption of that part. For example, if all products require exactly one unit of a part at the second level, then the variability of the consumption of this part will be zero, i.e., the sequence of the products will not make a difference for the consumption of this part. Also consider a part at the second level that is required by the end-products by one, two or three units, with two being the average. Then, sequencing an end product that requires exactly two units of this part does not make a difference, whereas the others make a (absolute) difference of exactly one unit. Similarly, we consider all workload requirements at all levels and incorporate them into the weight of each product. The larger a weight coefficient is, the more responsible that product is for the variability in the system.

In what follows we (following Kubiak and Sethi's [10] [11] footsteps) show how the re-formulated problem is transformed into an assignment problem. First, we define an ideal position for each copy of each product and a cost function which increases as a copy deviates from its ideal position. Let Z_t^{i*} be the ideal position of the tth copy of product i and $C_{t,k}^i$ be the cost of assigning the tth copy of product i to the kth stage of the sequence. Then, the following formulation is defined:

$$Z_t^{i*} = \left\lceil \frac{(2t-1)D_1}{2d_{i,1}} \right\rceil$$

$$C_{t,k}^i = \begin{cases} W_i \sum_{q=k}^{Z_t^{i*}-1} \Psi_{t,q}^i, & \text{if } k < Z_t^{i*} \\ 0, & \text{if } k = Z_t^{i*} \\ W_i \sum_{q=Z_t^{i*}}^{k-1} \Psi_{t,q}^i, & \text{if } k > Z_t^{i*} \end{cases}$$

Where:

$$\Psi_{t,q}^i = \left| \left(t - q\frac{d_{i,1}}{D_1} \right)^2 - \left(t - 1 - q\frac{d_{i,1}}{D_1} \right)^2 \right|$$

Here, $\lceil x \rceil$ denotes the smallest integer that is greater than or equal to x and $|x|$ denotes the absolute value of x.

Let $S_{t,k}^i \in \{0,1\}$ be the decision variable denoting whether the tth copy of product i is assigned to the kth stage of the sequence. The assignment problem formulation of our problem is then given by:

$$\text{Minimize} \sum_{i=1}^{n_1} \sum_{t=1}^{d_{i,1}} \sum_{k=1}^{D_1} C_{t,k}^i S_{t,k}^i \quad (11)$$

Subject to:

$$\sum_{i=1}^{n_1} \sum_{t=1}^{d_{i,1}} C_{t,k}^i S_{t,k}^i = 1,\ k = 1,..,D_1. \quad (12)$$

$$\sum_{k=1}^{D_1} C_{t,k}^i S_{t,k}^i = 1,\ i = 1,..,n_1,\ t = 1,..,d_{i,1}. \quad (13)$$

Constraint set (12) assures that each copy of each product is assigned to exactly one position. Similarly, constraint set (13) assures that exactly one copy is assigned to each position. This assignment problem (AP) with $2D_1$ nodes can be solved in $O(D_1^3)$ time using the well-known Hungarian method or any other optimization algorithm for the assignment problem [27] [28].

5. Computational study

5.1. Experiment plan

In this study, we pseudo-randomly create 1000 test instances. Each instance is defined with n_l, m_l, $d_{i,1}$, $b_{i,j,l}$, $p_{h,j,l}$, $w^U_{j,l}$ and $w^L_{h,l}$ for each l = 1,..,4, i =1,..,n_1, j = 1,..,n_l, and h = 1,..,m_l. As explained below, all data are created using the discrete uniform distribution with different intervals for each parameter.

The number of parts/products at each level are determined according to the discrete uniform distribution such that $n_l \in \{3,..,5l+2\}$ and $m_l \in \{1,2\}$.

The demand for each end product is chosen between six and 25 ($d_{i,1} \in \{6,..,25\}$). Part requirements of the end products are created using $b_{i,j,l} \in \{0,..,2^l-1\}$.

Processing time requirements on each critical manufacturing source is created between six and 15 time units ($p_{h,j,l} \in \{6,..,15\}$, for all l = 1,..,4, j = 1,..,n_l, and h = 1,..,m_l).

Weights $w^U_{j,l}$ and $w^L_{h,l}$ are the only continuous parameters. Initially, we create them according to the uniform distribution such that $w^U_{j,l} \in [0,1]$ and $w^L_{h,l} \in [0,1]$. Then we analyze them in four groups and normalize the weights in each group. More specifically,

$$\sum_{i=1}^{n_1} w^U_{i,1} = \sum_{l=2}^{4}\sum_{j=1}^{n_l} w^U_{j,l} = \sum_{h=1}^{m_1} w^L_{h,1} = \sum_{l=2}^{4}\sum_{h=1}^{m_l} w^L_{h,l} = 0.25 \cdot$$

Our instance generation method is very flexible, i.e., it allows having a very diverse set of instances in order to represent a rich variety of scenarios that may be faced in practice. We support this approach by creating a large number (1000) of instances to obtain reliable results in lieu of real data.

5.2. Methods

We implement four heuristic methods on each instance. These methods are the three heuristics adopted from the existing literature in Section 3 and the assignment problem transformation proposed in Section 4. Since the problem is a computationally very difficult problem, we are not obtaining the exact optimal solutions of the test instances in this study. Instead, we compare the performance of the four alternative methods with each other.

All the algorithms are coded in C# language, which is included in the Microsoft Visual Studio.NET 2003 package. The experiments are performed on a personal desktop computer with a P4 3.4GHz CPU and 2GB memory. For the solution of the emerging assignment problems, a commercially available optimization package - CPLEX 9.1 is used.

5.3. Results

The computation times for all four methods are found negligible, i.e., under a minute, in our study. Thus, we do not draw conclusions about solution times here. For the solution quality, since we do not know the exact optimal solutions of the instances, we first find the best result and then report the performance of each method relative to the best of the four. Table 1 presents a summary of our results on all 1000 instances. The methods are in the rows and the performance measures are in the columns. The first column reports the number of times a method found the best solution of the four methods. The second and the third columns report the average and maximum ratios of the objective function values obtained by a method to the best (lowest) objective function value.

Table 1. Relative performances of the heuristics

Method	*Number* of hits	*Ratio* Avg.	Max.	*Time (min)* Avg.	Max.
1SH	37	2.05	7.42	0.01	0.05
2SH	425	1.26	3.96	0.11	0.94
EDD	144	1.16	2.38	0.01	0.04
AP	510	1.08	2.31	0.07	0.26

6. Discussion

The largest number of hits is obtained by the AP method. Furthermore, the lowest average and maximum ratios to the best solution is achieved by the AP method. Consequently, we state that the AP transformation proposed in this paper is comparable to the existing heuristic approaches adapted to the problem.

Among the adapted heuristics, the one-stage heuristic is the worst with respect to all three performance measures. Considering that its more-sophisticated version, the two-stage heuristic, has a better solution quality as well as a negligible computational time requirement, we claim that 1SH is not comparable to the others and can be dropped.

Among the EDD and 2SH methods, we observe an interesting trade-off. That is, 2SH finds the best solution of all four methods in a larger number of instances, whereas EDD more consistently approximates the best solution.

It is also an interesting observation that the EDD approach, not considering the part consumption variations or workload variations, approximates the best solution of the four methods rather closely.

In this study, no attempt has been made to obtain the exact optimal solutions of the instances. Therefore, development of an exact solution method for the problem is an important future research direction on the problem. In the existing literature, dynamic programming has been proposed for simpler versions of the PSP. Therefore, adaptation of dynamic programming to the problem studied in this paper appears to be a straightforward and promising research direction.

Development of an exact solution procedure will also enable a more accurate performance evaluation of the heuristics. Moreover, the heuristic procedures can be improved by performing a local search on the solutions obtained.

In summary, this paper addresses an important production scheduling problem arising in just-in-time manufacturing systems and develops a general framework thereof. For the solution of the general problem, the paper presents how to adapt three known heuristics in the literature and also proposes a new heuristic approach. The new approach is based on re-formulating the problem and then solving the re-formulation as an assignment problem in polynomial time. Computational results show that the proposed approach is promising in solving the generalized problem effectively.

7. References

[1] Monden Y., 1983, Toyota Production System, Industrial Engineering and Management Press.

[2] Kubiak W., 1993, Minimizing Variation of Production Rates in Just-In-Time Systems: A Survey, European Journal of Operational Research, 66, 259-271.

[3] Yavuz M., Akcali E., 2007, Production Smoothing in Just-In-Time Manufacturing Systems: A Review of the Models and Solution Approaches, International Journal of Production Research, 45, 3579-3597.

[4] Miltenburg J., 1989, Level Schedules for Mixed-Model Assembly Lines in Just-In-Time Production Systems, Management Science, 35, 192-207.

[5] Miltenburg J., Steiner G., Yeomans S., 1990, A Dynamic Programming Algorithm for Scheduling Mixed-Model Just-In-Time Production Systems, Mathematical and Computer Modeling, 13, 57-66.

[6] Sumichrast R., Russell R., Taylor B., 1990, Evaluating Mixed-Model Assembly Line Sequencing Heuristics for Just-In-Time Production Systems, Journal of Operations Management, 9, 371-390.

[7] Inman R., Bulfin R., 1991, Sequencing JIT Mixed-Model Assembly Lines, Management Science, 37, 901-904.

[8] Ding F., Cheng L., 1993, A Simple Sequencing Algorithm for Mixed-Model Assembly Lines in Just-In-Time Production Systems, Operations Research Letters, 13, 27-36.

[9] Cheng L., Ding F., 1996, Modifying Mixed-Model Assembly Line Sequencing Methods to Consider Weighted Variations

for Just-In-Time Production Systems, IIE Transactions, 28, 919-927.
[10] Kubiak W., Sethi S., 1991, A Note on Level Schedules for Mixed-Model Assembly Lines in Just-In-Time Production Systems, Management Science, 37, 121-122.
[11] Kubiak W., Sethi S. P., 1994, Optimal Just-In-Time Schedules for Flexible Transfer Lines, The International Journal of Flexible Manufacturing Systems, 6, 137-154.
[12] Miltenburg J., Sinnamon G., 1989, Sequencing Mixed-Model Multi-Level Just-In-Time Production Systems, International Journal of Production Research, 27, 1487-1509.
[13] Kubiak W., Steiner G., Yeomans S., 1997, Optimal Level Schedules for Mixed-Model, Multi-Level Just-In-Time Assembly Systems, Annals of Operations Research, 69, 241-259.
[14] Miltenburg J., Sinnamon G., 1992, Algorithms for Scheduling Multi-Level Just-In-Time Production Systems, IIE Transactions, 24, 121-130.
[15] Inman R., Bulfin R., 1992, Quick and Dirty Sequencing for Mixed-Model Manufacturing Systems, International Journal of Production Research, 30, 2011-2018.
[16] Cakir A., Inman R., 1993, Modified Goal Chasing for Products with Non-Zero/One Bills of Material, International Journal of Production Research, 31, 107-115.
[17] Bautista J., Companys R., Corominas A., 1996, Heuristics and Exact Algorithms for Solving the Monden Problem, European Journal of Operational Research, 88, 101-113.
[18] Duplaga E., Hahn C., Hur D., 1996, Mixed-Model Assembly Line Sequencing at Hyundai Motor Company, Production and Inventory Management Journal, 37, 20-25.
[19] Aigbedo H., Monden Y., 1997, A Parametric Procedure for Multicriterion Sequence Scheduling for Just-In-Time Mixed-Model Assembly Lines, International Journal of Production Research, 35, 2543-2564.
[20] Duplaga E., Bragg D., 1998, Mixed-Model Assembly Line Sequencing Heuristics for Smoothing Component Parts Usage: A Comparative Analysis, International Journal of Production Research, 36, 2209-2224.
[21] Zeramdini W., Aigbedo H., Monden Y., 2000, Bicriteria Sequencing for Just-In-Time Mixed-Model Assembly Lines, International Journal of Production Research, 38, 3451-3470.
[22] Ding F., Zhu J., Sun H., 2006, Comparing Two Weighted Approaches for Sequencing Mixed-Model Assembly Lines with Multiple Objectives, International Journal of Production Economics, 102, 108-131.
[23] Korkmazel T., Meral S., 2001, Bicriteria Sequencing Methods for the Mixed-Model Assembly Line in Just-In-Time Production Systems, European Journal of Operational Research, 131, 188-207.
[24] Miltenburg J., Goldstein T., 1991, Developing Production Schedules Which Balance Part Usage and Smooth Production Loads for Just-In-Time Production Systems, Naval Research Logistics, 38, 893-910.
[25] Tamura T., Long H., Ohno K., 1999, A Sequencing Problem to Level Part Usage Rates and Work Loads for a Mixed-Model Assembly Line with a Bypass Subline, International Journal of Production Economics, 60-61, 557-564.
[26] Steiner G., Yeomans S., 1996, Optimal Level Schedules in Mixed-Model, Multi-Level JIT Assembly Systems with Pegging, European Journal of Operational Research, 95, 38-52.
[27] Kuhn H. W., 1955, The Hungarian Method for the Assignment Problem, Naval Research Logistics Quarterly, 2, 83-97.
[28] Dell'Amico M., Toth P., 2000, Algorithms and Codes for Dense Assignment Problems: The State of the Art, Discrete Applied Mathematics, 100, 17-48.

Investigation of Operator Behavior Using Haptic Controlled Backhoe Simulator

Benjamin Osafo-Yeboah, Ritson Delpish, and Xiaochun Jiang
North Carolina A & T State University
xjiang@ncat.edu

Abstract

Research in haptic feedback and control has received a lot of attention in recent years because haptic provides force feedback that allows users to have the sensation of 'feel.' This study is an attempt to investigate the actions and behaviors of human operators as they manipulate a backhoe simulator using the Phantom haptic device. Sequential data of operator behavior were collected in this exploratory study. The most salient finding of this research is that there was interdependence among certain operator behavior events as defined in the study.

1. Introduction

Fluid power systems play a very important role in today's economy. Since the 1940's, fluid power has been used primarily in various machinery providing power in the form of hydraulics and/or pneumatics for industries such as manufacturing, transportation, agriculture and aerospace. Among them, a backhoe, powered by hydraulics, is a piece of excavating equipment with a digging bucket on the end of an articulated arm (Figure 1). Like many other earthmoving equipment, operating a backhoe is not easy. One common problem is that operators need to solve the inverse kinematic relationships between lever displacement and bucket trajectory [1]. It usually takes quite a long time before an operator can feel comfortable operating the machine and solve the inverse kinematic relationships subconsciously. It certainly is not easy for a novice operator to have a "feel" for the non-intuitive level motions since the only feedback available to the operator is the observed bucket speed, the engine's response to a load, and/or pressure waves propagated back to the user's hand. Consequently, earthmoving tasks are often performed by professionals no matter how small the task may be. To overcome this costly and inconvenient solution, haptic interface is being considered to replace the traditional direct-acting valve control levers. This haptic interface will free the operators from solving the inverse kinematic relationship and therefore, can shorten the training time for a novice operator [1-2].

Figure 1. A John Deer backhoe in operation

Haptics, when used in operating fluid power systems such as a backhoe, can be referred to the "feel", or the tactile sensations of the lever, joystick, wheel, and other controls, and the force felt when operating the machine [3]. The idea of introducing haptic interface to earthmoving equipment is to provide the operators a more complex "feel" that not only includes the input, but also the feedback of the equipment to the input [1]. Such a haptic feedback can be very crucial in helping novice operators, who lack experience to do a better job. The value of such a feedback can be explained by relating to our experience in gardening using a shovel. Furthermore, the intuitiveness of the interface can reduce operator mental workload and stress level, leading to improved performance. In addition, haptic feedback can help earthmoving equipment operators to avoid damaging utility lines during excavation, by providing a force feedback that alerts the operator to the presence of unusual obstacles, and, therefore, lead to a safer use.

The ease of use is very important when designing the new generation earthmoving equipment interfaces. There has been a shortage in qualified operators for operating earthmoving equipment. With the large number of baby boomers retiring from the industry, this problem becomes even more serious. One of the reasons is that the training process of the operator is quite long due to the non-intuitive interface. Therefore, introducing the haptic interface to the earthmoving equipment may provide an answer to some of the questions facing the industry.

Another concern is this new E-generation is far more accustomed to computer games than operating machines. It will be much easier for them to use this new haptic interface that resembles interfaces they may have encountered from their gaming experience.

This concept of using haptic interface in a backhoe has been tested at the Georgia Institute of Technology where a haptically operated backhoe simulation tool was developed [1-2]. The haptic input device is PHANToM 1.0, originally designed by Massie and Salisbury in 1994 and subsequently commercialized by SensAble Technologies [4-5]. The Phantom is a device consisting of a rod attached to a robotic arm that is capable of providing force feedback. This device is capable of six degree of freedom position sensing and three degree of freedom force.

The significance of this conceptual design of the backhoe haptic interface is that it can be easily expanded to other fluid power systems such as excavators.

Since traditional earthmoving equipment like many other machines mainly rely on visual (and some auditory) feedback, operators are often used to those two modalities. In this new design, the third modality, haptic feedback is also included with the expectation that this extra modality will improve operator performance.

Two members of the NSF-funded Engineering Research Center for Compact, Efficient and Effective Fluid Power, North Carolina A & T State University and Georgia Institute of Technology are currently collaborating on exploring this new generation interface for earthmoving equipment.

Among the research issues, one critical issue in this research is that the impact of the haptic interface on operator behavior is not clear. Since operators now need to coordinate his or her vision, hearing, and tactile feedback, it is necessary to investigate how operator behavior might be affected in this new system.

Operating earthmoving equipment is a complex task. Like many other complex human machine interaction tasks, it usually can be defined as a set of discrete events such as actions, eye movement, and/or think aloud verbalizations. Simply analyzing each individual event independently is no longer sufficient. Instead, the relations and interdependence of these events need to be analyzed. For instance, the fact that action A is always followed by action B may reveal important information. In general, sequential data analysis can have important implications for human machine interface design [6]. One tool to analyze sequential data is the Pathfinder network scaling algorithm. Pathfinder is a structuring technique. It has been widely used to analyze sequential data in human machine systems [7-8].

The objective of this study is to study the operator behavior patterns using the haptic backhoe simulator.

2. Methodology

An exploratory study was conducted at North Carolina A & T State University to study the operator behavior using the haptic interface of the backhoe simulator. The following sections will describe the participants, stimulus materials, equipment, and procedure of the study.

2.1 Participants

Six participants (3 males and 3 females) between the ages 20 and 30 years old (mean age = 23.5 and standard deviation = 3.6), enrolled at North Carolina Agricultural and Technical State University (NCA&T) participated in the study. All of them were African Americans. The lack of diversity in the participant pool reflects the fact that NCA&T is a Historically Black University. All of the participants had some experience using a joystick and playing videogames. None of them had operated a backhoe or any other earthmoving equipment before the experiment.

2.2 Stimulus Material

The haptically controlled backhoe simulator developed by Georgia Institute of Technology [1-2] was used in this study. A screenshot of the simulator can be seen in Figure 2.

The participants were instructed to use Phantom Omni device to control the backhoe and dig the dirt from the trench and to drop the dirt into a designated area. The Phantom has 6 degrees of freedom. The participant can use it to dig the dirt from the trench as indicated by two white stripes and move the bucket towards the right of the backhoe, and dump the dirt to the designated area (Figures 2 and 3).

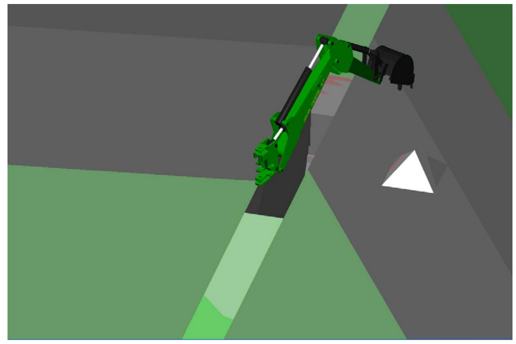


Figure 2. Backhoe simulator

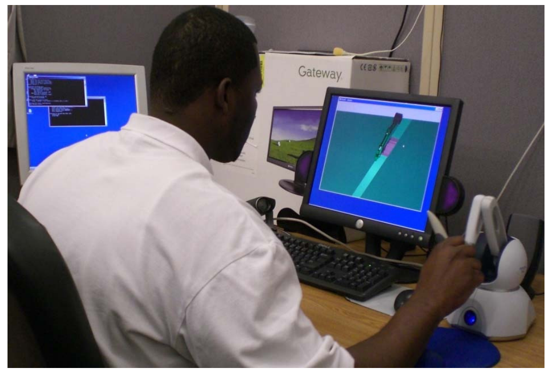


Figure 3. A participant is operating the backhoe using the Phantom

2.3 Equipment

Four Dell computers with Windows XP operating systems were used in this study. Phantom Omni was used to develop the haptic interface. This study was conducted in the usability lab (Noldus Usability Lab with two cameras and a microphone) at North Carolina A & T State University. The usability lab contains two rooms. On the experimental side, there were three computers connected using a local network. The first computer was used for Phantom Omni. The second is for graphical display, and the third one is for xPC-target simulation that is used to run the backhoe simulation. On the control side, there was one computer and a video monitor that displays video feed from the experimental room (Figure 4).

2.4 Procedure

Upon arrival at the usability lab, the participants were briefed the purpose of the study and asked to sign an informed consent form. A pre-test questionnaire was administered. Demographic information such as gender, age, and ethnicity was collected. Other information related to the experiment such as their experience in operating any earthmoving equipment, and playing videogames was also solicited. A short demo was then given and each participant was given five minutes to play with the simulator. Questions from the participants were answered before the experiment started.

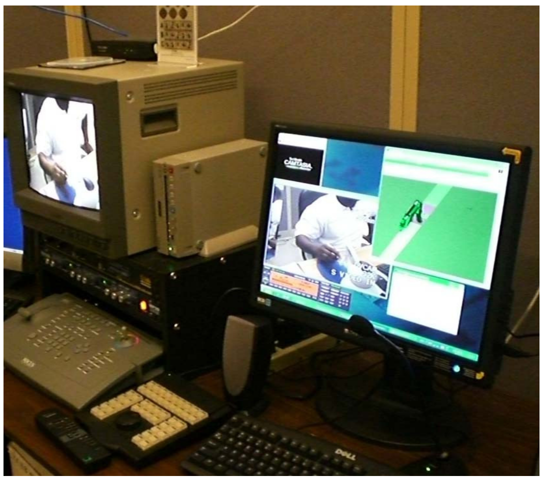


Figure 4. Control room in the usability lab

The participants were informed that the experiment would be taped for behavioral analysis. They were also encouraged to think aloud during the experiment. The task was to dig dirt from the trench and dump it to the designated area. Each participant was asked to perform the task twice.

Upon completion of the task, the participants were thanked and debriefed. A post-test questionnaire was administered afterwards. Each participant was asked about their experience using the haptic interface, their comfort level, and suggestions about the interface were solicited. The experiment lasted about 40 minutes.

3. Results and Analysis

All tapes containing experimental data were viewed and coded. The following events were defined in order to study the operator behavior: (1) Phantom Omni movement horizontal, (2) Phantom Omni movement vertical, (3) Phantom Omni movement turning, (4) eye fixating on monitor, (5) eye movement between monitor and Phantom Omni, (6) eye fixating on Phantom Omni, and (7) verbalization. Frequencies of each event were counted.

Pathfinder algorithm was used to analyze the sequential data in this study. Prior to use pathfinder, a transition matrix containing frequencies of transitions between the events was constructed. Conditional transition probability for each pair of events (A, B) was calculated as follows:

$$P = \frac{\#\ of\ \ event\ \ B\ \ follow\ \ event\ \ A}{\#\ of\ \ event\ \ A}$$

The transition matrix based on the coded data is shown in Table 1 where 1-7 represents event 1 through 7 as defined in the previous paragraph. For instance, the conditional transition probability for event 2 follows event 1 is 0.13.

Table 1. Transition matrix (in percentage)

Event	1	2	3	4	5	6	7
1	0	0	0	20	0	0	0
2	13	0	10	53	0	0	0
3	0	0	0	0	0	5	0
4	0	0	0	0	3	0	0
5	0	25	13	0	0	0	0
6	0	0	0	0	5	0	4
7	10	18	15	0	2	0	0

Using the pathfinder algorithm, the network can be derived as shown in Figure 5.

In constructing this pathfinder network, a cutoff probability of 0.02 was used. Any links that fail to satisfy this minimum strength cutoff value were omitted from the network.

Results clearly indicated that there was some interdependence among the seven events in the study. For instance, there was often a verbalization event following a horizontal Omni movement event (with a probability of 0.10) and there was usually a vertical Omni event following an Omni turning even (with a probability of 0.10). Those links clearly demonstrated that there are certain rules and knowledge in the network. Therefore, those events cannot be treated separately. There are certain behavior patterns as shown in the network.

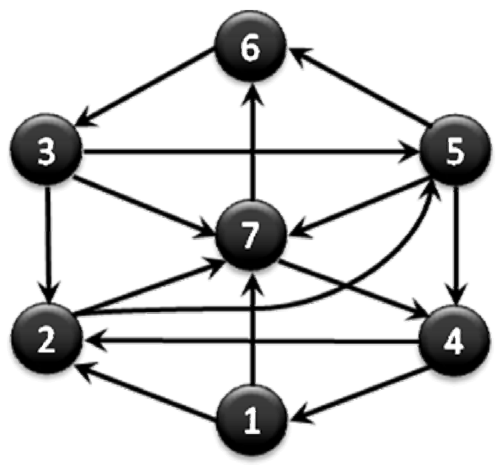


Figure 5. Pathfinder network

4. Discussion and Conclusion

Sequential data of operator behavior were collected from the usability lab in this exploratory study. The data was coded into seven discrete events. Conditional transition probabilities were calculated for all pairs of events and a transition matrix was constructed. This matrix was then submitted to Pathfinder. A network of linked events was generated using Pathfinder. Through this analysis, sequential data were reduced and complex sequential relationships were revealed.

The most salient finding of this research is that there was interdependence among certain operator behavior events as defined in the study.

The findings of this research, although preliminary, provide insights into how operators act when operating the haptic interface using the simulated backhoe. The existence of incorrect or awkward links can provide even more valuable information since they usually imply some difficulties. For instance, in this study, the link between Omni turning and Eye movement between the monitor and the Omni indicates operators struggling to coordinate between his or her hand and eye movements.

Even though this study is constrained with a small sample size, the methodology can be easily applied to a larger sample in a real environment in the future research. In addition, usability study on the simulation can be done to further understand operator behavior.

These findings will help haptic interface designers to improve the interface design.

5. Acknowledgement

This research is partially funded by the NSF Engineering Research Center for Compact, Efficient, and Effective Fluid Power. The views and conclusions contained in this paper are those of the authors and should not be interpreted as presenting the official policies or position, either expressed or implied, of the National Science Foundation.

The authors would also like to thank Professor Wayne Book from Georgia Institute of Technology and Dr. Matthew Kontz from Caterpillar Inc. for the backhoe simulator and their technical support.

6. References

[1] Frankel, J., (2004) "Development of a Haptic Backhoe Testbed," Master's. Thesis, Georgia Institute of Technology, Atlanta, GA, May 2004.

[2] Kontz, M., and Book, W., (2008) "Electronic control of pump pressure for a small haptic backhoe," International Journal of Fluid Power, 8(2), pp.5-16.

[3] Wickens, Lee, Liu, and Sallie Gordon-Becker (2004), Introduction to Human Factors Engineering, 2ND Edition, Prentice Hall.

[4] Massie, T.H, Salisbury, JK (1994) "The PHANToM Haptic Interface: A Device for Probing Virtual Objects" Proceedings of the ASME Winter Annual Meeting Chicago, Illinois.

[5] Cavusoglu, M. C, Feygin, D, Tendick F (2002) "A Critical Study of the Mechanical and Electrical Properties of the PHANToM Haptic Interface and Improvements for High-Performance Control." The MIT Press Journal, Volume 11, Number 16.

[6] Cooke, N., Neville, K, Rowe, A., (1996) "Procedure network representation of sequential data." Human computer Interaction, 11, 29-68.

[7] Goldsmith, T., Davenport, D., (1990) "Assessing structural similarity of graphs", In R. Schvaneveldt (Ed.), Pathfinder associate networks: Studies in knowledge organization (pp.75-87). Norwood: New Jersey: Ablex.

[8] Kellog, W., Breen, T. (1990) "Using pathfinder to evaluate user and system models" In R. Schvaneveldt (Ed.), Pathfinder associate networks: Studies in knowledge organization (pp.179-195). Norwood: New Jersey: Ablex.

The Journal of Management and Engineering Integration
Vol. 1, No. 2, December, 2008

Economics of Quality Improvement in Manufacturing Processes Based on Attributes

Gamal S. Weheba[*], Akale M. Tola[**], and Abu S. Masud[*]
[*]*Wichita State University*
[**]*Hawker Beechcraft Corporation*
gamal.weheba@wichita.edu

Abstract

Traditional models of the cost of quality combine prevention and appraisal expenses and consider their coupled effect on failure costs. This approach conceals the real economic value of continuous process improvement and undermines its scope and objectives. In this paper we represent a modified cost function based on the revised model developed by Weheba and Elshennawy [1]. The modified function can be used to estimate the total reactive costs at a given stable level of operation utilizing attribute data. A case study is presented with numerical evaluations to illustrate its performance.

1. Introduction

Beginning in the 1950's, quality-oriented staff departments were urged to evaluate quality costs to sell their activities to the company managers [2]. Early contributions of Feigenbaum [3] have resulted in what is known today as quality related costs. As pointed out by Campanella [4], quality costs represent the difference between the actual cost of a product or service and what the reduced cost would be if there were no possibility of substandard services, failure of products, or defects in their manufacture. More specifically, quality costs are the total costs incurred by (a) investing in the prevention of nonconformances to requirements, (b) appraising a product or service for conformance to requirements, and (c) failing to meet requirements. A number of conceptual models were proposed and used to depict the relationship between these cost elements [5]. However, there is a growing need for more realistic models that best represent current practice of successful organizations in improving quality and reducing the total cost. The following section represents a review of the published literature pertaining to quality cost models and recent developments made in this area.

2. Literature Review

Gryna [5, 6] represented two conceptual models for the cost of conformance. Each model shows three curves representing failure costs, operation (prevention and appraisal) costs, and the resultant total cost of quality. The first model depicted the prevailing views during much of the twentieth century. A major aspect of this model is the infinite cost required to attain perfection. The second model, however, indicated a more realistic view of the total cost of quality. In contrast to the older model, the total cost curve indicated that higher conformance costs less. However, the tendency to combine appraisal and prevention costs in both models has been questioned by Diallo et al, [7] and Fine [8, 9]. During a study of the collection and use of quality related costs, Plunkett and Dale [10] found wide differences between the models and real data. They concluded that the models are inaccurate and can be misleading. Ittner [11] examined the hypothesis that conformance expenditures must continue to be increased to achieve ongoing reductions in nonconformance costs. Based on a time series analysis of quality costs reported by 49 manufacturing units of

21 companies, he observed that nonconformance cost reductions could be achieved with little or no subsequent increase in conformance expenditures. He also pointed out that a micro-level examination of quality cost behavior could provide a better understanding of the underlying economics of quality improvements. Despite these limitations, a recent review of research on the cost of quality models [12] indicated that the method most commonly implemented is the classical prevention-appraisal-failure model.

In 2003, Weheba and Elshennawy [1] proposed a revised model for the cost of quality. The model incorporated two cost functions as shown in Figure 1. The first, termed the reactive cost of quality, estimates the cost of maintaining stable levels of operation at existing levels of conformance (L), while the second, termed the proactive cost, estimates the cost of achieving improved levels. The difference between these cost estimates represents incremental savings, which can be used in evaluating process improvement alternatives. However, they assumed that a single normal variable x can be used to characterize the process performance relative to the part design nominal m with tolerance requirements given as $m \pm \Delta$.

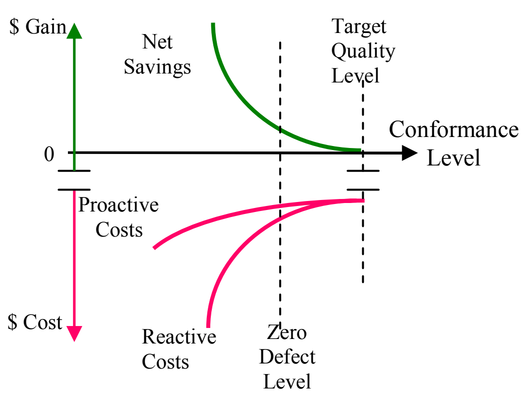


Figure 1. Cost of quality: Revised model [1]

In this paper, the reactive cost function is modified to model scenarios where production units are classified as either conforming or nonconforming to design requirements. Statistical monitoring of such characteristics requires the use of control charts for attributes. The following section represents the mathematical development of the modified model, whereas, Section 5 represents numerical results based on a case study.

3. Process model and cycle time

The process considered here is one of discrete manufacturing of one or more quality characteristics. Following the same modeling approach and notations in [1], it is assumed that a Shewhart-type control chart for attributes is being used to monitor performance over time. A production unit may be classified as nonconforming with respect to one or more quality characteristics. The charting scheme consists of drawing samples of n items, every h hours and plotting the calculated statistic against control limits placed k-standard deviation units around the centerline. The probability that a sample point falls outside these control limits when no change is made to the process is α, whereas the probability that a sample point falls within the control limits when a change has actually been made is β. As the process continues to be in control, formulated lots of N items are released to an inspection station for appraisal. Lot acceptance is based on inspecting a random sample of size n_S. We denote the probability of acceptance by Q, and that of rejection by P. Accepted lots are passed to further processing stages, while rejected lots are routed to screening and nonconforming units are replaced. Non-destructive testing is assumed in both sampling and screening activities. Delay costs are incurred when the lot completion time λ_t exceeds its processing time λ. Assuming that the time to inspect a unit is e and that to replace a nonconforming unit is λ_r, then the time required to inspect a sample of n_s units and replace nonconforming units is, $\lambda_S = n_S(e + \lambda_r p)$, where p is the average proportion of nonconforming units. Similarly, the time required to inspect the sample, screen the lot and replace nonconforming units is, $\lambda_{sc} = N(e + \lambda_r p)$. As such, the expected cycle time can be expressed as:

$$E(\lambda_t) = \lambda + (1-\alpha)[\lambda_s + P(\lambda_{sc} - \lambda_s)] + \alpha\,\lambda_{sc} \quad (1)$$

The expected delay that would be experienced in delivering N units to the customer is:

$$E(d) = (e + \lambda_r p)[N - Q(1-\alpha)(N - n_s)] \quad (2)$$

Hence, the mean squared delay can be estimated as:

$$\upsilon^2 = (e + \lambda_r p)^2[N^2 - Q(1-\alpha)(N^2 - n_s^2)] \quad (3)$$

Equations (1-3) above are comparable to those derived in [1] under the assumption that it takes longer to screen rejected lots than to investigate a false signal from the control chart.

4. Modified reactive cost function

In this section, we derive the modified function of the cost of maintaining a stable level of operation at a given average proportion of defective units. This includes the cost of statistical monitoring, product inspection cost as well as the loss due to the production of nonconforming units.

4.1. Statistical monitoring cost

This includes the cost of inspecting samples of n units of the product and that of investigating occasional false alarms. If B is the cost of inspecting a unit and W is the cost of investigating a chart signal, then the average monitoring cost per lot of size N is:

$$E(C_m) = NB\left\{(1-\alpha)\frac{n}{N} + \alpha(1 + \frac{W}{NB})\right\} \quad (4)$$

It is important to note here that this element represents the minimum unavoidable portion of the total reactive costs. In applications where the control charts are used as archives of the final inspection results, Equation (4) should be reduced to $\alpha(NB + W)$.

4.2. Product inspection cost

Included in this element are the costs incurred due to inspecting lots of the product after processing. Two components are considered; namely initial costs C_{i1} and downstream costs C_{i2}. The first represents initial cost of inspecting samples of n_s units plus the occasional cost of screening rejected lots. The assumption of constant lot size requires the replacement of any nonconforming unit found. Since these replacements have a proportion p nonconforming units, it will take on the average $1/(1\text{-}p)$ units to find a conforming one. Consequently, the costs involved are:

$$E(C_{i1}) = \frac{NB}{(1-p)}\{P + Q\frac{n_s}{N}\} \quad (5)$$

Downstream costs represent those of further processing nonconforming units in accepted lots. An estimate of such cost is based on the number of remaining units (N-n_s), and the probability of acceptance Q. These remainders contain p nonconforming units, the processing of which represent a loss of A' per unit. This loss is equivalent to the value added on subsequent processes until the nonconforming unit is detected. The replacement cost is $B/(1\text{-}p)$ as before. Hence, the downstream cost can be estimated as:

$$E(C_{i2}) = \frac{NB}{(1-p)}\{(1 - \frac{n_s}{N})Q\,p[\frac{A'}{B}(1-p) + 1]\} \quad (6)$$

Adding the two equations and rearranging, the expected inspection cost per lot is:

$$E(C_i) = \frac{NB}{(1-p)}\left\{P + Q\frac{n_s}{N} + Qp(1 - \frac{n_s}{N})[\frac{A'}{B}(1-p) + 1]\right\} \quad (7)$$

Equation (7) is in agreement with that given by Weheba and Elshennawy [1]. The function captures the relationship between the process average and the economics of sampling.

4.3. Cost of Deviation

This cost element includes the direct in-plant cost of producing nonconforming units and the consequent cost of delay. If A is the in-plant cost of disposing or reworking a nonconforming unit, the expected direct cost per lot can be expressed as:

$$E(C_{d1}) = A\,p\,N \tag{8}$$

On the other hand, the consequent delay cost is estimated using the quadratic-loss function proposed by Taguchi, et al., [13]. Since nonconforming units are replaced, the main effect of rejecting a lot or investigating a signal is to increase the lot completion time λ_t. This would result in internal losses due to overtime operation and increased handling and inventory costs. If a is the delay cost per lot and ρ the time slack defined as the difference between the lot delivery date D and its processing time λ, then the process delay can be represented as an S-type (smaller-the-better) characteristic with zero target value. Thus, the indirect cost of delay per lot is:

$$E(C_{d2}) = \frac{a}{\rho^2}\upsilon^2 \tag{9}$$

Where υ^2, is the mean squared deviation given by Equation (3). Adding Equations (8) and (9), the expected total cost of deviation is:

$$E(C_d) = A\,p\,N + \frac{a}{\rho^2}\upsilon^2 \tag{10}$$

The total reactive cost of quality per lot at any given level of performance can now be obtained by adding Equations (4), (7) and (10). The expected reactive cost of quality per unit time of operation at the stable level $\{L\}$ is:

$$E(RC)_L = \frac{NB}{\lambda}\left\{(1-\alpha)\frac{n}{N} + \alpha\left(1+\frac{W}{NB}\right) + \frac{1}{1-p}\left[1+Q(1-p)\left(p\frac{A'}{B}-1\right)\left(1-\frac{n_s}{N}\right)\right] + \frac{A}{B}p + \frac{a}{\rho^2}\frac{1}{NB}\upsilon^2\right\} \tag{11}$$

Equation (11) above can be used to estimate the total reactive cost at two levels of stable operation: one representing the current level $L = 0$, and the other representing an improved level of conformance $L = 1$. Any serious attempt to improve quality must result in a reduction in the total reactive costs. In other words, the difference between $E(RC)_0$ and $E(RC)_1$ should always reflect positive net savings. The following section provides an illustrative example from the industry.

5. Case study

To illustrate performance of the modified model, the following example is used. A company manufactures aluminum castings that are used on its assembly line. A number of quality characteristics are inspected, and parts are classified as conforming or nonconforming. Lots of N =150 units are produced and submitted for inspection. An attributes sampling plan is used with samples of size 20 units and acceptance number (c) of zero. Accepted lots are passed for further processing, while rejected lots are screened and nonconforming units are replaced. The cost of inspection per unit (B) is $1.00, while in-plant cost of rework (A) is $10.00 per unit. In addition, the cost of replacing a nonconforming casting after assembly (A') is $50. It takes on the average 0.2 hour to process a unit, 0.05 hours to inspect it, and 0.3 hour to replace a nonconforming unit after assembly. The company faces delay costs a = $2000, when lot completion time exceeds λ = 32 hours due to interruptions of assembly operations. This limits the time slack to 2 hours. Due to a high rejection rate, a process improvement team was formed to improve the process. The team decided to construct a proportion defective chart to evaluate the state of operation. Historical data from inspection results were used to calculate sample proportions. These were plotted against 3-sigma limits placed at 0.000 and 0.2347. The results indicated that the process was operating in a state of statistical control at an average proportion ($\overline{p}$) of 0.067. Based on these results, the team leader obtained the following estimates:

$$\alpha = 1 - \sum_{x=0}^{4} \frac{n_s!}{x!(n_s - x)!} \, \overline{p}^x (1-\overline{p})^{n_s - x}$$
$$= 1 - (0.99106) = 0.00894$$

$$Q = \frac{n_s!}{c!(n_s - c)!} \, \overline{p}^c (1-\overline{p})^{n_s - c}$$
$$= (1-\overline{p})^{20} = 0.2498$$

In addition, the cost of investigating a false signal from the *p*-chart was estimated at $500 per occurrence. Since the lot-screening time is much larger than that required to investigate a chart signal, Equations (2) and (3) resulted in the following estimates:

$$E(d) = (e + \lambda_r p)[N - Q(1-\alpha)(N - n_s)] = 8.26$$

$$\upsilon^2 = (e + \lambda_r p)^2 [N^2 - Q(1-\alpha)(N^2 - n_s^2)] = 83.68$$

Using Equation (11) the expected reactive cost of quality $E(RC)_0$ is estimated at $1406 per hour at the current process level. It is interesting to note here that as much as 98% of this total cost is attributable to the inability to meet delivery schedules. This attests to the ability of the modified model to include significant cost elements that were traditionally overlooked by the classical models. Typically, an application of the prevention-appraisal-failure model would have resulted in an estimation of the direct failure cost only. This could hinder the economic feasibility of any attempt to improve the process. In this application, it was clear that improvement efforts were worthy and the team was able to identify two quality characteristics as major contributors to part rejection. A root-cause analysis indicated that both characteristics correlated to the orientation of the parts during initial operations. Team members from the Engineering Department recommended that corrective measures be taken to repair the fixture to assure accurate orientation of parts before machining. It was anticipated that these changes could reduce the average proportion of nonconforming units to the 1% level. Consequently, a decision was made to inspect each part after processing to confirm the effectiveness of these corrective measures. Using Equations (2) and (3) the team leader estimated the average delay at the new level to be 2.37 hours with associated mean squared delay of 12.88 hours^2. Using Equation (11), this was translated into an estimated total reactive cost $E(RC)_1$ of $219 per hour of operation. The team reported its findings with an estimated 84% reduction of the total reactive cost. Nonetheless, it became obvious that the original acceptance sampling plan and the *p*-chart have to be revised upon achieving the new level of performance. One of the options considered is to terminate acceptance sampling activities and utilize a control chart for monitoring the time between occurrences of nonconforming units.

6. Summary and conclusions

This paper offers a modified version of the revised model for the cost of quality proposed by Weheba and Elshennawy [1]. Cost elements of the reactive function were modified to accommodate applications involving attributes. The resulting function portrays three cost elements and reflects the same logic advocated in [1] regarding the economics of process improvements. An example from the industry was added to demonstrate the ability of the modified model to capture hidden losses and hence justify improvement efforts. It is important to note here that applications of the modified model are subjected to the same limitations typically reported in analyzing this type of data using control charts for attributes. These charts tend to require samples of large sizes as the average proportion of defective units decreases. In scenarios involving high yield processes, the required sample sizes may not be feasible. In such scenarios, measurements on selected quality characteristics are more likely to be economically desirable. However, the proposed model can be used during the initial steps, when multiple quality characteristics are involved, to justify the need for improvement. It is likely that these initial steps, when successfully completed, lead to applications involving measurements on single quality characteristics where the revised model [1] can be utilized.

References

[1] Weheba, G., and Elshennawy, A., (2003). "A revised model for the cost of quality," International Journal for Quality and Reliability Management, Vol. 21, No. 3, P291-308
[2] Juran, J., (1951). Quality Control Handbook, 1st edition, McGraw-Hill, New York, NY
[3] Feigenbaum, A., (1991). Total Quality Control, 3rd ed., McGraw-Hill, New York, NY
[4] Campanella, J., and Corcoran, F.J., (1999). Principles of Quality Costs, 3rd ed. ASQ Quality Cost Committee, ASQ Quality Press, Milwaukee, Wisconsin
[5] Gryna F., (1988). Quality costs, in Juran, J.M., Gryna, F.M., Quality Control Handbook, 4th ed., McGraw-Hill, New York, NY, pp. 4.1-4.30
[6] Gryna, F.M., (2001). Quality Planning and Analysis from Product Development Through Use, 5th ed., McGraw-Hill, New York, NY, pp. 20-30
[7] Diallo, A., Khan, Z.U., and Vail, C.F., (1995). "Cost of quality in the new manufacturing environment," Management Accounting, Vol. 77, No. 2, pp. 20-25
[8] Fine, C. H., and Porteus, E. L., (1988). "Dynamic process improvement," Operations Research, Vol. 37, No. 4, pp. 580-591
[9] Fine, C.H., (1986). "Quality improvement and learning in productive systems," Management Science, Vol. 32, No. 10, pp. 1301-1315
[10] Plunkett, J.J., and Dale, B.G., (1988). "Quality costs: a critique of some economic cost of quality models," International Journal of Production Research, Vol. 26, No. 11, pp. 1713-1726
[11] Ittner, C.D., (1996). "Exploratory evidence on the behavior of quality costs," Operations Research, Vol. 44, No. 1, pp. 114-130
[12] Schiffauerova, A., and Thomson V., (2007). "A Review of Research on Cost of Quality Models and Best Practices," International Journal of Quality & Reliability Management, Vol. 23, No. 6, pp. 647-669
[13] Taguchi, G., E.A. Elsayed and T.H. Hsiang (1989), Quality Engineering in Production Systems, McGraw-Hill, New York, NY

The Journal of Management and Engineering Integration
Vol. 1, No. 2, December, 2008

Gender Role Portrayals in Television Commercials: An Inquiry Into Stereotyping in an Economy in Transition

Michael Stoica
Washburn University

Darryl W. Miller
University of Wisconsin – River Falls

Dan Ardelea
University Politehnica Bucharest

Abstract

This is a pilot study that examines gender role portrayals in Romanian television commercials and compares the outcomes to the findings in Sweden and Japan published previously. Results revealed both progressive and traditional gender role portrayals for both women and men reflecting a society and economy in a state of transition toward European Union membership. The study tested the masculinity index (Hofstede 1988) comparing gender role portrayals in the three countries. The results show a mixture of traditional and modern roles for both men and women.

1. Introduction

The portrayal of gender roles in advertising is an area of inquiry that has received a significant attention in the marketing literature over the past several decades. Many studies of gender roles in advertising have been conducted largely to determine whether the increased economic, political, and social power of women in many parts of the world, particularly in the industrial and post-indusrial nations, is reflected in advertising images (Courtney and Lockeretz, 1971, Snegupta, 1995, Shimp, 200). This research is of interest not only because it can be an important element of advertising strategy, but because advertising is an artifact that may reflect many of the attitudes, values, and behaviors that reinforce the consumer lifestyles of a society (Pollay 1986).

Most of the research on this topic has been conducted by scholars in the United States but there has been some examination of the topic in other countries as well (Milner and Higgs 2004). Several authors have suggested that much of the research in this area lacks an encompassing theoretical perspective (e.g., Albers-Miller, 1996). In response to this criticism, some recent studies have employed Hofstede's (1984) masculinity/femininity scale and its theoretical basis upon which to formulate hypotheses. This study extends the line of research to Romania, a country that is currently in a state of transition from a centrally-planned socialist economy characterized by virtually no consumer advertising to a free market-based economy featuring a robust and globally connected advertising industry. We use the Hofstede masculinity/feminity scale to formulate some of our hypotheses. Since television tends to make the most impact of all advertising media (Shimp, 2000), we chose to analyze commercials from this medium.

2. Literature Review

The earliest research on gender roles in advertising was conducted in the United States. In response to feminist charges that advertisers portray women in terms of a limited number of negative stereotypes, Courtney and Lockeretz (1971) conducted an exploratory study to determine whether such stereotypes could be identified. Employing a sample of ads from major American magazines published in 1970, they found that while very few ads portrayed women in a very negative way, many seemed to reflect stereotypes related to the traditional domestic roles of women. Several studies employing American samples appeared throughout the 1970s and into the 1980s. Bretl and Cantor (1988) reviewed this literature and compared it with results from their own sample.

They found persistent differences in level of employment depicted with men portrayed as having higher status jobs, being involved in outdoors activities, possessing credibility from the role of product authority, and performing the role of voiceover even in commercial for products targeted at women. Women were associated with the roles of spouse and parent, being at home, and using products used in the home. However, the authors found that the magnitude of many of the differences were diminishing and concluded that that the trend in American advertising seemed to be moving toward a less sexist and more equal representation of men and women in society.

Gilly (1988) compared gender roles in television commercials across samples from Mexico, Australia and the United States. She suggested that values consistent with female subordination seem to be universal. However, cultures vary in attitudes toward the proper roles for women in society and that this may be reflected in advertising images. She found that commercials from all three samples reflected traditional gender stereotypes. Men were used more often as voiceovers, shown more in independent roles while women were depicted as younger and in relation to others.

Comparing gender roles for women in commercials from Japan with those in the United States, Sengupta (1995) discovered a pattern of results largely consistent to that found by Gilly (1988). In the American commercials, men were more likely than women to appear in high status occupations. Women were more likely than men to appear in a decorative role. Cheng (1997) extended the comparison of American commercials with Chinese commercials. He found that in both samples men were more likely than women to appear in high status occupations while women more likely appeared in mid-level status occupations.

A pair of studies comparing gender roles in television commercials between Malaysia and Singapore rendered similar results. In the first, Wee, Choong, and Tambyah (1995) found that their results reflected the cultures of each nation. Tan, Ling, Theng, (2002) obtained similar results. They also found much traditional stereotyping of both men and women in Malaysian and Singapore.

This stream of research cited here has rendered many interesting results. The findings confirm the virtually universal subordinate role portrayals of women but also the trend in many countries toward a more egalitarian status with men. Also, the results are quite consistent with respect to advertising images reliably reflecting the values of national cultures and differences across national cultures as they pertain to gender roles. However, Albers-Miller (1996) has suggested that taken as a whole, cross-cultural research in advertising this area has rendered fragmented and often contradictory results because it lacked a comprehensive theoretical perspective for guiding hypothesis development. She suggested using Hofstede's typology of cultural values as a theoretical base for such research and provided empirical evidence linking advertising appeals to values within the typology (Albers-Miller and Gelb, 1996).

3. Cultural Values In Advertising

Of particular interest in this study is the Masculinity/Femininity dimension. According to Hofstede, masculine values include achievement, heroism, assertiveness and material success. Feminine values, alternatively, emphasize relationships, modesty, caring for the weak, and the quality of life. Masculine values have been linked with the advertising appeals of effectiveness, convenience, and productivity while feminine values have been linked with natural, frail, and modesty appeals (Albers-Miller and Gelb, 1996). Advertising appeals reflecting the gender values listed above should also be expected to occur (Moon and Chan 2005).

Milner and Collins (1998, 2000) have reported a pair of cross-cultural investigations of gender roles in advertising using Hofstede masculinity/femininity scores (MAS) across nations. Overall, the results are consistent with their predictions. In the first of these studies Milner and Collins (1998) compared gender role portrayals in Turkey, a nation with a slightly feminine Hofstede score of 45 with portrayals in commercials from three countries with masculine scores – Mexico (MAS = 69), United States (MAS = 62), and Australia (MAS = 61).

The data for the latter three samples was drawn from Gilly (1988). Results were consistent with predictions that, compared with commercial in the masculine countries, commercials from Turkey would reflect characters more often portrayed in relationships and less often in work situations. Taking a similar approach, Milner and Collins (2000) compared gender roles in commercials across sample from Japan, Russia, Sweden, and the United States. They found that characters in commercials from the feminine countries of Russia (MAS = 36) and Sweden (MAS = 5) than in the masculine countries of Japan (MAS = 95) and the United States (MAS = 62). The prediction that characters in the commercials from masculine countries would reflect more work situations than would commercials from the feminine countries was not supported. Romania's MAS score, which is estimated to be 42, reveals a slight tilt in favor of feminine values. Thus, one might expect to find a mixture of both traditional and modern portrayals of men and women in advertising.

The cultural differences between Sweden, Japan and Romania are summarized by Hofstede using five dimensions: power distance, individualism, masculinity, uncertainty avoidance, and long-term orientation. While Japan shows a very high masculinity index (of over 90), Sweden and Romania have much lower scores (of 5 and 42 respectively). Japan is a society where power distance is relatively high while individualism has a relatively low score. In Romania, while power distance is high, the society behaves more like a collective group.

Based upon the discussion above, we pose the following research question and hypotheses:

H1a: *Characters are more likely to be portrayed in relationships in Romanian commercials than are characters in Japanese commercials.*

H1b: *Characters are less likely to be portrayed in relationships in Romanian commercials than are characters in Swedish commercials.*

H2a: *Characters are less likely to be portrayed in employment situations in Romanian commercials that are characters in Japanese commercials.*

H2b: *Characters are more likely to be portrayed in employment situations in Romanian commercials that are characters in Swedish commercials.*

H3a: *There are fewer significant differences in gender role portrayals in Romanian commercials than there are in Japanese commercials.*

H3b: *There are more significant differences in gender role portrayals in Romanian commercials than there are in Swedish commercials.*

4. Method

Sample. A sample of 351 commercials was captured within 37 hours of programming from popular public and nonpublic cable networks in Romania -- ProTV and Antena 1 (both private) and TRV1 (public). Programming was recorded in two day parts – mid afternoon and prime time during May of 2005 and 2006. Duplicate commercials were eliminated from the analysis.

Coding. Categories of analysis along with their operationalization were virtually all adopted from earlier research. Coding procedures were conducted based upon widely accepted standards for content. Two adults fluent in both Romanian and English independently coded the characteristics of each commercial and of each character. Preceding the main data collection phase they participated in two training sessions. Disagreements were resolved through discussion between the judges. Initial reliability coefficients are all well above the minimum necessary for adequate reliability analysis (Neuendorf 2002).

5. Results

The categories with the highest proportions of commercials included food and beverages (29.2%) and personal care and cosmetics (23.6%). A total of 590 central characters were examined. The ratio of characters to commercials was 1.68. The sample included a few more male characters (308, 52.2%) than female characters (282, 47.8%). Results indicate a significant difference between men and women emerged with respect to the product categories of the commercials in which they appear (χ^2 (10df) = 43.15, $p \leq .01$). Women appear at a higher rate than did men in commercials for personal care and cosmetics (67.6 to 32.4%) and commercials for household cleaning agents (62.5 to 37.5%).

A significant difference between the gender groups was also found according to the target market for the product (χ^2(2df) = 50.95, $p \leq .01$). Compared with men, women were more likely to appear in commercials for products targeted primarily at women (81.3 to 18.7%). Men were more likely to appear in commercials for products targeted primarily at men (81.8 to 18.2%) and at both or either gender group (57.8 to 42.2%). There was a significant difference between the gender groups in the setting of the ad (χ^2(4df) = 22.82, $p < .01$). Women were more likely to appear in ads with home settings (43.3% to 34.1%) and studio settings (11.0% to 3.2%). Men were more likely to appear in work settings (12.0 to 7.1%) and outdoor settings (36.7 to 28.4%). The groups were about equally likely to appear in store/restaurant settings (18.20% for women and 17.4% for men). The use of female and male voiceovers in the commercials was quite similar. Twenty-six percent of the commercials featured a female-only voice over and 24.5% featured a male-only voice over. A large proportion (42.5%) featured both a female and male voiceover and 6.8% included no voiceover. Finally, there was a significant difference between the gender groups appearing in commercials using a sexual appeal (χ^2(1df) = 15.47, $p \leq .01$) with women appearing at a higher rate than did men (33.0% to 18.8%).

Results from Milner and Collins (2000) study for Sweden and Japan were used to compare with the results of this study. The Milner and Collins study employed a very similar methodology so the results are directly comparable. The first hypothesis predicted that compared with characters in Romanian commercials (moderate Hofstede masculinity score) those in Japanese commercials (very high Hofstede masculinity score) are less likely to be shown in relation to others and those in Swedish commercials (very low Hofstede masculinity score) are more likely to be depicted as being in relation with others. Consistent with the prediction more characters are shown in relation to others in Romanian commercials than in Japanese commercials (63.7% to 37.2%). However, contrary to the hypothesis, more characters are also shown in relation to others in Romanian commercials than in Swedish commercials (63.7% to 57.4%). Thus, hypothesis one receives only partial support.

Hypothesis two predicted that characters in Romanian commercials are less likely than those in Japanese commercials to be shown in at work settings and more likely than those in Swedish commercials to be shown in at work settings. Consistent with this hypothesis characters in Romanian commercials are shown in at work setting less than in Japanese commercials (9.7% to 16.8%). Contrary to the hypothesis characters in Romanian commercials are shown in at work setting less than in Swedish commercials (9.7% to 25.7%). Hypothesis 2 receives only partial support.

The third hypothesis predicted that there would be fewer significantly different gender role portrayals in Romanian commercials than there would be in Japanese commercials and more than there would be in Swedish commercials. When comparing measures on variables that were taken in both the Milner and Collins (2000) and the current study we find five significant differences in the Japanese sample, three in the Romanian sample and three in the Swedish sample (Table 1). Thus, Hypothesis 3 is partially confirmed.

A comment regarding the difference in coding approach for marital status is in order. Milner and Collins (2000) coded portrayals as married, not married, or not identified. We decided to code a character as married only if it was fairly obvious that the character was depicted as married. We decided that we could not reliably code a character as not married vs. not identified. So we decided that if a character was not coded as married, we would code them as not married. Perhaps it would have been better to code the not married as not identified. If we move the “not married data to the not identified cells this is corrected and the 2X2 Chi-square test renders the same results.

The relatively high proportion of *active* portrayals within the Romanian sample reflects another difference in coding approach. Milner & Collins (2000) following Gilly (1988) coded the active as engaging in physical activity or sport. Our active category included recreational activity, interacting with family members or friends, and doing household chores. Thus, our scope of activity is much greater than was theirs and we therefore got a much higher percentage of portrayals as active.

Marital Status

We decided to code a character as *married* only if it was fairly obvious that the character was depicted as married. We decided that we could not reliably code a character as *not married* vs. *not identified*. So we decided that if a character was not coded as *married*, we would code them as *not married*. Perhaps it would have been better to code the *not married* as *not identified*. If we move the "*not married* data to the *not identified* cells this is corrected and the 2X2 Chi-square test renders the same results.

6. Discussion and Conclusion

As hypothesized, the gender portrayals exhibited in Romanian television commercials reflect a mixture of traditional and modern roles for both groups. These findings are consistent with the results of the majority of previous studies with commercials from other countries that show an increasingly varied portrayal of women reflecting an increase in their economic power and a concomitant shift in social norms. Consistent with more modern gender roles were the findings that female and males are used equally often as voiceovers and as spokespersons in commercials. Further, the lack of significant difference in portrayals of men and women performing household chores seems consistent with modern values. Finally, the greater proportion of portrayals of men as frustrated and as foolish compared with women seem consistent with modern values.

Yet the more traditional stereotypes are still apparent. Traditional gender roles for women are reflected in their association with sexual appeals and with domestic products, especially personal and beauty care products and household cleaning products. These associations reflect stereotypes related to concern for physical appearance and the housekeeping role. Compared with men, women were shown in higher proportions in home settings and in inactive roles. Traditional male roles are reflected in their association with alcoholic beverages and tobacco, work settings, outdoor settings, product authorities, active roles, and recreational activities.

Finally, except for the comparison with commercials from Milner and Collins' (2000) sample of Swedish commercials these results are consistent with predictions based on the Hofstede masculinity index and add support to the validity of using the index to predict patterns in advertising within national cultures. As expected commercials in Romania, a culture with a slight leaning toward feminine values, reflected more relationship and fewer work portrayals than did commercials from the highly masculine culture of Japan. Lastly, the predicted ranking in the number of significant differences in commercials across the three countries partially emerged. The highly masculine culture of Japan had the most, followed by slightly feminine Romania, and highly feminine Sweden.

References

Albers-Miller, N.D. (1996), 'Designing Cross-Cultural Advertising Research: A Closer Look at Paired Comparisons,' International Marketing Review, 13 (5), pp. 59-75.

Albers-Miller, N.D. & Gelb, B.D. (1996), 'Business Advertising Appeals as a Mirror of Cultural Dimensions: A Study of Eleven Countries,' Journal of Advertising, 15 (4), pp. 57-70.

Bretl, D.J. & Cantor, J. (1988), 'The Portrayal of Men and Women in U.S. Television Commercials: A Recent Content Analysis and Trends over 15 Years,' Sex Roles, 18(9/10), pp. 595-609.

Cheng, H. (1997), "Holding Up Half Of The Sky"? A Sociocultural Comparison Of Gender-Role Portrayals In Chinese And US Advertising,' International Journal of Advertising, 16(4), pp. 295-319.

Courtney, A.E. & Lockeretz, S.W. (1971), 'A Woman's Place: An Analysis of the Roles Portrayed by Women in Magazine Advertisements,' Journal of Marketing Research, 8(Feb.), pp. 92-95.

Europa World Year Book (2007), London: Taylor and Francis.

Federal Research Division (1989), Romania: Country Report, Library of Congress.

Ford, J.B., Voli, P.K., Honeycutt, E.D. & Casey, S.L.. (1998), 'Gender Role Portrayals in Japanese Advertising: A Magazine Content Analysis,' Journal of Advertising, 27(1), pp. 113-124.

Furnham, A. & Mack, T. (1999), 'Sex-Role Stereotyping in Television Commercials: A Review and Comparison of Fourteen Studies Done on Five Continents over 25 Years,' Sex Roles, 41(5/6), pp. 413-437.

Gilly, M. (1988), 'Sex Roles in Advertising: A Comparison of Television Advertisements in Australia, Mexico, and the United States,' Journal of Marketing, 52 (2), pp. 75-85.

Hofstede, G. (1984), Culture's Consequences, Thousand Oaks, CA: Sage.

Hofstede, G. & Bond, M.H. (1988), 'The Confucius Connection: From Cultural Roots to Economic Growth,' Organizational Dynamics, 16(4), pp. 4-21.

Hofstede, G. Kolman, L., Nicolescu, O., & Pajumaa, I. (1996), 'Characteristics of the Ideal Job Among Students in Eight Countries,' in Grad, H., Blanco, A., & Georgas, J. (eds.) Key Issues in Cross-Cultural Psychology, Lisse, Netherlands, Swets & Zeitlinger, pp. 199-216.

Milner, L.M. & Collins, J.M. (2000), Sex-Role Portrayals and the Gender Of Nations, Journal of Advertising. 29(1), pp. 67-79.

Milner, L.M. & Collins, J.M. (1998), 'Sex Role Portrayals in Turkish Television Advertisements: An Examination in Global Context,' Journal of Euromarketing, Vol. 7 (1), pp. 1-28.

Milner, L.M. & Higgs, B. (2004), 'Gender Sex-Role Portrayals in International Television Advertising over Time: The Australian Experience,' Journal of Current Issues and research in Advertising, 26(2), pp. 81-96.

Moon, Y.S. & Chan, K. (2005), 'Advertising Appeals and Cultural Values In Television Commercials: A Comparison Of Hong Kong And Korea,' International Marketing Review, 22 (1), pp. 48-66.

Neuendorf, K.A. (2002), The Content Analysis Guidebook, Thousand Oaks, Ca:

Sengupta, S. (1995), 'The Influence of Culture on Portrayals of Women in Television Commercials: A Comparison between the United States and Japan,' International Journal of Advertising, 14 (4), pp. 314-333.

Shimp, T.A. (2000), Advertising and Promotion, Fort Worth TX: Dryden.

Stoica, M., Ardelea, D. & Miller, D.W. (2006), 'Trends in the Romanian Advertising Industry: Positioning, Competition, and Selection,' Review of International Comparative Management, The Academy of Economic Studies Bucharest, 8, in press.

Tan, T.T.W., Ling, L.B., & Theng, E.P.C. (2002), 'Gender-Role Portrayals In Malaysian And Singaporean Television Commercials: An International Advertising Perspective,' Journal of Business Research, 55(10), pp. 853-861.

Wee, C.H., Choong, M.L. and Tambyah, S.K. (1995), 'Sex Role Portrayal in Television Advertising: A Comparative Study of Singapore and Malaysia,' International Marketing Review, 12 (1), pp. 49-64.

Table 1: Number of Differences Across Romanian, Japanese, and Swedish Commercials

Characteristic	Sweden*		Romania		Japan*	
	Women	Men	Women	Men	Women	Men
Age						
Young	77.9	35.1	90.4	86.0	70.9	46.2
Middle	20.3	50.5	7.1	11.7	26.5	39.4
Old	1.9	14.4	2.5	2.3	2.6	14.4
	$\chi^2 = 68.19$		n.s.		$\chi^2 = 30.12$	
Marital Status						
Married	12.7	6.3	36.5	32.1	26.9	9.1
Not Married	19.6	23.6			22.7	28.8
Not Identified	67.7	70.2	63.5	67.9	50.4	62.1
	n.s.		n.s.		$\chi^2 = 16.49$	
Spokesperson						
Yes	82.8	89.0	5.0	3.9	55.6	50.0
No	17.2	11.0	95.0	96.1	44.4	50.0
	n.s.		n.s.		n.s.	
Credibility						
User	20.0	45.8	98.5	97.2	95.2	79.4
Authority	80.0	54.2	1.5	2.8	4.8	20.6
	n.s.		n.s.		$\chi^2 = 11.42$	
Role						
In Relation with Others	60.8	54.8	61.3	65.9	41.1	30.3
Independent	36.1	44.7	38.7	34.1	50.7	61.4
	$\chi^2 = 6.13$		n.s.		n.s.	
Activity						
Active	12.7	9.1	79.4	88.0	5.2	20.5
Inactive	87.3	90.9	20.6	12.0	94.9	79.6
	n.s.		$\chi^2 = 7.97$		$\chi^2 = 20.68$	
Frustrated						
Yes	8.2	89.9	13.5	22.1	6.0	10.6
No	91.8	10.1	86.5	77.9	94.0	89.4
	n.s.		$\chi^2 = 7.39$		n.s.	
Work						
Yes	13.9	34.6	7.1	12.0	10.5	27.9
No	78.5	64.6	92.9	88.0	75.4	50.4
	$\chi^2 = 27.81$		$\chi^2 = 4.08$		$\chi^2 = 25.48$	

*Milner & Collins (2000)

A 2-D Clustering Algorithm to Aid Mine-Warfare Analysts in Determining Navigable Lanes over Cluttered Seafloors

Marlin L. Gendron and Maura C. Lohrenz
Naval Research Laboratory, Stennis Space Center
mgendron@nrlssc.navy.mil, mlohrenz@nrlssc.navy.mil

Abstract

Naval operations, especially in shallow-water regions, require safe navigable lanes (i.e., clear of explosives). During mission planning, mine warfare experts analyze sidescan sonar imagery of the seafloor in regions of interest to help characterize the seafloor bottom type. One of the environmental parameters derived from sidescan is seafloor "clutter", or density of mine-like echoes. Regions of low clutter predict safer passage with less time-consuming, lane-clearance effort. A measure of clutter density traditionally was manually estimated, but this tends to be labor-intensive and produced inconsistent results (e.g., depending on the analysts' training and level of fatigue). The authors have developed an automated algorithm to cluster mine-like echoes quickly and consistently. Numerous tests have been performed during mine warfare missions, and the Naval Oceanographic Office has concluded that results generated by this clustering algorithm are equivalent or superior to manual clustering methods. Because the algorithm is also more reliable (repeatable), faster, and requires fewer personnel, it has been adopted by the Navy for their mine warfare operations.

1. Introduction

The Naval Oceanographic Office (NAVOCEANO) characterizes the seafloor environment during military Mine Warfare (MIW) operations to determine the least time- and resource-consuming geographic areas to clear assault lanes of mines so naval ships can pass safely. A primary component of bottom-type estimation, derived from sidescan sonar imagery, is seafloor clutter density. Seafloor regions with low clutter density are easier to clear of mines because there are fewer mine-like echoes (MILECs), or clutter, that might be mistaken for mines.

An object on the seafloor is defined as clutter in MIW if it appears in grayscale sidescan imagery as a bright region adjacent to a shadow. The bright region must be closer to the sonar than the shadow. Clutter density is the number of MILECs per cluttered region. To determine clutter density, imagery analysts view many hours of recorded sidescan imagery to find MILECS and plot their latitude and longitude coordinates. An analyst manually divides the plotted MILECs into bounded clusters, estimates the area of each cluster, and calculates the clutter density based on the number of MILECs in the cluster per area of the cluster.

To reduce the time to determine clutter density and minimize inconsistent results, the authors developed an automated clutter detection algorithm to determine the location of MILECs and an automated clustering algorithm to cluster the MILECs and determine clutter density.

This paper presents NRL's automated clustering algorithm, which is capable of producing polygons and computing clutter density from a set of MILEC locations. The results are repeatable, do not depend on the ordering of MILEC locations, and closely match clusters produced manually.

2. Clustering with geospatial bitmaps

The Geospatial Bitmap (GB) clustering algorithm is non-iterative and well-suited to autonomous clustering applications, because the order in which elements are input to the algorithm has no effect on the resulting clusters, no seed point is required to initiate clustering, and no Euclidean distance is needed for cluster selection.

2.1 GB overview

The algorithm relies on a patented GB concept based on simple bitmaps – binary structures (e.g., a binary image) in which bits are turned on (set = 1) or off (cleared = 0) [1]. The index of each bit is unique and denotes its position relative to other bits in the bitmap. A set bit indicates an element of interest exists at that location, accurate to within the resolution of the bitmap. A cleared bit indicates the absence of any element at that location. The patent describes a two-dimensional (2-D) GB, in which a bit is indexed by column and row. Although a GB can be defined for an entire finite space, memory is only allocated (dynamically) when groups of spatially close bits are set, resulting in a compact data structure that supports very fast Boolean and morphological operations.

Binary images have been used in combination with clustering techniques [2],[3],[4] but GBs have never been used to drive the clustering directly. GB clustering is similar to cluster analysis by binary mathematical morphology, which is iteratively applied to detect cluster cores and eliminate irrelevant details in the cluster shapes [5].

2.2 GB algorithm parameters

The algorithm provides an efficient way to cluster any set of unique data elements in 2-D space into bounded polygons. In this section, a derivation is presented to show how a set of elements are represented in a GB and then "expanded" by a specified mapping called an expansion shape. The expansion produces bounded regions or clusters, which are stored in a new GB. When the expansion mapping is uniform, this process is equivalent to dilation of a structuring element, as defined in mathematical morphology [6]. The GB process differs in that the expansion mapping does not have to be uniform. For example, given a priori knowledge of the dataset, the dimensions of the expansion shape could depend on the distance from an element of interest to a specified origin.

The algorithm requires five arguments:

1. A set D of unique data elements;
2. A mapping P' used to expand the elements of D;
3. A mapping M' that maps the input space S into a continuous Cartesian space GS;

4-5. Resolutions rx and ry, in which the elements of D will be represented in a discrete Cartesian space GB.

P' is applied to each element d_i of D to produce set T_i:

$T_i = \{P'(d_i)\} \mid d_i \in D$, where $T_i \subset S$ (eq. 1)

Let U_i be the set created by applying M' to T_i:

$U_i = \{M'(t)\} \; \forall \; t \in T_i$ (eq. 2)

Let GS be the Minimum Bounding Box (MBB) of the union of the expanded and mapped elements of D:

$GS = MBB(U_i \cup \ldots \cup U_n)$ (eq. 3)

A "cleared" GB is then created as a discrete approximation of the continuous space GS at resolutions rx and ry.

2.3 Examples in 2-D Cartesian space

The following example illustrates the GB clustering algorithm using continuous values {latitude, longitude} to cluster elements in 2-D Cartesian space. First, a GB is defined such that each bit contains at most one element (figure 1). Each element is represented as a “set bit” (value=1) with an index to identify its position. All other bits in the GB are cleared (bit value=0).

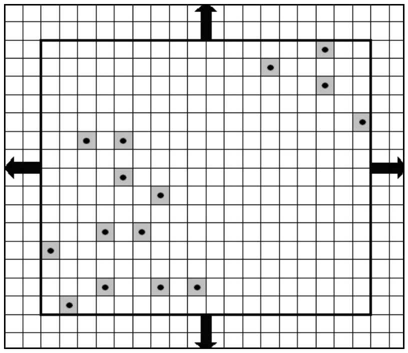


Figure 1. Elements are mapped to a GB, which is enlarged to accommodate expansion of outermost set bits in region of interest.

Next, an expansion shape is defined to control how clusters will form around the elements. Any shape with dimension less than or equal to that of the element space can define a cluster. The expansion shape used in this example is a 5x5-bit square (figure 2). Using a circle would maintain an equidistant expansion around each set bit. Other shapes could be chosen to preserve or accentuate a pattern in the data. For example, a long, narrow ellipse might be an appropriate expansion shape to cluster a set of elements known to lie roughly along a line.

The GB must be enlarged just enough to accommodate the expansion shape in any direction. The new GB (GB1) is determined by the MBB of the expansion of all elements to be clustered. Here, the GB is enlarged by two bits along each edge (figure 1). Each element is expanded by setting the surrounding bits according to the expansion shape. Geographically close elements will cluster together (figure 3). The size of the expansion shape dictates the size of the resultant clusters: a larger expansion results in larger clusters with greater maximum element spacing, while smaller expansions result in smaller, tighter clusters.

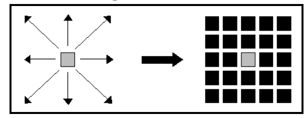


Figure 2. Example of expanding a set bit.

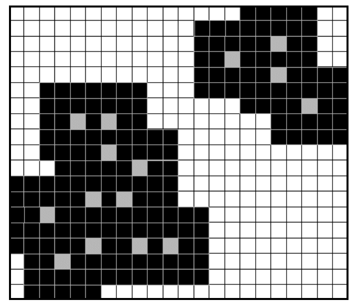


Figure 3. Geospatially close elements are clustered together as expanded bits merge.

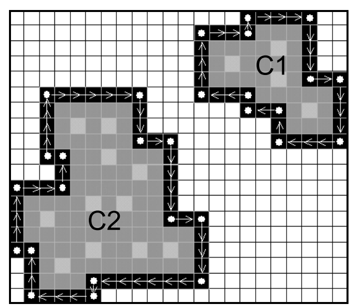


Figure 4. Boundary vertices are obtained.

The next step is to reduce each cluster to a set (or graph) of 1-bit vertices that defines the cluster’s bounding polygon. A copy of the bitmap is made, and all internal bits of each cluster in the copy are cleared (e.g., the gray bits in figure 4). Traversal of the cluster starts at a convex vertex, for which the interior angle is less than 180 degrees, and for which there is knowledge of the vertex’s relationship to the polygon. For example, if traversal starts at an upper-left vertex of the cluster, the edge leaving that vertex in a clockwise direction is to the right (figure 4).
From this vertex, traversal continues clockwise around the cluster: at each bit along the boundary, the algorithm tests

whether the next set bit is in one of three directions from the current bit, in the following order: 1) 90° counter-clockwise from the current bit, or 2) in the same direction as the previous iteration, or 3) 90° clockwise from the current bit. An allowable direction is based on whether the new bit is both set and not part of a different polygon.

For example, in figure 4, clockwise traversal of clusters C1 and C2 (one at a time) would progress as follows:

1. Start at the upper-left-most bit in the cluster, which is the first vertex in the traversal, and move one bit to the right (clockwise). Set this bit, since traversal started with the upper-left-most bit.
2. Test whether the bit immediately above (90° counter-clockwise) is set (here, no).
3. Test whether the bit immediately to the right (continuing in the same direction) is set: if so, this becomes the current bit.
4. Repeat steps 2-3, moving one bit at a time along the top edge of the cluster, until the next vertex is reached. At that point, the tests in steps 2 and 3 will fail, and the algorithm will try the third directional test: whether the bit immediately below (90° clockwise) is set, which it is in this example. When the direction of traversal changes, the bit is tagged as a vertex (figure 4) and the process continues.
5. Traversal is complete when it returns to the upper-left-most vertex in the cluster.

One vertex can belong to more than one cluster. In figure 5a, C1 and C2 are stored as two clusters, since the algorithm does not check for set bits diagonally adjacent to each other. In figure 5b, C1 and C2 share a vertex, but they are still considered two clusters. In figure 5c, the regions overlap by more than one bit and are considered one cluster.

The algorithm determines whether there is one cluster or two adjacent clusters by examining the 8-bit neighborhood around the bit in question. (This test only occurs if more than one direction could be taken from the current bit; i.e., if the bits 90° counter-clockwise *and* straight ahead from the current bit are set, as in figures 5b and 5c). If two cleared bits are diagonally opposite each other and flank the current bit, the last possible direction is taken (e.g., 90° clockwise in figure 5b). If not, the first possible direction is taken (e.g., 90° counter-clockwise in figure 5c). In each case, traversal continues until it returns to the first vertex, at which point the cluster's bitmap is reduced to a vector boundary, and a graph of all vertices defining the bounding polygon is created (figure 6).

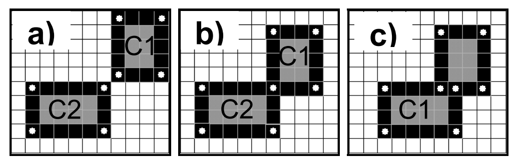


Figure 5. Test for one or two clusters.

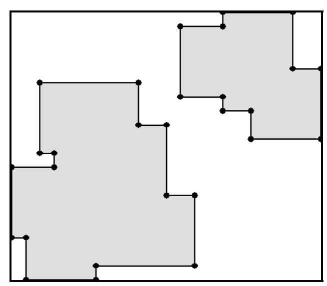


Figure 6. Convert bitmaps to vector boundaries.

To determine the number and locations of all elements in each cluster, a logical AND is performed on each GB with GB1; e.g., GB2 = GB AND GB1. The resulting GB2 is traversed to obtain the number of set bits and their locations.

Each bounding region, defined by its vertices, is smoothed with a polyline-smoothing algorithm [7] to reduce the number of vertices needed to define the cluster. Finally, the GB algorithm calculates cluster density as the number of elements in the bounded cluster region divided by the region's area.

4. GB algorithm vs. manual clustering

Many tests have been performed to compare results of the GB clustering algorithm with manual clustering by expert analysts. The following presents a typical comparison conducted by NAVOCEANO.

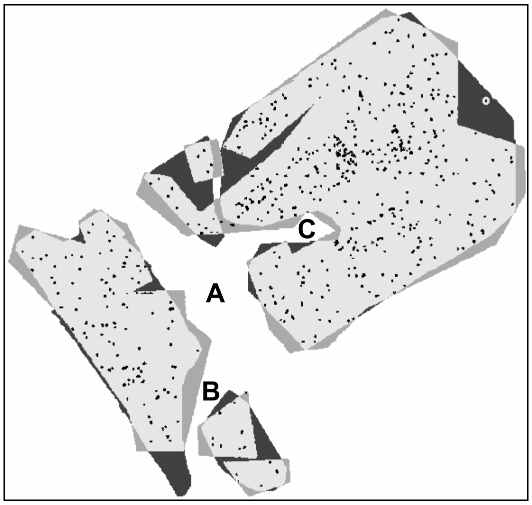


Figure 7. Comparison of clustering results from manual (by NAVOCEANO analyst) and automated (using GB algorithm) methods.

Table 1. Comparison of clustering methods.

Description	#pixels	%image
Match	105325	82%
GB only	11337	9%
Manual only	11874	9%

Figure 7 depicts a large MIW dataset (550 objects detected in acoustic imagery) manually clustered by a Navy analyst and automatically clustered by the GB algorithm using an expansion circle with radius 250m. Previous GB clustering trials on similar datasets determined 250m is an appropriate expansion size for MIW applications, based on the desired width of object-free zones, discussed below.

The lightest gray areas in figure 7 represent cluster regions that overlap (i.e., where manually drawn clusters coincide with GB-generated clusters). The medium gray areas represent portions of clusters that GB generated but not the analysts. The darkest gray areas represent portions of clusters that the analysts drew but GB did not include.

A comparison of clusters in figure 7 (and corresponding data in table 1) shows that GB-generated clusters matched manual clusters for 82% of the operational area. Mismatches were equivalent (9% each) between manual-only and GB-only cluster regions. This comparable performance is particularly important when examined from a MIW perspective, in which a primary goal is to clearly mark object-free zones (to assist military vessels in navigating more safely through potentially hazardous areas). Both methods reveal one primary object-free lane (labeled “A” in the figure), one narrower lane (B), and an inlet (C).

After clustering the data, analysts count the number of objects in each cluster and determine the area of each bounded region to calculate cluster densities. Military operators use this information to decide whether a region is too dangerous to penetrate or clear of mines. The GB algorithm calculates cluster density automatically by keeping track of which objects belong to each cluster and determining the area of each bounded region after generating the region’s vertices.

Manual clustering outcomes (and the time required to complete clustering) can vary significantly, depending on the experience of the analyst and the complexity of the dataset. Two different analysts might cluster the same dataset completely differently. The GB algorithm produces identical results on every run, given the same dataset and the same expansion shape. After numerous successful trials, satisfied analysts at the Naval Oceanographic Office have replaced their manual techniques with the GB algorithm to assist them in clustering detected objects in acoustic imagery more consistently and efficiently.

5. Conclusion

This paper presents the GB clustering algorithm, which provides a repeatable method to cluster elements into bounded regions and calculates cluster density. The GB clustering algorithm was compared with manual clustering of a large dataset consisting of mine-like objects detected in sidescan sonar imagery. With an appropriately sized expansion shape, the GB algorithm succeeded in mimicking an expert analyst's manual clustering, in which a primary goal was to emphasize object-free zones. A circular expansion was used with 250m radius, chosen to maximize the number of object-free zones that would be wide enough for a Navy MIW vessel to navigate through. In this case, the expansion shape size was clearly dictated by Navy requirements. One limitation of the GB technique for other clustering applications might be determining how best to set this parameter.

Because the GB clustering algorithm is efficient and repeatable, significantly reduces workload for MIW analysts, generates cluster boundaries, and calculates cluster density, analysts at the Naval Oceanographic Office have adopted this algorithm to automatically cluster mine-like objects detected in side-scan imagery and determine the density of these objects in areas of interest, significantly aiding naval MIW operations. Numerous tests have been performed during MIW missions, and NAVOCEANO has concluded that results generated by the GB clustering algorithm are equivalent or superior to manual clustering methods. Because the algorithm is also more reliable (repeatable), faster (especially for large datasets), and requires fewer personnel, it has been adopted by NAVOCEANO for their MIW operations.

The authors have applied for U.S. and foreign patents (Navy Case Numbers 84,921 and 97,458, respectively) and now are investigating its potential for quantifying clutter in electronic displays [8],[9].

6. Acknowledgements

This work was sponsored under Program Element 602435N by the NRL 6.2 Base Program. The authors thank Mr. Jim Hammack and Mr. Ron Betsch at NAVOCEANO for their support of this project. We thank Ms. Peggy Haeger of NAVOCEANO for her tireless testing of this algorithm.

7. References

[1] Gendron, M., P. Wischow, M. Trenchard, M. Lohrenz, L. Riedlinger and M. Mehaffey (2001). "Moving Map Composer." U.S. Patent 6,218,965, April.

[2] Hobby, J. and T.K. Ho (1997). "Enhancing degraded document images via bitmap clustering and averaging. "Proceedings of the 4th International Conference on Document Analysis and Recognition, Ulm, Germany, 18-20 August.

[3] Ho, T.K. and G. Nagy (2000). "OCR with no shape training," Proceedings of the 15th International Conference on Pattern Recognition, Barcelona, Spain, 3-8 September.

[4] Yoon, J.P., V. Raghavan, and V. Chakilam (2001). "BitCube: a three-dimensional bitmap indexing for XML documents," J. Intelligent Information Systems 17.

[5] Postaire, J.G., R.D. Zhang, and C. Lecocq-Botte (1993). "Cluster analysis by binary morphology," IEEE Trans. on Pattern Analysis and Machine Intelligence, 15:2.

[6] Gonzales, R.C. and R.E. Woods (2002), "Digital Image Processing, 2nd Ed." Prentice Hall, New Jersey.

[7] Layne, G.J., M.L. Gendron and M.C. Lohrenz (2004). "POS polyline smoothing: reduction of polyline vertices," Proceedings of the 10th International Conference on Industry, Engineering and Management Systems, Cocoa Beach, Florida, March.

[8] Lohrenz, M., G. Layne, S. Edwards, M. Gendron, and J. Bradley (2006). "Feature clustering to measure clutter in electronic displays." Proceedings of the 12th International Conference on Industry, Engineering, and Management Systems, Cocoa Beach, FL, March.

[9] Lohrenz, M. and M. Gendron (in press). "A 3D clustering algorithm to model clutter in electronic geospatial displays." J. of Management and Engineering Integration. Winter, 2009.

Comparing the Intelligibility of Bone-Conducted Speech Using Different Test Materials

Benjamin Osafo-Yeboah, Ritson Delpish, and Xiaochun Jiang
North Carolina A & T State University
xjiang@ncat.edu

Abstract

Speech intelligibility is the degree to which a speaker's intended message is accurately interpreted by a listener. The goal of this research is to compare the intelligibility of bone conducted speech using two different speech intelligibility testing materials (the Callsign Acquisition Test (CAT) and the Modified Rhyme Test (MRT)) and compare the results to previous studies. Results indicated that there was no significant difference between the CAT test and the MRT test with the babble noise as the background. Considering that soldiers are more comfortable with the CAT test, it indicates that the CAT test is a good option to use in military applications when conducting bone conducted speech studies.

1. Introduction

Two-way radio communication plays a very important role in many applications such as in construction, traffic control and battleground. There are two alternative two-way radio communication systems: via Air conduction and via bone conduction. In military applications, air conducted radio communication is often used to receive and transmit information among infantry and between command and control on the battlefield. The importance of having an effective and efficient communication cannot be overemphasized. To a large extent, the soldiers' lives may rely on their ability to hear and understand communicated speech clearly on the battleground. Delivering the right information to soldiers at the right time and in a form that is clear can be the difference between life and death for these brave soldiers [1]. In other words, the intelligibility of the speech transmitted needs to be very high and reliable. Unintelligible speech is worthless, and low intelligible speech can be even worse.

Air conduction systems convey sounds from a sound source to a microphone and from a loudspeaker or earphones to the listeners' ear. For example, headphones can be used to deliver auditory signals to the cochlea through the medium of air. The sound waves are collected by the pinna and transmitted to the inner ear through the ear canal. Traditionally air conduction is more commonly used in industry.

Besides air conduction, the human auditory system is also sensitive to pressure waves that are transmitted through the bones in the skull [2]. This mechanism of sound transmission is called bone conduction. In bone conduction, the sound waves are transmitted to the inner ear by the cranial bones without passing traveling through the ear canal. This enables bone conduction headsets to deliver sound through direct application of vibrators to the skull and provide privacy and portability similar to those offered by headphones while at the same time leaving the ear canal and pinna unobstructed [3].

These features of bone conduction make it a good candidate for military applications. Comparing with the air conducted communication systems, bone conducted systems have several advantages in military applications. Among them, the most important advantage is the improvement of situation awareness and the ability to localize ambient noise. These advantages can make soldiers safer on the battleground. In addition, the transducers in bone conducted communication systems are lightweight, inconspicuous, and can easily be integrated into military headgear [4].

One important question for both bone conducted and air conducted speech is the intelligibility of the speech. Speech intelligibility can be defined as the percentage of speech units that can be correctly identified by a listener over a given communication system in a given acoustic environment or the degree to which speech can be understood within a given acoustic environment [5-7].

Speech intelligibility is a measure of sound clarity that indicates the ease of understanding of a speech. It can be expressed as a percentage of words, sentences or phonemes (speech sounds making up words) correctly identified by a listener, and it is a complex function of psychoacoustics, signal-to-noise ratio of the sound source, and direct-to-reverberant energy within the listening environment. Speech intelligibility can be affected by many factors such as background noise, and the type of applications. For instance, in some clinical tests, some of the information communicated through speech is contained within contextual, visual and gestural cues, it is therefore possible to understand meaning even if only a fraction of the discrete speech units are heard correctly [6-7]. On the other hand, visual and gestural cues may be hard to detect in military environments. Soldiers must therefore, rely more heavily upon the sound that is actually transmitted over their communication system on the battlefield.

To help study issues related to speech intelligibility, speech intelligibility test materials are often used. Commonly used speech intelligibility tests include the Modified Rhyme Test (MRT), Diagnostic Rhyme Test (DRT), Northwestern University Test number 6 (NU-6), Diagnostic Medial Consonant Test (DMCT) and Central Institute for the Deaf Test (W-22) among others [8]. However, these tests have been criticized for having poor validity in military settings due to the fact that the tests were developed primarily for clinical diagnostic testing. There is a test material developed by the Auditory Research Team (ART) of the Army Research Laboratory Human Research and Engineering Directorate (ARL-HRED) specifically for military applications. This test material is the Callsign Acquisition Test (CAT). The CAT uses military code words familiar to soldiers and therefore has greater appeal among military personnel [6-7].

Despite the advantage of the CAT test, its effectiveness remains to be seen. It is therefore, necessary to compare the speech intelligibility between CAT and traditional test materials. Since the MRT is the most widely used speech intelligibility test available. Although there were studies in the past that compared CAT with MRT, they were carried out using air conduction. In this study, bone conduction will be used instead of air conduction. Since the focus of this study was on the comparison, a babble background was used for both tests due to the fact that babble noise most resembles actual battleground environment.

Previous studies [8,12] have shown that condyle location (Figure 1) is appropriate for the bone conducted speech studies. Therefore, condyle was chosen in this study to be the location where the bone vibrator was place.

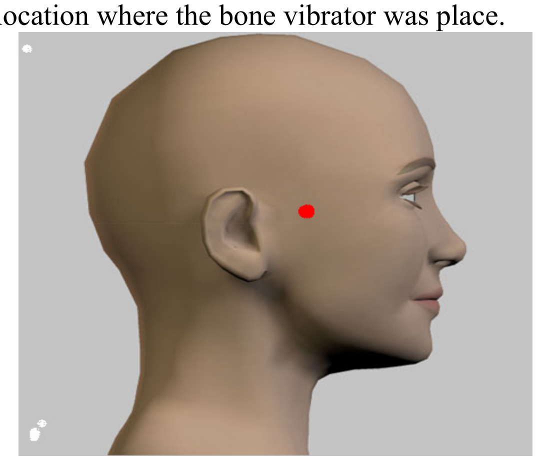


Figure 1. Condyle location on a skull

The objective of this study is to compare the intelligibility of bone conducted speech using two different speech intelligibility testing materials (CAT and MRT) with babble background noises and compare the results to previous studies.

2. Methodology

2.1 Participants

Six undergraduate students (3 males and 3 females) between the ages 20 and 26 years old (mean age = 22.5 and standard deviation = 2.1), enrolled at North Carolina Agricultural and Technical State University participated in the study). All of them were African Americans. The lack of diversity in the participant pool reflects the fact that NCA&T is a Historically Black University. All participants in this study were required to have a normal hearing defined by a threshold better than or equal to 20 dB HL (hearing level) at audiometric frequencies of 250 Hz to 8 kHz (ANSI S3.6-1996) [9]. For each individual, the difference between the left and right ear thresholds had to be 10 dB or less to ensure hearing symmetry. The audiometric screening was performed prior to the experiment. The screening involved standardized clinical equipment and procedures and complied with the ANSI S3.1-1991[10] requirements for audiometric testing under earphones. This was carried out with a Fonix Hearing Evaluator FA-12 Digital Audiometer in a sound treated booth. Participants who passed the audiometric screening were invited to participate in the study.

2.2 Stimulus Materials

CAT-60 with sixty test items was used in this experiment. Gripper [6] reported that this list has the capability of providing the same predictive power as the full CAT with good test-retest reliability. MRT with 50 six-word lists of rhyming or similar-sounding monosyllabic English words were used as the speech intelligibility test materials as well. Figures 2 and 3 provide the screenshots for the CAT and MRT tests that were randomly presented to listeners.

The test items presented at 80 dB SPL were mixed with babble background which was presented at 80 dB SPL using the Sound Forge 8.0 software to produce a signal-to-noise ratio (SNR) of –9dB. In order to ensure accurate measurement of the sound intensity produced by the bone vibrator, an artificial mastoid was used together with a sound level meter to measure the intensity of the sound output of the bone vibrator.

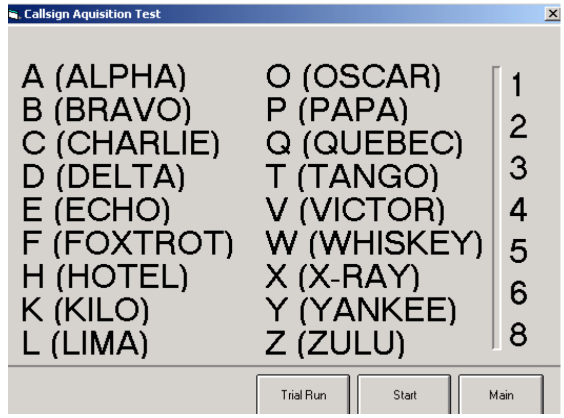


Figure 2. Screenshot of the CAT test

Figure 3. Screenshot of the MRT test

2.3 Experimental Design

A completely randomized design was used in this experiment. Since each participant went through both CAT and MRT test, a within subject design was used. One independent variable with two levels (CAT and MRT) and one dependent variable (speech intelligibility score as measured by percentage of complete callsign correctly identified by participants) were used.

2.4 Equipment

The equipment used for this study included a Desktop Computer, Sound Forge 8.0® Software, CAT software developed at the U.S. Army Research Laboratory to administer CAT items, a Desktop Monitor and Keyboard inside an acoustically treated sound chamber (Acoustic Systems 143 MC), a step attenuator (Kay Elementrics 839 Attenuator), a Radioear B-71 bone vibrator, a pair of Telephonics TDH-39 earphones, a digital force gauge, a headband, and a suite of calibration equipment (Larson Davis Precision Sound Level Meter Kit containing a(n) 824 Sound Level Meter, AMC 493 Artificial Mastoid, and Precision Acoustic Calibrator).

2.5 Procedure

On the day of the experiment, participants were briefed on the purpose of the research. Participants were asked to sign an informed consent form before proceeding to start the experiment. All participants were required to pass the hearing test. Participants were seated at the listener's station in a sound treated booth (shown in Figure 3) in front of a monitor (that displays the CAT items) with a keyboard for data input. A bone vibrator was placed at the listener's condyle and fastened in place with a headband (Figure 4). The condyle was used in this study because it appears to have the most favorable intelligibility score compared to other locations on the head [4, 12].

A digital force gauge was used to measure the force exerted by the headband to ensure that adequate, but no unnecessarily high and uncomfortable force was applied to the head of the listener. For this experiment, a static force of 3.5N to 3.9N (to ensure consistency) was used in accordance with audiology related literature, which recommends that the static force applied by bone vibrator to the human head must fall within the range 2.5N (minimum required for stable position) and 5.9N (level of discomfort) (ANSI S3.5-1997 [11]).

For the CAT-60 test, participants were asked to listen to incoming CAT items displayed on the computer monitor in front of them and identify what they hear by pressing the appropriate keys on their computer keyboard. For example, if the listener heard the callsign "Charlie Six", the correct response would be to press the "C" key, followed by the "6" key and then press the "Enter" key. For the MRT test, a random list of six MRT words was presented to the listener each time. The words were numbered one through six (1- 6), listeners were asked to identify the word they hear by entering the number that corresponds to the word. For example, if a listener hears "tear" in a word list "teak, team, teal, teach, tear, tease", numbered 1-6 with "teak" corresponding to 1 and "tease" corresponding to 6 respectively, then the listener will press key "5" followed by the "Enter" key.

Upon completion of the experiment, participants were debriefed and thanked for their participation.

Figure 1. A participant is taking the test

2.6 Data Collection

Both CAT and MRT can be measured by more than one measure. For instance, CAT can be measured using the percentage of correctly identified word, number, or total. To ensure a fair comparison, only the percentage of correctly identified complete (total) phrase from each test were collected and analyzed.

3. Results

3.1 Subjective Measures

About 82% of the participants felt comfortable using the bone vibrator. In addition, half of the participant stated that it would be comfortable using the bone vibrator for a long term.

3.2 Objective Measures

Each Participant's responses were calculated and converted to the percentage of total (word number combination). Data from the experiment can be seen in Table 2 and Figure 5.

A paired t test was used to compare the speech intelligibility of the two test materials. Results indicated there is no statistically significant difference between the two test materials (t_5=0.99, p>0.05).

Table 2. Data Collected from the Experiment

Subject #	CAT	MRT
1	20	28
2	41	30
3	50	28
4	23	30
5	53	20
6	28	36
Mean	35.8	28.7
Std. Dev.	14.1	5.2

4. Discussion

One of the important findings of this research is that there was no significant difference between the CAT test and the MRT test with the babble noise as the background. Considering that soldiers are more comfortable with the CAT test, it indicates that the CAT test is a good option to use in military applications when conducting bone conducted speech studies.

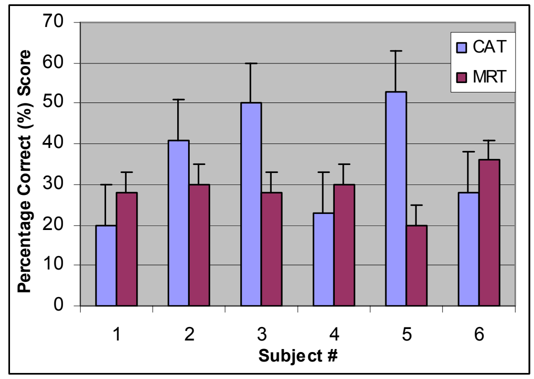


Figure 5. Graphical Display of CAT and MRT data

However, it needs to be pointed out that there are some limitations about this study. First, the sample size is too small and the participant pool is very narrow. Those factors might affect the results of the experiment, and hence, it needs to be taken into consideration. Second, only babble noise was used in this study. In the future, other background noises need to be included.

Even with the limitations of this study, it indicates that the CAT is a good alternative test material for the military applications. Results from this research will have the potential to help designers of military communication systems to improve their designs.

5. Acknowledgment

The authors would like to thank Dr. Elmar Schmeisser from the US Army Research Office for his assistance of this research. The research reported in this paper was performed in connection with Grant number W911NF-05-1-0518 with the U.S. Army Research Office. The views and conclusions contained in this document are those of the authors and should not be interpreted as presenting the official policies or position, either expressed or implied, of the U.S. Army Research Office.

6. References

[1] Rao, M. and Letowski, T. (2003), "Speech intelligibility of the callsign acquisition test (CAT) for Army communication systems." the 114th Convention of the Audio Engineering Society. Amsterdam, the Netherlands

[2] Tonndorff, J. (1972), Bone Conduction. In J. V. T. (Ed.), Foundations of Modern Auditory Theory (Pp. 195-237). Academic Press, New York.

[3] Walker, B. N, Stanley, R.M (2005), "Evaluation of bone conducted headsets for use in multitalker communication environments," Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting. September 26-30, 2005, Orlando, Florida

[4] Osafo-Yeboah, B., McBride, M., Gripper, M., and Jiang, X. (2006), "Effect of Bone Conduction Vibrator Placement on speech intelligibility using the Callsign Acquisition Test (CAT) – A pilot study," 2006 Annual Industrial Engineering Management Systems Conference, Cocoa Beach, FL.

[5] Eagan, J. P. (1948), "Articulation Testing Methods." Laryngoscope, *58*, 955-991.

[6] Blue M, Ntuen C, Letowski, T. (2004), "Speech intelligibility of the callsign acquisition test in a quiet environment." International Journal of Occupational Safety and Ergonomics, 10(2), pp. 179-189.

[7] Gripper, M. A. (2006), Evaluations of the Callsign Acquisition Test CAT in acoustic environments. Doctorial Dissertation, North Carolina A&T State University, May 2006.

[8] Osafo-Yeah, B., Jiang, X., McBride, M., Mountjoy, D., and Park. E., "Using the Callsign Acquisition Test (CAT) to investigate the impact of background noise, gender, and bone vibrator location on the intelligibility of bone-conducted speech," International Journal of Industrial Ergonomics, In Press, Corrected Proof, Available online 11 September 2008.

[9] ANSI S3.6 (1996). American National Standard Specification for Audiometers. *American National Standards of the Acoustical Society of America*. New York.

[10] ANSI S3.1 (1997). Criteria for Permissible Ambient Noise during Audiometric Testing. *American National Standards Institute*. New York.

[11] ANSI S3.5 (1997). Methods for the Calculation of the Speech Intelligibility Index. *American National Standards Institute*. New York.

[12] McBride, M.; Letowski, T. R. and Tran P. K. (2005). Bone Conduction Head Sensitivity Mapping: Bone Vibrator. Army Research Laboratory, Aberdeen Proving Ground.

The Journal of Management and Engineering Integration
Vol. 1, No. 2, December, 2008

The Effect of Visual Distraction on Visual Inspection Performance

Michael J. Eberhart and Mohammad T. Khasawneh
State University of New York at Binghamton
mkhasawn@binghamton.edu

Abstract

Distractions are a natural occurrence in the manufacturing workplace. While performing a visual inspection one must sustain focused visual attention to perform the task quickly and accurately. This study was conducted using a simulated inspection task under multiple speeds of a non-relevant visual distraction (i.e., blinking monitor background): no distraction, 3, 6, and 9 blinks/second. The accuracy and search times were collected to study the effect of distraction on human visual search performance. Increased accuracy was found with increased distraction, with the maximum reached at 6 blinks/second. Search time, however, was not affected by the increased stimulus. This study agrees with previous studies that suggest that central nervous system (CNS) arousal increases humans performance during simple tasks, a concept known as the inverted-u hypothesis.

1. Introduction

The manufacturing industry involves humans at all levels of production including assembly and inspection. The technological advances in automation that have come over the years have decreased the amount of necessary human involvement within the assembly stage. Automation in the industry has increased productivity and has decreased imperfection and, more importantly, accidents involving humans. However, there are still many aspects of manufacturing where humans outperform machines, one of which is visual inspection. Visual inspection is the process in which an inspection is performed on a product to see if it meets the manufactures' criteria. This determines if a product can go on to the next step of production or be sent to the consumer.

Many attempts have been made to integrate machines into the inspection process, but because humans have better decision making ability and the ability to see events that were not predicted [3], the use of the human visual inspection process in the future is expected to continue [10].

Much research has gone into what affects human performance while performing visual inspection and how performance can be improved. However, the most common discussions address the effects of noise. For example, Gawron [7] reviewed 58 noise experiments in which 29 showed a decrease in performance, 22 showed no effect, and 7 showed an increase in performance with exposure to noise. The wide range of performance effects is due to the range of tasks performed and measures taken [7]. It is not clear what the effect of noise is on performance when it is generalized. Additionally, many recent studies have shown that an increased level of visual distraction contributes to poorer performance and slower reaction times in complex scenarios [12]. However, because there has not been an extensive amount of research dedicated to the effects of a visual distraction during a visual inspection task, this study examines that relationship.

The issue with a distraction during visual inspection is maintaining focused attention. Focused attention is the process of maintaining attention for an extended period of time to the task at hand while ignoring other surrounding factors. Manufacturing environments are not the most ideal place for focused attention tasks, particularly because they are usually filled with loud noise and extensive movement from people, tools, and machines. Even in the most ideal conditions where inspection is done away from most of the distractions, interruptions can still occur. It has been shown that visual attention is not perfect and can still be affected when attention is focused before a distraction [4]. This study is therefore designed to investigate the effects of visual distraction, if any, on human performance during a visual inspection task. The methodology adopted in the study is discussed in the following sections.

2. Method

2.1. Subjects

Nine subjects, who worked at a Lowe's Home Improvement store, were randomly selected to participate in this study. The age of the subjects ranged from 20 to 50 with a mean age of 28. Seven subjects were male and two subjects were female. All have had experience using computers. While all of the males play video games on a regular basis, none of the females play video games.

2.2. Equipment

The experiment was conducted on a Dell laptop with an AMD 1.8 GHz processor and a 15.4" display. Each subject used a mouse to respond to the stimulus material.

2.3. Stimulus

The inspection task was simulated using a Visual Basic program, as shown in Figure 1, similar to that used in a pervious visual inspection study [6]. The distraction was a program written in Visual Basic that flashed the color red behind the visual inspection program at a selected rate (blinks/sec). Non-relevant distractions have been shown to affect performance while within ones focus of attention [9]. Research has also shown that the color red is stimulating and disagreeable and diverts attention to the outward environment [8]. The distraction rates of 3, 6, and 10 blinks per second were chosen with duration of 0.05 seconds per blink. The recommended blink rate to gain attention is between 3 and 10 seconds with duration of at least .05 seconds [11]. There have been many studies investigating the effects of intermittent light, computer displays, and fluorescent lights, as a distraction and how it affects saccades. Saccades are the fast eye movement humans make between fixations. Many studies have shown that saccade latency, the delay in time from determining a new target and the start of the actual eye movement, is delayed during this type of distraction [1,5].

2.4. Experimental Design and Procedure

A within-subject design with single factor (distraction rate) at four levels was used in this study. The study took place over a three day period, wherein 4 subjects were gathered on day 1, 3 subjects on day 2, and 2 subjects on day 3. The same procedure was performed on each day of the experiment. Each subject was first asked to fill out a prescreening application which determined if they wore corrective lenses, had color blindness, or had reflex epilepsies, among other information that could lead to further analysis or conclusions. One subject at a time was brought into a quite room with adequate lighting and was briefed on the experiment and was given a consent form to sign. Each subject was then given a chance to train with the visual inspection program with no feedback on performance.

The subject's task was to find a defect that was present in each screen that would be presented as the letter 'X'. The background density was chosen to be forty percent and a self-paced time of 16 seconds was given for each screen. All aspects of the design for the inspection task were chosen based on a previous experiment [6]. All subjects performed the simulated visual inspection under no distraction, 3 blinks/sec, 6 blinks/sec, and 10 blinks/sec. Each distraction rate was performed twice (8 trials per subject) with 25 screens. All trials were random and were broken up into two sections, each consisting of 4 trials, separated by an hour break. After both sections were completed a questionnaire was used to determine if the distraction had affected the subject and if there were any other unforeseen factors that may have been unaccounted for. Subject's compensation involved being able to perform this experiment while being on the clock at work. The store manager provided the room and the permission of payment.

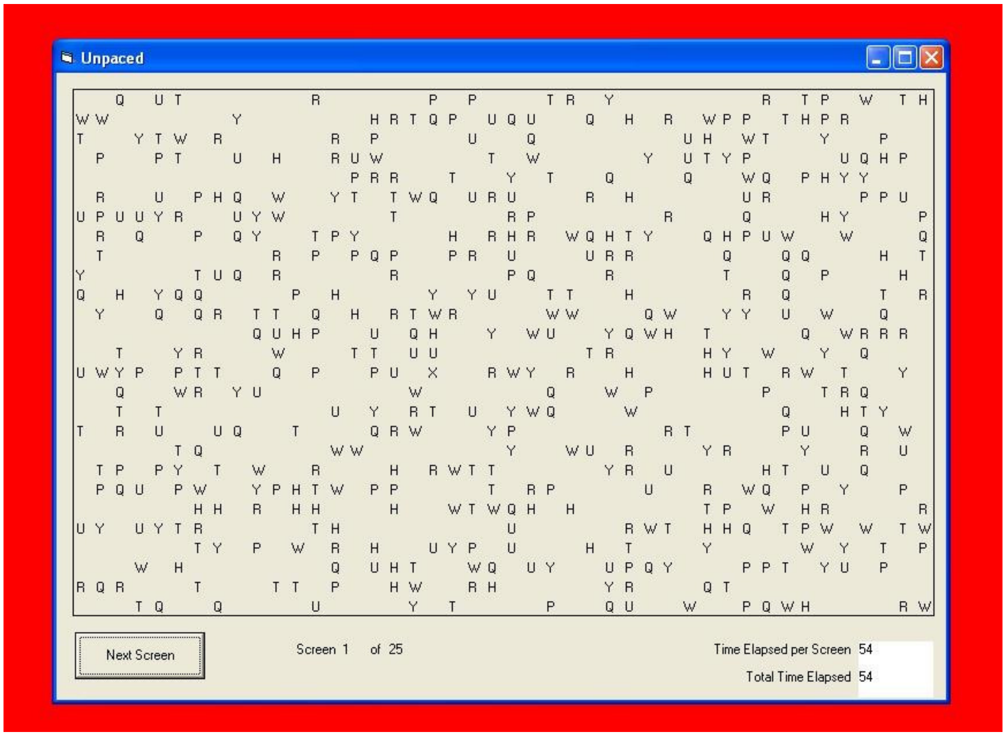


Figure 1. Screenshot of the visual inspection program.

3. Results

The study measured accuracy and speed of search time. Search time is the time taken to find the defect that existed and does not account for the time when a defect was not found. Table 1 shows the mean values and standard deviations of speed and accuracy for the different distraction rates. Analyses of variance (ANOVA) were conducted on the two performance measures.

Table 1. Means and (standard deviations) of performance measures.

Distraction Rate	None	3 blinks/sec	6 blinks/sec	10 blinks/sec
Accuracy (%)	69.11(11.96)	76.44(11.86)	83.78(16.35)	80.67(12.92)
Inspection Time (seconds)	7.26(.718)	6.58(1.35)	6.26(1.28)	6.85(1.25)

3.1 Search Time and Accuracy

The results showed that the distraction rate had no significant effect on search time ($F(3,24) = 1.8457$, $p>.05$). Moreover, the results showed that the distraction rate had a significant effect on performance ($F(3,24) = 4.251$, $p < .05$). A post hoc analysis was conducted on the accuracies using Tukey's test. The test revealed that the mean accuracy rate at a distraction rate of 6 blinks/sec was different than that measured with no distraction, as seen in Figure 2. It is important to note that subjects were not informed about the equal importance of speed and accuracy.

3.2 Questionnaire

In addition to the performance data collected during the experimental trials, subjective data was collected using pre- and post-experiment questionnaires. The responses on the task questions lead to no other conclusions or analysis. The data in Figure 3 shows how many subjects said 'yes' the blinking screen was a distraction or 'no' the blinking screen was not a distraction.

Distraction Rate	No Distraction	3 blinks/sec	6 blinks/sec	9 blinks/sec
No Distraction				
3 blinks/sec				
6 blinks/sec	[redacted]			
9 blinks/sec				

Figure 2. Comparison of mean accuracies at different distraction rates (blinks/s)

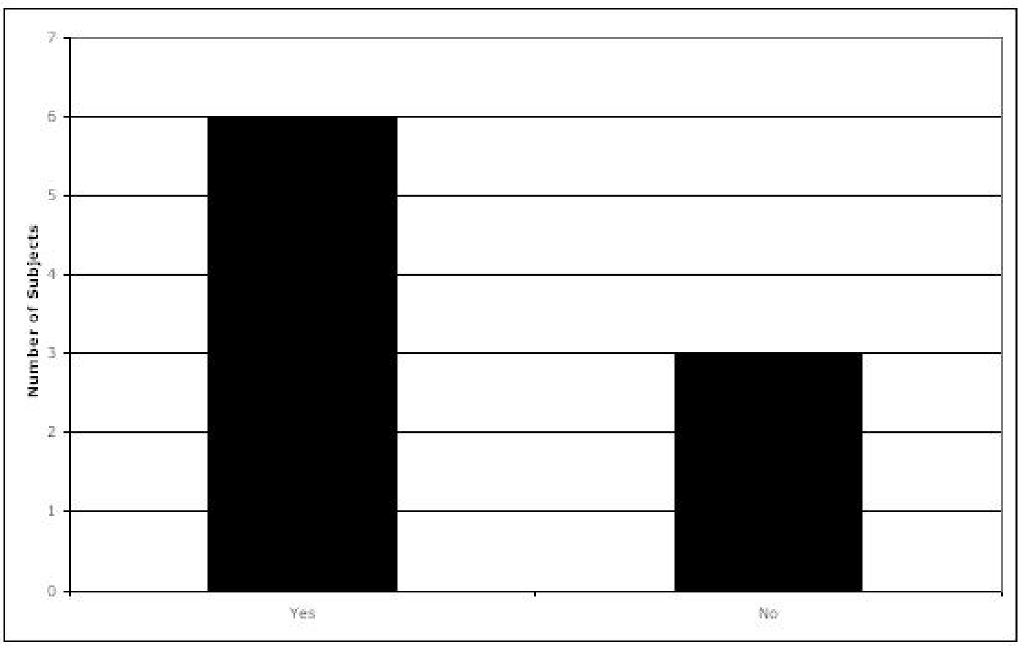


Figure 3. Subjects' response to questionnaire

4. Discussion and Conclusion

The purpose of this study was to show the effect of a non-relevant visual distraction inside the focused attention of an inspection task. The study has shown some interesting evidence that distraction during an inspection task may actually increase the accuracy of the inspector and has no affect on the inspection time. This evidence is very similar to the effects of sound on performance during a visual task.

There has been research that shows that visual detection tasks have an increase in performance while noise is used as a distraction. Britton and Delay [2], for example, explored the effects of frequency and decibel (db) level of noise and how it affects a focused visual attention task. Specifically, they showed that, under lower frequencies, an increased dB level increased the accuracy of detection, whereas, under a higher frequency, performance decayed while under an increased dB level [2]. They suggest that these results show that there is an optimal level of CNS arousal that can be achieved. At the optimal level of CNS arousal one's focused attention is at its most productive point and can be maintained for a longer period of time. When the CNS arousal is above or below this point performance degrades. This type of response to a distraction is called the inverted-u hypothesis. The data displayed from this research, as illustrated in Figure 4, shows the inspection accuracy as a function of the distraction rate. As the distraction increases the accuracy of the subjects also increases.

Briton and Delay [2] also discussed that an effect happens when one's attention is narrowed, by showing that the accuracy to detect the stimulus in the peripheral view was worse than directly in front of the subject. This theory can also be applied to the current study. The hypothesis in this research claimed the red blinking screen would focus the subject's attention outward and saccade latencies would increase, therefore decreasing accuracy and increasing inspection time. The distraction, however, could be assumed to actually narrow the subject's attention down to the screen so that all peripheral content was disregarded.

This effect could be the reason that during a certain amount of visual distraction the subjects had increased accuracy. The highest level of distraction (i.e., 10 blinks/sec) has a negative impact on the accuracy of inspection, similar to the effect of the higher frequencies and dB level of noise in Britton and Delay's [2] research. The CNS of the subject seems to overload, and performance degrades. The results show that the distraction rate had a significant effect on accuracy.

The distraction had no significant effect on search time. The mean search time remained relatively the same for each trial. When there was a distraction there was a slight decrease in the search time as seen in Figure 5. This slight decrease in the search time could be an effect of the blink rate or color. The increased distraction rate could make the subject think that more time has elapse than actually did or the color of the screen could irritate the subject enough to make them perform a faster inspection. However, further research would be needed to make a valid assumption.

Future research should focus on the: integrated effect of the type, duration, and location of visual distraction. These three factors may be analyzed to determine if the response can be repeated. Moreover, the transfer effects into realistic industrial environments should be explored.

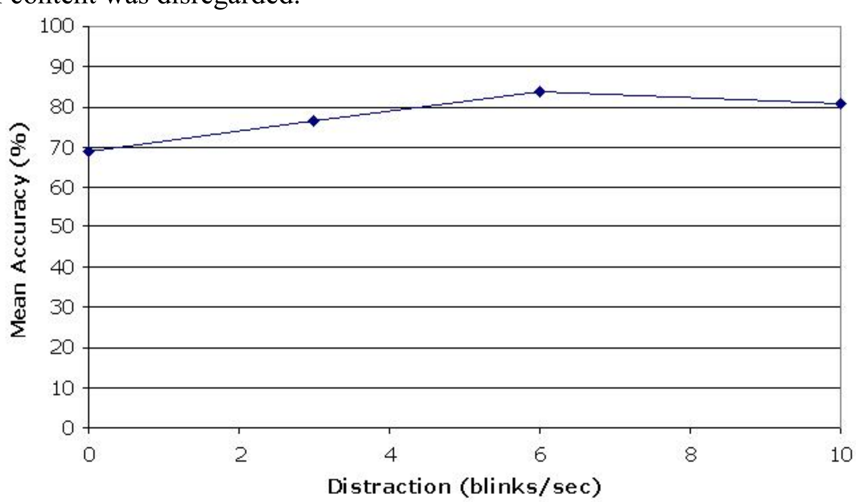


Figure 4. Mean accuracy as a function of distraction rate

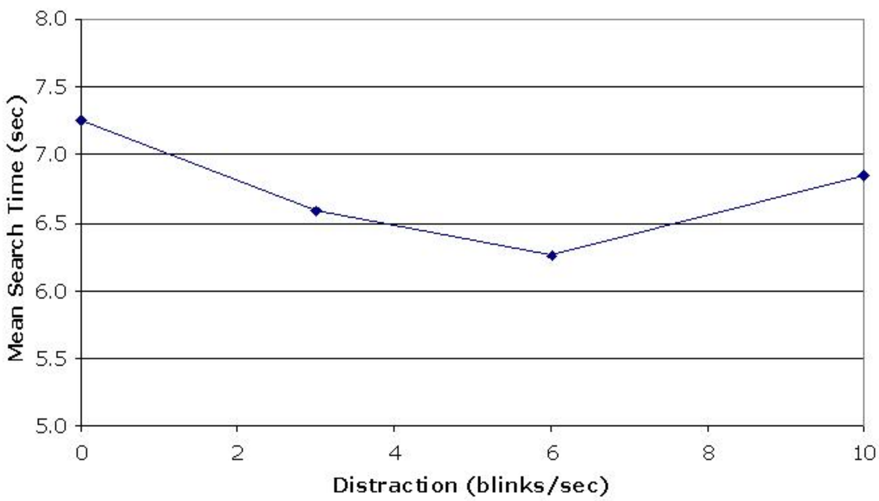


Figure 5. Mean search time as a function of distraction rate

5. References

1. Baccino T, Wolfgang J, and Bussolon J. (2001). The influence of bright background flicker during different saccade periods on saccadic performance. Vision Research, 41: 3909-3916.

2. Britton, L.A., Delay, E.R. (1989). Effects of noise on a simple visual attentional task. Perceptual and Motor Skill, 68(1), 875-878.

3. Drury, C.G., Sinclair, M.A., (1983). Human and machine performance in an inspection task. Human Factors, 25(4), 391-400.

4. Eriksen, B.A., Eriksen, C.W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. Perception & Psychophysics, 16, 143-149.

5. Findlay, J. M. and Walker, R. 1999, A model of saccade generation based on parallel. processing and competitive inhibition, Behavioral and Brain Sciences, 22, 661 – 721.

6. Garrett, S.K., Melloy, B.J., Gramopadhye, A.K. (2000). The effects of per-lot and per-item pacing on inspection performance. International Journal of Industrial Ergonomics, 27(5), 291-302.

7. Gawron, V. (1982). Performance effects of noise intensity, psychological set, and task type complexity. Human Factors, 24(2), 225-243.

8. Goldstein, K. (1942). Some experimental observations concerning the influence of colors on the function of the organism. Occupational Therapy and Rehabilitation, 21, 147-151.

9. Gronau, N., Sequerra, E., Cohen, A., Ben-Shakhar, G. (2006). The effect of novel distractors on performance in focused attention tasks: A cognitive-psychophysiological approach. Psychonomic Bulletin & Review, 13(6), 570-575.

10. Thapa, V.B., Gramopadhye, A.K., Melloy, B., Grimes, L., (1996). Evaluation of different training strategies to improve decision-making performance in inspection. International Journal of Human Factors in Manufacturing, 6(3), 243-261.

11. Woodson, W.E., Conover, D.W. (1964). Human engineering guide for equipment designers (2nd ed.). Berkley: University of California Press.

12. Zhang, H., Smith, M.R.H., Witt, G.J. (2006). Identification of real-time diagnostic measures of visual distraction with an automatic eye-tracking system. Human Factors, 48(4), 805-821.

Preventing WMSDs Among Medical Sonographers Using Digital Human Modeling: An Evaluation Study

Kumar Sudarsanam, Bhawana Bhopte, Mohammad T. Khasawneh
State University of New York at Binghamton
mkhasawn@binghamton.edu

Abstract

The purpose of this paper is to study the efficacy of the National Institute for Safety and Health's (NIOSH) recommended control measures to prevent Work related Musculoskeletal Disorders (WMSDs) among medical sonographers using digital human modeling. Medical sonographers experience significant pain and discomfort from performing trans-abdominal and trans-vaginal exams in their neck, shoulder, and back. To address these issues, NIOSH recommends a number of control measures to reduce the incidence of risk factors that facilitate injury, including the use of height adjustable scanning beds to enable the sonographers to position the patient at a reach distance that ensures neutrality in their muscle exertions. To test the efficacy of such measures, a practical methodology should be adopted and applied before health care facilities invest in such costly workplace redesign. Therefore, this study uses "Jack", a digital human modeling software, to evaluate the efficacy of using height adjustable beds in ultrasound centers. The ultrasound procedure was modeled with and without adjustable beds and the results were analyzed using available ergonomics assessment tools. The results indicated that using height adjustable scanning beds resulted in lesser low back compression force, better scanning posture, and much lesser shoulder abduction angle, thereby conclusively proving the efficacy of NIOSH's recommendations.

1. Introduction

Ergonomics evaluations in the modern workplace have become more and more common as necessitated by the exponential growth in WMSDs. WMSDs are the most important cause for the rise in worker's compensation claims, thereby causing immense losses for businesses and governments. Governments face problems in increasing litigation costs against businesses which further necessitates broad based interventions. While specific ergonomics interventions and safety measures undertaken have reduced the incidence of WMSDs, the need for best practices in this area is imperative.

Among the sectors that has grown prominently in the past two decades is health care. NIOSH's initiatives have especially increased in the health care sector whose rates of occupational injury have risen over the past decade. Two highly unsafe industries, construction and manufacturing, are less hazardous than they were historically. The national rates for non-fatal injuries and illnesses resulting in lost-workdays were 8.1/100 for full time hospital workers (8.8 in 2001) [1], and 9.1/100 workers in nursing and personal care facilities (13.5 in 2001). On the other hand, construction workers had an injury rate of 6.3/100 (7.1 in 2001), and manufacturing workers 6.3/100 (8.1 in 2001) [1].

The purpose of this research is to evaluate and quantify the efficacy of a number of ergonomic interventions prescribed by NIOSH [12] in preventing sonographers' WMSDs in a health care facility, by employing computer-aided ergonomics and digital human models. To analyze the efficacy of a prescribed ergonomic intervention, a typical ultrasound center, with the workstation, personnel, and the task, is modeled and simulated.

Sonography is commonly used in most medical specialties [10]. The Society of Diagnostic Medical Sonography's (SDMS) study [13] indicated that 80% of sonographers are scanning in pain, with 20% of them eventually experiencing a career-ending injury. Moreover, the average time a sonographer is in the profession before experiencing pain during scanning is 5 years [13]. Such injuries can be very costly for hospitals and imaging centers, considering not only the lost workdays, but, in many cases, the permanent loss of skilled talent. Organizations like the Occupational Safety and Health Administration (OSHA), SDMS, and NIOSH have aided the development of industry standards and best practices for ergonomic design of equipment, workplaces and refinement of workplace tasks to prevent such injuries. In its most recent workplace solutions publications, NIOSH published the results of an investigation at a hospital's antenatal unit [12], which reinforced previous findings by SDMS and OSHA, where most sonographers reported neck, arm and shoulder injuries while performing ultrasound tasks.

To control the associated risk factors, NIOSH recommended appropriate work practices and engineering controls which reflected the industry standards set by SDMS in its consensus report released in 2003. The salient among those practices include the following [12]:

- Provide adequate work space for personnel, sonography equipment, the patient bed, and other equipment.
- Use adjustable tables (including those equipped with drop-down side rails).
- Minimize awkward and extreme postures.
- Schedule different types of exams for each sonographer in a workday to decrease strain on musculoskeletal tissues.
- Provide periodic training and reassessment regarding the above ergonomic interventions as well as incorporate new best practices.

One limitation of such a study is that the demographics of the test subjects may not be representative of the facility trying to adopt those practices or the population of medical sonographers, thereby emphasizing the importance of testing and quantifying the impact of the proposed measures on a representative percentile of the working population, which motivated this research.

3. Literature Review

Workplace task evaluation and human performance assessment has been carried out using virtual humans in a variety of contexts. For example, Pitarch et al. [16] illustrated the use of virtual humans in evaluating the ergonomics of operators' usage of joysticks in a material handling task in a manufacturing setup. An ergonomic analysis of joystick design and its relation to the torque exerted on the fingers was performed and appropriate design criteria for minimum risk of injury to fingers were predicted. Kumar [9] published a comprehensive ergonomics evaluation of work practices, tasks and design of tools in the cleaning occupation, which included several studies to evaluate physiological, subjective, and postural loads while employing traditional and redesigned workstations.

Furthermore, Okumoto [14] published a study detailing the employment of Jack for simulation of human tasks in shipbuilding industry in Japan. Static strength prediction analyses and energy expenditure estimations were conducted for several physical tasks, such as welding, grinding and lifting. Based on these analyses, preferable work methods were proposed. Du [6] followed a computer-based ergonomic analysis using Jack to evaluate a lifting task while being conducted both manually and using an assist-device at a workstation in an automobile assembly plant. Four analysis tools, namely static strength analysis, rapid upper limb assessment, metabolic energy expenditure analysis, and NIOSH lifting equation, were used to evaluate the potential risk of injury. The results revealed that there were relative stresses on the trunk and arm areas when the task was performed manually and that using the assist-device decreased injury risk potentially and adversely impacted assembly productivity.

Kim et al. [8] conducted a lifting posture analysis using virtual human modeling using an optimization-based motion prediction method. The results showed that different amount of external loads and tasks lead to different human postures and joint torque distribution, thus causing different risk levels of injury. Chaffin [4] briefly describes a method using computerized simulations of human physical exertions to assist in reducing the risk of low back injuries during the early phases of designing manual material handling tasks in industry. The method is applied to a variety of common conditions to document how certain workplace design changes can reduce low back stresses, and allow tasks to be performed by additional workers. Oxley et al. [15] investigated and identified risk factors for WMSDs in box recycling collection schemes and provided recommendations for reducing the risks of WMSDs to waste/recycling collectors using digital human and workstation modeling in the Jack software. Similar ergonomics evaluation and assessment studies were performed by Chaffin [2,3] and Chang and Wang [5].

3. Problem Statement

As can be seen in Figure 1, a sonographer's workstation consists of the ultrasound system, the monitor, the control panel, the scanning bed/table, and the chair.

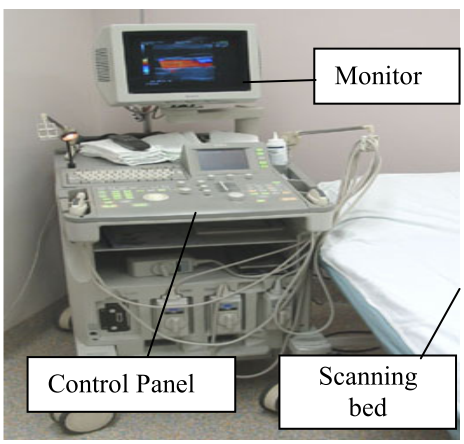


Figure 1. An ultrasound workstation [13].

The basic task of a sonographers generally involves maneuvering the ultrasound system and the patient's scanning bed to bring them side by side, and then placing the transducer near the area of the body to be imaged. Viewing the screen during the scan, the sonographer looks for subtle visual cues that contrast healthy areas from unhealthy ones, for example. He/she then decides whether the images are satisfactory for diagnostic purposes and selects the ones to be shown to the physician. To achieve minimum shoulder/arm abduction, NIOSH recommends that adjustable tables be provided to optimize the positioning of the patient and the sonographer. Therefore, the aim of this study is to conduct an ergonomic evaluation of the ultrasound scanning procedure pre- and post-intervention, with (Scenario I) and without (Scenario II) height adjustable equipment, to validate and, most importantly, quantify, the efficacy of NIOSH's recommendations.

4. Methodology

As stated earlier, the present study addresses the following hypotheses: 1) there is no significant difference associated with potential ergonomic injury risk between the scanning task when the sonographer operates with adjustable scanning table/bed; and 2) there is a significant decrease in injury risk when the sonographer operates with adjustable scanning table/bed. A typical work schedule comprises of five days per week, with seven scanning hours per day [7]. During a typical workday, each sonographer performs about 10 ultrasounds, most of which are trans-abdominal, and, hence, the model presented here is restricted to a trans-abdominal examination task. To maintain appropriate fidelity with the actual working posture of an ultrasound operator, the postures shown in Figure 2 were used in this research.

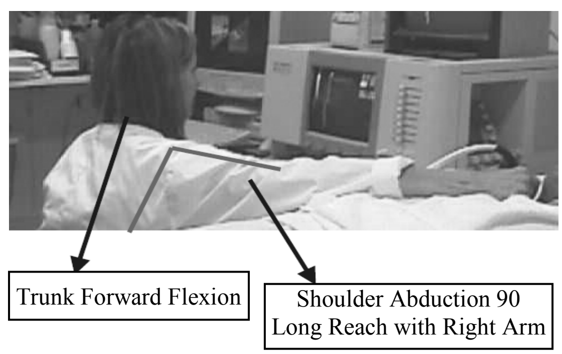


Figure 2. Illustration of shoulder abduction and trunk flexion [11].

It is important to note that the ultrasound worker adjusts his or her scanning posture to reach the entire abdominal area while requiring minimum movement from the patient. Sonographers pay maximum attention to the comfort of the patient. In NIOSH's study, one of the scanning rooms had improvised a stretcher into a scanning bed, thereby utilizing its limited height-adjusting capabilities. The working posture of a sonographer operating that equipment can be seen in Figure 3. Although this arrangement reduced the abduction angle, the limited adjustability of the bed reduced the benefits that could be obtained. However, this posture served as reference in the second scenario modeled in this study.

Figure 3. Example of reduced shoulder abduction using stretcher of limited adjustability [11].

Digital human modeling software, such as Jack, provides the advantage of employing databases of anthropometrically representative population of human models. Therefore, for modeling purposes, three different ultrasound workers, each corresponding to the 5th, 50th, and 95th percentiles, were used in the analysis. The anthropometric database that Jack uses is based on the ANSUR database, and, hence, is assumed to approximate well the population of medical sonographers in the US. Those models were therefore used to accurately model the scanning task for each of the four experimental replicates of the animation/simulation. The demographics of the experimental human models are shown in Table 1.

Table 1. Demographics of test models.

Percentile	Height (cm)	Weight (lbs)	Upper Arm (cm)	Lower Arm (cm)
5	164.69	61.590	35.5	44.6
50	175.49	75.69	38.0	47.4
95	186.65	98.070	40.7	50.4

Three traditional ergonomics assessment tools in Jack were used in this study. First, *Low Back Analysis* was used to calculate the spinal forces acting on the back. Second, *Ovako Working Posture Analysis* was used to determine the comfort of the working postures and the urgency of taking corrective measures. Third, the *Static Strength Prediction* tool was primarily used to observe the shoulder abduction angles. The simulation was modeled using Jack's animation tool to represent the scanning task in terms of motion per second. A 31.60 second simulation model was developed using the two scenarios for which all three analysis tools were used.

5. Results

5.1. Low Back Analysis

The results of this analysis for all three anthropometric profiles of humans are shown in Table 2 for both males and females. The results indicated that the low back compression forces for both scenarios are significantly different. This suggests (and further justifies) that the height adjustability of the bed is an important factor in the design of new (ergonomically-designed) ultrasound workstations and workplaces. This is intuitively evident since lesser torso flexion results due to the operator being able to maneuver the patient to a height of his/her comfort. Moreover, the reach required for accomplishing the scanning motion is lesser.

Table 2. Digital human model results.

Gender/ Percentile		Low Back Compression Force			
		Scenario I		Scenario II	
		Mean	SD	Mean	SD
Male	5	1256.80	24.50	756.80	25.96
	50	1435.00	17.85	872.00	18.49
	95	1543.60	13.58	1228.00	20.76
Female	5	1475.75	18.66	877.50	19.89
	50	1607.00	24.67	916.00	23.99
	95	1678.50	26.44	1028.00	29.99

Although the injury risk is below NIOSH's back compression limit of 3400 N, the frequency of the scanning task and the sustained pressure incurred is an important factor contributing to low back injuries. In fact, a NIOSH survey [11] revealed that 58% of sonographers indicated that low back was a major anatomical site of discomfort for them, which was third only to shoulder (59%) and wrist (76%). Hence, these results not only show that the measures recommended by NIOSH to reduce the risk of injury to sonographers are indeed effective in reducing the risk of low back injury, but also quantify the amount of improvement, which further justifies the advantage of digital human modeling.

5.2. Ovako Working Posture Analysis Results

Figures 4 and 5 illustrate the results of the working posture analysis for the two scenarios. It is evident that the posture of the sonographer when using a height adjustable bed seems normal and natural and the postural load on the musculoskeletal system is acceptable, thereby implying no need for further corrective measures. These results indicate that the scanning posture in the original position has harmful effects on the musculoskeletal system, which further proves the efficacy of the intervention.

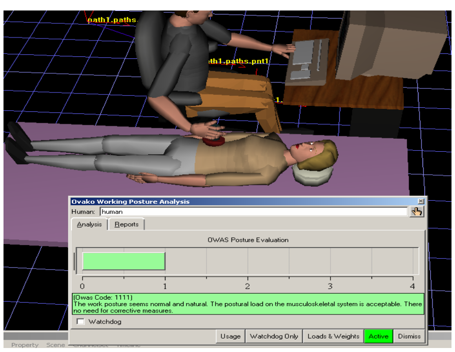


Figure 4. OVAKO results for Scenario I.

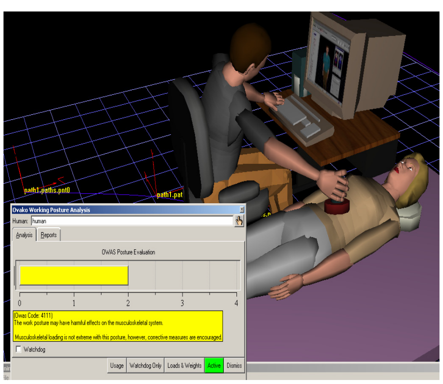


Figure 5. OVAKO results for Scenario II.

5.3. Shoulder Abduction Angle Results

Figures 6 and 7, obtained using the *Static Strength Prediction* tool, confirm the reduction in shoulder abduction angles as a result of the intervention, particularly because of having lesser reach to perform the scanning task as a result of being able to position the patient at a comfortable height. The results clearly show that the shoulder abduction angle in scenario one is clearly reduced from 90 degrees to 40 degrees in scenario two, which is consistent with NIOSH's findings.

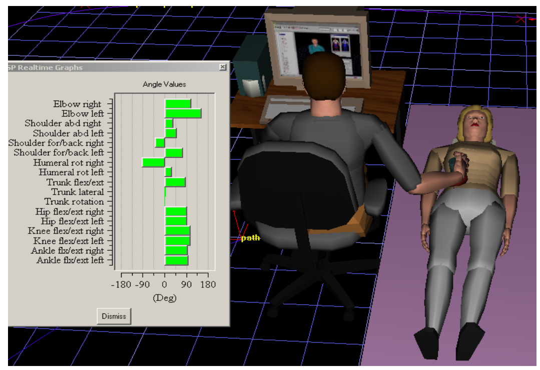


Figure 6. Shoulder abduction angles for Scenario I.

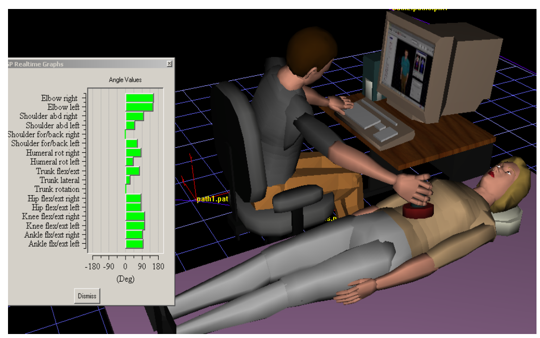


Figure 7. Shoulder abduction angles for Scenario II.

6. Conclusions and Final Remarks

Apart from the above results, the following observations were made. First, high torque at both the ankles was observed in both scenarios, which is consistent with NIOSH's findings at the antenatal testing site [11]. This can be alleviated with the provision of proper adjustable footrests. Second, the high elbow angles indicate the need for support in both scenarios. However, there are different opinions in the literature on what kind of support is necessary [7]. The support can be as simple as a wedge or a pillow placed between the patient and the sonographer or a saddle-type cushion placed over the patient and resting on the edges of the bed, for example. From the above results and discussion, one can conclude that the initial hypothesis presented in this evaluation study is well supported. There is a significant difference in injury risk associated with the scanning task when medical sonographers operate with height-adjustable patient beds.

This is primarily due to a much reduced shoulder abduction angle that dramatically reduces the muscle activity in the shoulder and the injury risk correspondingly. This study also proves the ergonomic efficacy of a NIOSH recommended control measure in reducing stress-induced injuries. Implementing this NIOSH recommended measure can result in a significant alleviation in injury risk and general discomfort for sonographers, thereby emphasizing the importance of having an adjustable scanning bed as a constituent of any ultrasound workstation.

Since the current study is limited to trans-abdominal examination task, future work in this area will involve more comprehensive ergonomics assessment for a wider range of medical sonography procedures, such as trans-vaginal exams, etc. Moreover, further research that evaluates the impact of other workplace design factors (and ergonomic interventions) on the risk of injury for extended time-periods (i.e., an 8-hour shift) is required.

7. References

[1] Bureau of Labor Statistics, Available at http://www.bls.gov/, Accessed December, 2008.

[2] Chaffin, D.B., "Digital Human Modeling for Vehicle and Workplace Design", Society of Automotive Engineers, Warrendale, PA, 2001.

[3] Chaffin, D.B., "On Simulating Human Reach Motions for Ergonomics Analyses", Human Factors and Ergonomics in Manufacturing, 12(3), pp. 235-247, 2002.

[4] Chaffin, D.B., "Primary Prevention of Low Back Pain Through the Application of Biomechanics in Manual Materials Handling Tasks", G Ital Med Lav Erg, 27(1), pp. 40-50, 2005.

[5] Chang, S.W., Wang, M.J., "Digital Workplace Evaluation Using Automobile Assembly as an Example", Proceeding of Human and Organizational Issues, The Digital Enterprise, Ireland, 2004.

[6] Du, J., "The Development of a Methodology for Assessing Industrial Workstations Using Computer-Aided Ergonomics and Digital Human Models", Thesis, Department of Industrial Engineering, Mississippi State University, 2005.

[7] Habes, D.J., Baron, S. "Ergonomic Evaluation of Antenatal Ultrasound Testing Procedures", Applied Occupational Health Environment, 15(7), pp. 521-528, 2000.

[8] Kim, J.H., Malek, K.A., Yang, J., Marler, T., Nebel, K., "Lifting Posture Analysis In Material Handling Using Virtual Humans", ASME International Mechanical Engineering Congress and Exposition, Orlando, FL, November, 2005.

[9] Kumar, R., "Ergonomic Evaluation and Design of Tools in Cleaning Occupation", Thesis, Division of Industrial Design, Lulea University of Technology, Sweden, 2006.

[10] Lacey, A., "Diagnostic Medical Sonographers: Seeing With Sound", Occupational Outlook Quarterly, Summer, 2001.

[11] Murphy C., Russo A., "An Update of Ergonomic Issues in Sonography", Healthcare Benefit Trust Report, July, 2000.

[12] NIOSH, "Preventing Work Related Musculoskeletal Disorders in Sonography", NIOSH Publication No. 2006-148, 2006.

[13] Consensus Report, Society of Diagnostic Medical Sonography, "Industry Standards for the Prevention of Work-Related Musculoskeletal Disorders in Sonography", Plano, TX, May, 2003.

[14] Okumoto, Y., "Simulation of Shipbuilding Works Using Virtual Human Model", The 9th Symposium on Practical Design of Ships and Other Floating Structures, Luebeck-Travemuende, Germany, 2004.

[15] Oxley, L., Pinder, A., Cope, M.T., "Manual Handling in Kerbside Collection and Sorting of Recyclables", Health and Safety Laboratory, Report No. HSL/2006/25, 2006.

[16] Pitarch, E.P., Yang, J., Malek, K.A., Kim, J., Marler, T., "Joystick Ergonomic Study In Material Handling Using Virtual Humans", Proceedings of ASME International Mechanical Engineering Congress and Exposition, Orlando, FL, November, 2005.

Statistical Methods for Improving Product Quality in a Process Industry

Angela P. Ansuj
Federal University of Santa Maria
Santa Maria (RS), Brazil
angelaansuj@yahoo.com

R. Radharamanan
Mercer University
Macon, GA 31207
radharaman_r@mercer.edu

Abstract

In this paper, factorial design experiments and Taguchi methods were used to monitor the quality of the incoming raw material, product quality during processing, and the final product quality of a process industry. The data collected and results obtained are presented and discussed. The analysis made on the experimental results provided information to improve the quality of the product in a process industry in all three phases with cost effectiveness.

1. Introduction

The methods developed by Taguchi to optimize the process of engineering experimentation emphasize the application of engineering knowledge rather than advanced statistical techniques [1, 2]. The greatest difference between the Taguchi methods and the classical methods is that in Taguchi's experiments orthogonal arrays are used to assure the reproduction of the effects of the parameters. Another difference is that in a Taguchi study various types of signal-to-noise (S/N) ratios are used to measure variability around the target performance [3]. Taguchi also proposed the robust design or parameter design for quality improvement. It is a cost-effective approach for reducing variation in products and processes. Efforts have been directed towards understanding Taguchi's ideas and developing more statistically efficient alternatives [4-7]. Several authors revealed weaknesses in two aspects of Taguchi's approach, one being the experimental format and the other his data analysis strategy [8, 9]. To design products with robustness, Taguchi applies experimental design methods and has proposed the signal-to-noise ratios that are summary measures over the noise factor settings that analyze the experiments with both sample mean and standard deviation. The use of signal-to-noise ratios is equivalent to an analysis of the logarithm of the standard deviation of the data [5]. Nair and Pregibon [7] proposed a structured data analytic approach with three phases of analysis: an exploratory phase, a modeling phase, and an optimization phase. Leon et al [6] pointed out the lack of connection between maximizing signal-to-noise ratios and choosing the settings of design parameters. Shoemaker et al [9] proposed a model that includes both effects of control factors and noise factors. In their approach, the response is modeled rather than expected loss.

In this paper factorial design experiments and Taguchi methods were used in controlling the quality of milk and the production process of small and medium size cooperative organizations located in the State of Rio Grande do Sul in Brazil. Based on the study results, recommendations were made for the effective use of Taguchi methods for inspection and control of the quality of milk and dairy products with cost effectiveness.

2. Methodology

This work was carried out in a cooperative milk processing organization located within the Federal University of Santa Maria, Brazil. The milk production process consists of:

- receiving milk from small and medium producers and verifying the incoming product quality;
- controlling the process of pasteurization and simultaneously verifying the product quality during pasteurization; and
- controlling the final product quality after pasteurization and before distribution to the customers [10].

The selection of factors and levels for controlling the quality of milk received from small and medium producers, during pasteurization, and before distribution was determined through brainstorming and group discussion between students, faculty, and representatives from small and medium size cooperative organizations. These factors were regarded as control factors (factors that can be freely specified) and had effects on the quality of milk.

Milk from vendors: Four factors (water content - A, fat content - B, acidity level - C, and density - D) each with two levels were considered with a total of 8 degrees of freedom - one for the overall mean, 4 for the four 2-level factors, and 3 for the 3 interactions between factors (A with B, C, and D). The low and high levels used were: A (0% and 3%), B (3% and 6%), C (15^o and 19^o Donic, and D (1.031 gm/cm^3 and 1.035 gm/cm^3). Eight experiments were carried out to estimate the effect of each factor. An orthogonal array of L_8 (2^4) was used with three replications for each experiment to minimize the cost related to experimental design and analysis.

During pasteurization: Three factors (temperature during pasteurization - E, phosphatase - F, and peroxidase - G) and two levels were considered with a total of 4 degrees of freedom – one for the overall mean and 3 for the 3 two-level factors. The low and high levels used were: E (72° C and 75° C), F (either negative - accept or positive - reject), and G (either positive - accept or negative - reject).

For the process of pasteurization, factorial design experiments, (2^3) were carried out to compare the results with that of Taguchi's experiments. The seven combinations between the three factors were: E, F, G, EF, EG, FG, and EFG. Seven columns and eight rows were needed for the factorial design experiments. The results were analyzed using the Analysis of Variance (ANOVA) table.

Final product: In addition to fat (B), acidity (C), and density (D), the following parameters were also controlled: total dry extract, H (11.5% and 12.5%), non-fat dry extract, I (8.5% and 9.0%), the weight of each milk package, J (1,032 grams and 1,033 grams), and the storage temperature, K (4° C and 5° C). A total of 7 parameters with 8 degrees of freedom were analyzed using 8 experiments, L_8 (2^7).

All measurements were made in the cooperative organization's laboratory located within the Federal University of Santa Maria.

3. Results and discussions

The results are presented and discussed under: quality of milk received from vendors; quality of milk during pasteurization; and quality of final product before distribution to customers.

3.1. Quality of milk received from vendors

The factors and the levels used to control the quality of milk received from vendors are listed in Table 1.

Table 1. Factors and levels for milk received from vendors

Factor	Variable	Level	
		Low	High
A	Water (%)	0	3
B	Fat (%)	3	6
C	Acidity (°D)	15	19
D	Density (gm/cm^3)	1.031	1.035

Table 2 shows the experimental design for the milk received from the vendors. The columns under factor levels list process control factors (A, B, C, and D), the

Table 2. Experimental design for milk received from vendors

Expt.	Factor Levels							Response Values (%)			Mean (%)	S/N Ratio (dB)
	A	B	AB	C	AC	AD	D					
	1	2	3	4	5	6	7	1	2	3		
1	1	1	1	1	1	1	1	7.5	8.3	7.9	7.9	-17.96
2	1	1	1	2	2	2	2	12.6	12.4	13.1	12.7	-22.10
3	1	2	2	1	1	2	2	11.8	12.2	12.0	12.0	-21.58
4	1	2	2	2	2	1	1	14.2	13.7	13.8	13.9	-22.86
5	2	1	2	1	2	1	2	10.2	9.7	9.5	9.8	-19.83
6	2	1	2	2	1	2	1	15.1	15.9	15.5	15.5	-23.81
7	2	2	1	1	2	2	1	10.8	10.6	11.3	10.9	-20.75
8	2	2	1	2	1	1	2	16.5	16.6	16.1	16.4	-24.30

interactions that were significant between factors (AB, AC, and AD), and the levels of variables 1 or 2 for percentage of water, percentage of fat, acidity in degree Dornic, and density in gm/cm^3 selected for each experiment. The rows represent different setting combination of each factor. The response values columns show replications of percentage rejection of milk received from vendors for each factor level combination. The last two columns represent the mean response and the signal-to-noise ratio in decibels (dB) for the situation *smaller-the-better,* $(S/N)_s$ for the three replications of each experiment.

The sample mean was calculated from the response values using the equation:

$$\overline{Y} = \sum_{i=1}^{n} Y_i / n \qquad (1)$$

where: Y_i is the sample value and n is the number of replications. The signal-to-noise ratio for the situation *smaller-the-better* $(S/N)_s$ was calculated using the following equation:

$$(S/N)_s = -10\log_{10} \sum_{i=1}^{n} (Y_i^2 / n) \qquad (2)$$

The response $\overline{Y}$ for the main factors (A, B, C, and D) and for the interactions between the factors that were considered significant (AB, AC, and AD) are presented in Table 3, for the levels 1 and 2, for each one of the factors, and for the combinations of the factors. The results are shown together with the absolute difference, Δ between level 1 and level 2.

The effects of each factor and its level were obtained as follows:

$A_1 = (7.9 + 12.7 + 12.0 + 13.9)/4 = 11.625;$
$A_2 = (9.8 + 15.5 + 10.9 + 16.4)/4 = 13.150.$

The factors from strong to weak for the situation *smaller-the-better* were: C_1, B_1, A_1, AC_2, AB_1, AD_1, and D_1. These factors correspond to acidity 15^{o}D, fat 3%, water 0%, and density 1.031 gm/cm^3 respectively. The overall average, μ = 12.3875. The expected value, (EV) for the factor levels, considering the first three strong factors was calculated using:

$$EV = \mu + [(C_1 - \mu) + (B_1 - \mu) + (A_1 - \mu)] \quad (3)$$

The optimal expected value for the combination C_1, B_1, and A_1 was found to be 8.475%. This value is the minimum percentage of rejection that can be expected when receiving the milk from the vendors, considering the mean with this combination.

The $(S/N)_s$ response for the main factors (A, B, C, and D) and for the interactions between the factors that were significant (AB, AC, and AD) are presented in Table 4 for the levels 1 and 2, for each one of the factors and their combinations. The results are presented together with the absolute difference, Δ between levels 1 and 2.

The factors from strong to weak for the situation *smaller-the-better* were: C_1, B_1, A_1, AD_1, AB_1, D_1, and AC_2. These factors correspond to acidity 15^{o}D, fat 3%, water 0%

Table 3. Response $\overline{Y}$ for milk received from vendors

Level	A	B	AB	C	AC	AD	D
1	11.625	11.475	11.975	10.150	12.950	12.000	12.050
2	13.150	13.300	12.800	14.625	11.825	12.775	12.725
Δ	1.525	1.825	0.825	4.475	1.125	0.775	0.675

Table 4. (S/N)s Response for milk received from vendors

Level	A	B	AB	C	AC	AD	D
1	-21.1250	-20.9250	-21.2775	-20.0300	-21.9125	-21.2375	-21.3450
2	-22.1725	-22.3725	-22.0200	-23.2675	-21.3850	-22.0600	-21.9525
Δ	1.0475	1.4475	0.7425	3.2375	0.5275	0.8225	0.6075

and density 1.031 gm/cm^3 respectively. The overall average, μ_s = –21.64875 dB. The expected value for the factor levels, considering the first three strong factors was calculated using:

$$EV=\mu_s+[(C_1-\mu_s)+(B_1-\mu_s)+(A_1-\mu_s)] \quad (4)$$

The optimal expected value of signal-to-noise ratio for the situation *smaller-the-better* for the combination C_1, B_1, and A_1 was –18.7825 dB. This combination and the signal-to-noise ratio value correspond to the minimum percentage of rejection that can be expected when receiving the milk from the vendors, when the analysis was done using signal-to-noise ratio.

3.2 Quality of milk during pasteurization

Table 5 shows the factors and levels considered during pasteurization. Table 6 shows the experimental design used during the process of pasteurization, L_4 (2^3). Table 7 presents the average values of all the experiments. These values correspond to the percentage rejection of milk during pasteurization.

Table 5. Factors and levels during pasteurization

Factor	Variable	Level	
		Low	High
E	Temperature	72°C (min)	75°C (max)
F	Phosphatase	Accept (-)	Reject (+
G	Peroxidase	Accpet (+)	Reject (-)

Table 7. Averaging of experimental values

Level	E	F	G	Mean
Level 1	10.95	11.75	10.15	10.95
Level 2	12.70	11.90	13.50	12.70
Mean	11.825	11.825	11.825	11.825

The chosen factor levels based on the minimum percentage rejection were E_1, F_1, and G1. These correspond to setting temperature during pasteurization at 72^{o}C, phosphatase at –ve, and peroxidase at +ve. These levels are in the orthogonal array L_4 (2^3). The expected value for the chosen factor levels was calculated using:

$$EV = \mu + [(E_1 - \mu) + (F_1 - \mu) + (G_1 - \mu)] \quad (5)$$

The expected optimum value was found to be 9.2%. It is same as the mean response on line 1 of Table 6.

Table 6. Experimental design during pasteurization

Expt. No.	Factor Levels			Factor Values			Response Values			Mean Response (%)	Standard Deviation
	E	F	G	E	F	G	1	2	3		
1	1	1	1	72	A	A	9.2	9.5	8.9	9.2	0.3000
2	1	2	2	72	R	R	12.3	12.8	13.0	12.7	0.3606
3	2	1	2	75	A	R	14.5	13.8	14.6	14.3	0.4359
4	2	2	1	75	R	A	11.4	11.1	10.8	11.1	0.3000

A = Accept; R = Reject

Table 8. Factorial design experiments during pasteurization

Expts.	Factor Values							Response Value			Mean (%)
	E	F	G	EF	EG	FG	EFG				
	1	2	3	4	5	6	7	1	2	3	
1	1	1	1	2	2	2	1	9.2	9.5	8.9	9.2
2	2	1	1	1	1	2	2	10.5	10.9	10.4	10.6
3	1	2	1	1	2	1	2	10.6	9.9	10.1	10.2
4	2	2	1	2	1	1	1	11.4	11.1	10.8	11.1
5	1	1	2	2	1	1	2	9.7	10.2	9.5	9.8
6	2	1	2	1	2	1	1	14.5	13.8	14.6	14.3
7	1	2	2	1	1	2	1	12.3	12.8	13.0	12.7
8	2	2	2	2	2	2	2	15.2	15.7	15.6	15.5

Table 9. ANOVA for the design of experiment with 3 factors

Factor	Factor Effects	Sum of Squares	DOF ν	MS	F	$F_{\alpha, v1, v2}$		Significant	
						α=1%	α=5%	α=1%	α=5%
E	2.40	34.560	1	34.560	307.20	8.53	4.49	Yes	Yes
F	1.40	11.760	1	11.760	104.53	8.53	4.49	Yes	Yes
G	2.80	47.040	1	47.040	418.13	8.53	4.49	Yes	Yes
EF	-0.55	1.815	1	1.815	16.13	8.53	4.49	Yes	Yes
EG	1.25	9.375	1	9.375	88.33	8.53	4.49	Yes	Yes
FG	0.65	2.534	1	2.534	22.52	8.53	4.49	Yes	Yes
EFG	-0.30	0.540	1	0.540	4.80	8.53	4.49	No	Yes
Error		1.800	16	0.1125					
Total		109.425	23						

3.2.1. Factorial design experiments:

Table 8 shows the calculation necessary for these experiments. The main effects were: E = 2.4; F = 1.4; and G = 2.8. The interactions between two and three factors were: EF = -0.55; EG = 1.25; FG = 0.65; and EFG = -0.30. The main effects from strong to weak were: G, E, and F. The effects of interaction between two and three factors from strong to weak were: positive effects of EG and FG, and negative effects of EF and EFG. As expected, there were some interactions between EF, EG, FG, and EFG.

3.2.2. Analysis of variance (ANOVA): Table 9 shows the ANOVA results obtained for the factorial design experiment including factor, factor effects, sum of squares (SS), degrees of freedom (ν), mean square (MS), calculated F value (F), tabulated F value ($F_{\alpha, v1, v2}$) for α = 1% and α = 5% and the result of significance for each value of α.

For the level of significance α=1%, the main factors E, F, and G as well as the interactions EF, EG, and FG were significant whereas the interaction EFG was not significant. For the level of significance α=5%, the main factors and all the interaction factors were significant. The results were similar to that of Taguchi methods, but the factorial design method required 8 experiments instead of 4.

3.3. Quality of the final product

The factors and levels considered for the final product are shown in Table 10.

Table 10. Factors and levels for the final product

Factor	Variables	Levels	
		1	2
B	Fat (%)	3.0	6.0
C	Acidity (° Dornic)	15.0	19.0
D	Density (gm/cm^3)	1.031	1.035
H	Total dry extract (%)	11.5	12.5
I	Non-fat dry extract (%)	8.5	9.0
J	Package weight (gms)	1,032	1,033
K	Storage temperature (°C)	4.0	5.0

Table 11. Experimental design for the final product

No.	B	C	D	H	I	J	K	Response Values (%)			Mean Response	$(S/N)_S$
	1	2	3	4	5	6	7	1	2	3	(%)	(dB)
1	1	1	1	1	1	1	1	9.8	9.3	9.7	9.6	-19.65
2	1	1	1	2	2	2	2	8.7	9.4	8.9	9.0	-19.09
3	1	2	2	1	1	2	2	10.4	11.0	9.8	10.4	-20.35
4	1	2	2	2	2	1	1	8.1	8.4	9.0	8.5	-18.60
5	2	1	2	1	2	1	2	11.3	10.8	10.9	11.0	-20.83
6	2	1	2	2	1	2	1	11.9	12.5	12.8	12.4	-21.87
7	2	2	1	1	2	2	1	10.1	9.5	9.8	9.8	-19.83
8	2	2	1	2	1	1	2	10.6	9.9	10.4	10.3	-20.26

Table 12. Response table for the final product

Level	B	C	D	H	I	J	K
1	9.375	10.500	9.675	10.200	10.675	9.850	10.075
2	10.875	9.750	10.575	10.050	9.575	10.400	10.175
Δ	1.5	0.75	0.9	0.15	1.1	0.55	0.10

Table 13. $(S/N)_s$ Response table for the final product

Level	B	C	D	H	I	J	K
1	-19.4225	20.36	-19.7075	-21.165	-20.5325	-19.835	-19.9875
2	-20.6975	-19.76	-20.4125	-19.955	-19.5875	-20.285	-20.1325
Δ	1.275	0.60	0.705	0.21	0.945	0.45	0.145

Table 11 shows the experimental design for the final product inspection, $L_8\ (2^7)$. The response is shown in Table 12 along with the absolute difference, Δ between level 1 and level 2. The effects of each factor and its level were calculated, for example, B_1 = 9.375 and B_2 = 10.875. A simple rule was used as a guideline to analyze which of the factors had strong effects on the process and which were merely natural variations. The largest difference divided by 2 = 0.75. All differences equal to or above 0.75 were considered to be strong effects. Because this experiment was designed for *smaller-the-better* performance, the strong factors and their levels were B_1, I_2, D_1, and C_2. These factors correspond to fat 3%, non-fat dry extract 9%, density 1.031 gm/cm^3, and acidity 19^{o}D. The expected value for the chosen factor levels was found to be 8%.

The $(S/N)_s$ values for the final product are shown in Table 13. The strong factors (1.275/2 = 0.6375) and their levels were B_1, I_2, and D_1. These factors correspond to fat 3%, non-fat dry extract 9%, and density 1.031 gm/cm^3. The expected value for the chosen factor levels was found to be –18.5975 dB.

The above illustrations clearly show how factorial design experiments and Taguchi methods can be applied in controlling the quality of milk received from the vendors, during pasteurization, and during final product inspection. The use of Taguchi methods not only found the optimal value but also was cost effective since less number of experiments were needed compared to the classical methods such as factorial design experiments.

4. Conclusions

Based on the illustrations, it is clear that factorial design experiments and Taguchi methods can effectively be used for inspection and control of variables that affect the quality of milk and its derivatives. The use of these methods for analyzing the variables can provide a preventive quality system and valuable data for controlling the process at all levels. It has been point out that the Brazilian dairy industries must adopt as soon as possible the Hazard Analysis Critical Control Points (HACCP) system in order to guarantee quality and safety of milk and its derivatives consumed by the population. It is necessary to establish short and long-term programs jointly with industries, universities, and official organizations of inspection to implement the best practices of production and adequate development of HACCP program. It is also necessary to promote training courses for qualifying the people who work in the inspection department in order to make proper decisions. Inspection services are essential for the dairy products processing enterprises for their survival. It is necessary to unify resources to improve the quality of dairy products thereby increasing the productivity of the industrial sector. The quality tools must be more frequently utilized in the dairy industries to detect quickly the existing problems and provide solution to the problems. This will provide opportunities for the small and medium size dairy industries to be competitive in the market place.

5. References

[1] R. N. Kackar, "Off-line Quality Control, Parameter Design, and the Taguchi Method", Journal of Quality Technology, 17 (4), 1985, pp. 176-188.

[2] G. Taguchi, E. A. Elsayed, and T. Hsiang, Quality Engineering in Production Systems, McGraw-Hill, New York, 1989.

[3] M. S. Phadke, Quality Engineering using Robust Design, P. T. R. Prentice Hall, Englewood Cliffs, New Jersey, 1989.

[4] G. Box, "Signal-to-Noise, Performance Criteria, and Transformations", Technometrics, v. 30, 1988, pp. 1-40.

[5] J. S. Hunter, "Statistical Design Applied to Product Design", Journal of Quality Technology, 17(4), 1985, pp. 210-221.

[6] R. N. Leon, A. C. Shoemaker, and R. N. Kackar, "Performance Measures Independent of Adjustments", Technometrics, 29, 1987, pp. 253-285.

[7] V. N. Nair, and D. Pregibon, "A Data Analysis Strategy for Quality Engineering Experiments", AT & T Technical Journal, 65, 1986, pp. 73-84.

[8] D. C. Montgomery, Design and Analysis of Experiments, John Wiley, New York, 1991.

[9] A. C. Shoemaker, K-L. Tsui, and C. F. Wu, "Economical Experimentation Method for Robust Design", Technometrics, 33, 1991, pp. 415-426.

[10] V. Tronco, Proposta de Melhorias na Inspeção de Leite, M. S. Thesis, Graduate Program in Production Engineering, Federal University of Santa Maria, Brazil, 1996.

The Journal of Management and Engineering Integration
Vol. 1, No. 2, December, 2008

An AGV Dispatching Algorithm with Look-ahead Procedure for a Tandem Multiple-load AGV System

Daebeom Kim
Kangnam University
dbkim@kangnam.ac.kr

Eui Park, Celestine Ntuen, and Younho Seong
North Carolina A&T State University
park@ncat.edu, ntuen@ncat.ed, and yseong@ncat.edu

Abstract

This paper proposes a new dispatching algorithm for a multiple-load vehicle in a tandem configuration AGV system. The algorithm prioritizes all move requests based on look-ahead information such as AGV travel time, the time blocking occurs at a workstation, the time an empty slot in a drop-off station becomes available, the time starvation is expected to occur at a workstation, and the lock condition of workstations. Through simulation experiments, the performance of the proposed algorithm is compared with several existing dispatching algorithms in terms of system throughput.

1. Introduction

AGVS (Automated Guided Vehicle System) is widely used for inter-process material handling in intelligent and automated production systems. Lately, high-performance AGV which can handle multiple-load is used. Due to the multiple-load characteristics of AGVs, research on the system design and operation has been restricted so far. To reduce the complexity of the operation of the multiple-load AGV, research on the tandem configuration is currently conducted [1]. As shown in Figure 1, the tandem AGV system - a method to fundamentally avoid the collision and congestion of vehicles - divides workshops into several groups, and one AGV is assigned to each workshop group.

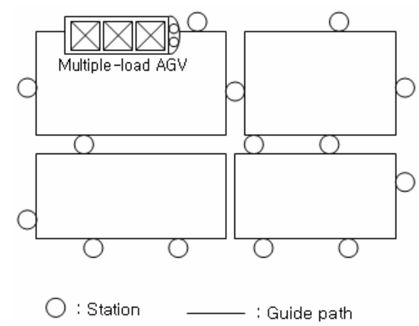


Figure 1. Tandem multiple-load AGV system

The First-Encountered-First-Served (FEFS) rule can be implemented for multiple-load AGVs where the AGVs pick up any load they encounter as long as there are empty positions on board. Loads are dropped off when their destinations are encountered [5]. Nayyar and Khator [6] examined the performances of single-load AGV's and multiple-load AGV's under various combinations of load pickup rules and AGV dispatching rule. Under batch manufacturing system, the results indicated that the 'shortest distance' dispatching rule has the best performance. The effectiveness of double AGV's on several layouts is examined [2, 7]. Also proposed was the multiple-load AGV operation method which considers the characteristics of production systems such as JIT environment, flexible manufacturing systems [4, 8, 9]. Based on the fuzzy set theory, Kim *et al*. [3] proposed an AGV dispatching algorithm for more flexible and adaptable operation of multiple-load AGVS.

The performance of a multiple-load capacity AGV is better than that of a single-load AGV, but the increase of the system performance cannot be expected if the system operation rule is set up incorrectly. In other words, if the near future system status is not considered in advance when loads are picked up or dropped off by multiple-load AGV, the performance can be similar to just that of a single-load AGV. In spite of the potential benefits of look-ahead information with multiple-load capacity, few papers on the dispatching algorithms with look-ahead function have been published.

This research proposes a new vehicle dispatching algorithm using look-ahead information in a tandem multiple-load AGV system. All move requests are prioritized based on dispatching efficacy and lock condition of workstation.

2. Algorithm development

The system we consider consists of multiple AGVs with multiple-load capacity and workstations, each of which has one pickup station and one drop-off station corresponding to the outgoing buffer and the incoming buffer, respectively. The guide path is configured with a tandem loop and a single AGV is assigned in each loop. This system has a many-to-many pickup/drop-off configuration in which pickup stations as well as drop-off stations are treated as distinct points.

If the AGV is empty, it selects several move requests based on the lock condition of workstation and the dispatching efficacy value of each move request. If the vehicle is empty and no move request is waiting from the outgoing buffers of the workstations, it stops to a predetermined point and dwells until new move request are available.

One dimension of the AGV operational problem is that the incoming and outgoing queue capacities at workstations are limited. Therefore, there is always a possibility that a particular workstation can be blocked or the system can be locked due to limited queue capacities. *Blocking* occurs when a workstation can not move its part to its outgoing buffer if the buffer is full. Whereas, *locking* occurs when the system is totally prevented from functioning, i.e. when no part movement can be achieved in the system. Therefore queue capacities must be taken into account in operation of AGVS.

A part from the head of the queue at a pickup station is loaded to an AGV (i.e. sequential pickup design). When the AGV arrives at a drop-off station, it is assumed that any part on the AGV can be unloaded regardless of its loaded position (i.e. direct drop-off design).

2.1. Dispatching efficacy

The dispatching efficacy is determined by the following three look-ahead system attributes: 1) travel distance from the AGV location to its pickup and drop-off station associated with the move request (to decrease the traveling time of the vehicle), 2) the expected duration of starvation or blocking that the workstation related to the move request will go through until the vehicle's arrival at the station. At each dispatching decision point, several move requests at neighbor workstations are selected. All pickups of the selected requests are accomplished first and then their drop-offs. To prevent too few communications with central computer, the concept of target dispatching time interval is introduced.

The following notations are made to define the dispatching efficacy.

q = index of move request
l = part type index, l=1,...,L
m = workstation index, m=1,...,M
i = index of move request q's pickup station (i=1,...,M) and
j = index of move request q's drop-off station (j= 1,...,N)
p_{lm} = required processing time of part type l at workstation m
I_m = incoming buffer capacity of workstation m [unit]
O_m = outgoing buffer capacity of workstation m [unit]
$I_m(t)$ = ordered set(FCFS base) of the indices of job types waiting at

incoming buffer of workstation m at decision time t

$OE_m(t)$ = number of empty buffer slots at outgoing buffer of workstation m at decision time t

$OF_m(t)$ = ordered indices of the first $OE_m(t)$ number of jobs in $I_m(t)$

$$X_m(t) = \begin{cases} 1 \text{ if workstation } m \text{ is busy at time} \\ 0 \text{ otherwise} \end{cases}$$

$\gamma_m(t)$ = the completion time of the job in process at workstation m at time t

T_i = travel time from current AGV location to pickup station i

$T_{i,j}$ = travel time from pickup station i to drop-off station j

m_i (m_j) = workstation index m corresponding to the pickup station i (drop-off station j)

TB_{m_i} = the time blocking occurs at workstation m_i in case no pickup is scheduled

$$= \begin{cases} t + \gamma_{m_i}(t) + \sum_{l \in OF_{m_i}(t)} p_{lm_i} & \text{if } \left|I_{m_i}(t)\right| + X_{m_i}(t) > OE_{m_i}(t) \\ \infty & \text{otherwise} \end{cases}$$

$TE_j(t)$ = the time an empty slot in drop-off station j becomes available

$$= \begin{cases} t & \text{if } \left|I_{m_j}(t)\right| < I_{m_j} \\ t + \gamma_{m_j}(t) & \text{if } \left|I_{m_j}(t)\right| = I_{m_j}, O_{m_j}(t) < O_{m_j} \\ \infty & \text{otherwise} \end{cases}$$

TS_{m_j} = the time starvation is expected to occur at workstation m_j in case no delivery is scheduled

$$= \begin{cases} t + \gamma_{m_j}(t) + \sum_{l \in I_{m_j}(t)} p_{lm_j} & \text{if } \left|I_{m_j}(t)\right| + X_{m_j}(t) + O_{m_j}(t) > O_{m_j} \\ \infty & \text{otherwise} \end{cases}$$

The dispatching efficacy is defined as

$$E(q) = \frac{1 + D_q^{\alpha}}{T_q}$$

where, $T_q = T_i + T_{i,j}$

$D_q = D_i + D_j$

$D_i = \max\{t + T_i - TB_{m_i}, 0\}$

$D_j = \max\{t + T_i + T_{i,j} - TS_{m_j}, 0\}$

α = scaling factor

D_i is the time duration of blocked status at pickup station i. D_j is the time duration of starved status at drop-off station j. D_q is the expected duration of starvation and locking of the workstation to which the move request is related will go through until the vehicle's arrival at the station. A move request with a higher $E(q)$ is ranked higher than one with a lower $E(q)$. A move request has higher priority if the travel distance is smaller from the current AGV location to pickup station i and from pickup station i to drop-off station j. A workstation which has or is expected to have blocking (starvation) needs the immediate service of the vehicle. Among those troubled workstations, we give higher priority to one whose expected time of blocking and starvation is longer. α is a user supply scaling constant. The effect of the blocking (starvation) compared to the AGV travel time can be controlled by the α .

2.2. Stepwise description of the algorithm

An AGV with the capacity of three unit loads may pick up three loads at near stations consecutively and then drop-off operations are carried out. As a result, the AGV may make less frequent communications with the central computer. To prevent too few communications, the concept of *target dispatching time interval(TDTI)* is introduced.

Let S be a set of move requests assigned to the vehicle at a dispatching decision point. *TDTI*, set by the system operator, determines the number of requests in S as follows:

$$\sum_{n \in S} \begin{pmatrix} \text{needed time to} \\ \text{accomplish the request } n \end{pmatrix} \le TDTI \quad (1)$$

Smaller values of *TDTI* result in more frequent communications and low utilization of an AGV's load space, while providing the central computer with more accurate system information for dispatching decisions. Longer *TDTI* results in less frequent communications and high utilization of an AGV's load space, but gives very rough information of the system. In prescribing the value of *TDTI*, the system operator has to trade off communication frequency, utilization of AGV's load space, and the accuracy level of system information for dispatching decisions.

Let

C = AGV load capacity

$S_q(t)$ = dispatching request list at dispatching decision point t

q_k^* = kth move request selected to delivery at dispatching decision, k=1 …C

$NP_{q_1^*}$ = neighbor set of q_1^*'s pickup station

$ND_{q_1^*}$ = neighbor set of q_1^*'s drop-off station

The neighbor means a pickup (drop-off) station located near another pickup (drop-off) station. The neighbor set of each pickup or drop-off station is predetermined by the system operator.

The proposed algorithm goes through a number of iterations to find job requests in *S* one by one as follows: Suppose the first request is selected as an element of *S* at an iteration. Then the next request is chosen only after dispatching efficacy of the unassigned job requests are adjusted reflecting the change of the vehicle's location and candidate move requests resulted from the previous selection.

Now, we present the algorithm steps.

Step 1: At decision point *t*, identify move request set $S_q(t)$

Step 2: Check the lock condition of workstation related each move request, $TB_m(t) \le t$ and $I_m(t) = I_m$. Dispatch AGV to the locked workstation and stop. Locked workstation means that workstation is blocked and its incoming buffer is full simultaneously.

Step 3: Calculate dispatching efficacy *E(q)* for each move request. If all move requests are impossible to be delivered because of the time $TE_j(t)$, the AGV waits at its current position. The algorithem considers the waiting of vehicle, instead of loading the relevant part, according to the look-ahead information, even if there is a move request for a part and loading space in the AGV.

Step 4: Select the first move request with the highest ranking value of *E(q).*

Step 5: Find the next request among the move requests in the neighbor workstation related with q_1^*. The next candidate requests are restricted to the $NP_{q_1^*}$ and $ND_{q_1^*}$. This iteration continues until the constraint (1), *TDTI,* and AGV load capacity limit *C* holds.

Step 6: Dispatch AGV to the requests in *S*.

Step 7: Repeat steps 1 to 6 at every decision point.

As the number of locked workstations becomes larger, the chance of shop locking may become larger. To reduce system locking, the AGV is assigned to the locked workstation with the highest priority.

Each pickup is carried out with its selection order of selected move requests, and drop-offs are accomplished according to its pickup sequence. For instance, at a dispatching decision point, an AGV with the capacity of two unit loads may pick up two loads consecutively before drop-off operations are carried out. The possible dispatching patterns are as follows: Case 1: Pickup A $\succ$ Drop-off A (Partially loaded), Case 2: Pickup A $\succ$ Pickup B $\succ$ Drop-off A $\succ$ Drop-off B (Full loaded), Case 3: Dwells at current position (Not dispatched until drop-off station becomes available). Table 1 shows a possible dispatching request list.

Table 1. Dispatching move request list

Order	Request position	Job type	Handling time(min)
1	W/S # 7	1(pickup)	1.5
2	W/S # 7	1(pickup)	0.5
3	W/S # 1	2 (drop-off)	1.5
4	W/S # 2	2 (drop-off)	1.0

(TDTI = 5 min)

3. Evaluation of the algorithm

Figure 2 shows the layout of a hypothetical job shop with nine workstations studied in this paper. Workstation 1 is the loading station for the raw material and workstation 9 represents the unloading station for the finished products. Parts are transferred in the system by an AGV system.

The experiments are carried out with the discrete simulator FACTOR/AIM. The following system parameter values and operational rules are use: Part pickup and drop-off time of AGV are 0.5 minute each. The AGV travels at the speed of 80 feet/min along the shortest distance path. Guide paths are bidirectional. Ten job types are to be produced whose routes and processing times are shown in Table 2.

The two well known dispatching rules are modified so that they can be adopted for the comparison. The two dispatching rules are the Nearest Request First (NRF) rule, and the Most Urgent Buffer Queue First (MUBQF) rule. A pickup request is generated when the processing of a part is completed at a workstation. A drop-off request is generated when a part is loaded on an AGV. NRF sends the AGV to the nearest workstation at which a pickup or drop-off action is required. Closeness can be defined in terms of AGV travel time. The move request which has the minimum value of urgency index has the highest priority in MUBQF. Urgency index is the value of [outgoing queue capacity-current outgoing queue length] for a pickup request and the length of current incoming queue for a drop-off request. If there is a tie, MUBQF selects the nearest and first request workstation. In the proposed algorithm (PROPOSED), the neighbor of each pickup (drop-off) station is defined as the station located within AGV travel time of 0.3 minute.

The performance of the proposed algorithm was studied under various capacities of the buffer queue and of the AGV. Throughput is defined as the total number of parts completed and removed from the shop floor during a unit time. The system throughput (in unit loads) over one unit time (8 hours) was obtained from five replications per rule on steady-state.

The buffer capacity was varied from one to five to examine the effects of buffer size. The load capacity of AGV was set to three. As shown in Figure 3, the proposed rule outperforms the other rules, especially with a larger buffer capacity. The effect of varying load capacity of AGV was also investigated and the results are shown in Figure 4. The buffer queue capacity used in this experiment was fixed and five. The proposed rule still performs better compared with the other rules. It can be seen that the performance of the proposed rule shows substantial improvements in throughput as the load capacity of AGV increases.

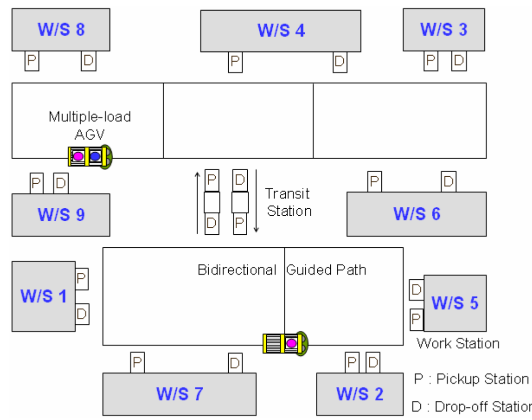


Figure 2. A hypothetical job shop

Table 2. Part routes and processing times

Part	Route	Process time(min)	Mix
1	1,6,4,7,9	1,15,25,20,1	0.1
2	1,7,5,4,2,8,3,9	1,5,20,15,15,5,15,1	0.1
3	1,8,6,2,5,3,7,9	1,8,10,20,15,5,15,1	0.1
4	1,6,2,5,9	1,20,10,15,1	0.1
5	1,3,5,2,7,8,9	1,15,5,15,18,20,1	0.1
6	1,2,3,9	1,20,25,1	0.1
7	1,4,7,3,6,5,9	1,15,20,12,8,15,1	0.1
8	1,5,3,8,4,2,9	1,20,5,15,20,8,1	0.1
9	1,3,5,8,7,4,6,9	1,8,20,10,10,5,5,1	0.1
10	1,8,6,7,4,9	1,15,15,20,13,1	0.1

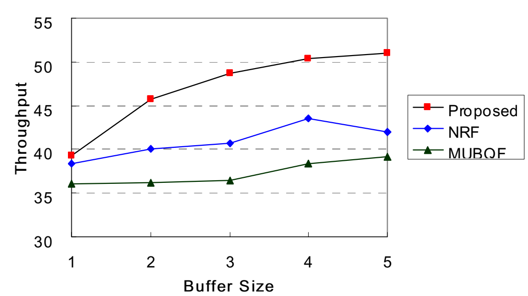


Figure 3. System throughput under varying buffer queue capacity

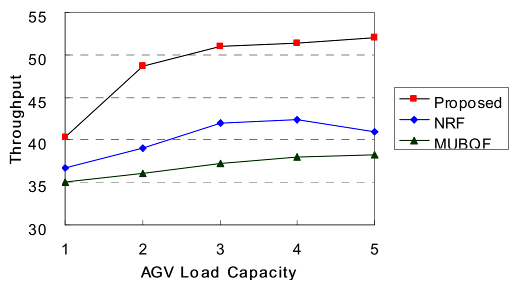


Figure 4. System throughput under varying load capacity of AGV

4. Conclusion

In this paper, an AGV dispatching rule using look-ahead system information in tandem multiple-load AGV configuration is proposed. The performance of the proposed algorithm is compared with several dispatching algorithms in terms of system throughput in a hypothetical job shop environment. Simulation experiments are carried out varying the capacity of the buffer queue and the load capacity of AGV. In all the cases, the proposed algorithm was shown to perform substantially better than the other rules. Although the results presented in this paper should be interpreted with reference to the hypothetical job shop and the experimental conditions described earlier, it is believed that the demonstrated advantages may be quite general.

Further research is needed on the issue of how to set the parameters in operation to enhance the performance.

References

[1] Chung, B. D. and Kim, K. S., "A Study on Design of a Tandem AGV System with Multi-Load AGVs", *IE Interface*, **15**(1), 2002, pp. 1-9.

[2] Hodgson, T. J., King, R. E. and Monteith, S. K., "Developing Control Rules for an AGVS using Markov Decision Processes". *Material Flow*, **4**, 1987, pp. 85-96.

[3] Kim, D. B. and Hwang, H., "A Dispatching Algorithm for Multiple-Load AGVS Using a Fuzzy Decision-Making Method in a Job Shop Environment", *Engineering Optimization*, **33**, 2001, pp. 523-547.

[4] Lee, J., Tangjarukij, M. and Whu, Z., "Load Selection of Automated Guided Vehicles in Flexible Manufacturing Systems", *International Journal of Production Research*, **34**(12), 1996, pp. 3383-3400.

[5] Liu, C. M. and Duh, S. H., "Study of AGVS Design and Dispatching Rules by Analytical and Simulation Methods", *Internal Journal Computer Integrated Manufacturing*, **5**(4), 1992, pp. 290-299.

[6] Nayyar, P. and Khator, T., "Operational Control of Multi-load Vehicles in an Automated Guided Vehicle System", *Computers and Industrial Engineering*, **25**(1-4), 1993, pp. 503-506.

[7] Ozden, M., "A Simulation Study of Multi-load-carrying Automated Guided Vehicles in a Flexible Manufacturing System", *International Journal of Production Research*, **26**(8), 1988, pp. 1353-1366.

[8] Occena, L. G. and Yokota,T., "Analysis of the AGV Loading Capacity in a JIT Environment" *Journal of Manufacturing Systems*, **12**(1), 1993, pp. 24-35.

[9] Sabuncuoglu,I., and Hommertzheim,D.L., "Dynamic Dispatching Algorithm for Scheduling Machines and Automated Guided Vehicles in a Flexible Manufacturing System", *International Journal of Production Research*, **30**,(5), 1992, pp. 1059-1079.

Modeling and Simulation of Total Ankle Replacement

Ha Van Vo
Mercer University
Macon, GA 31207
Vo_HV@mercer.edu

Barry Tuvel
Ankle & Foot Care
Center of Miami
DRTPOD@aol.com

R. Radharamanan
Mercer University
Macon, GA 31207
radharaman_r@mercer.edu

Abstract

The purpose of this study was to propose a new method for implantation of an ankle joint endoprosthesis. Key to this novel methodology was the customization and sizing of the implant using patient-specific geometry that is obtained using a combination of computerized tomography (CT) and finite element modeling. Computer modeling of the customized ankle implant also provided information related to the forces that will act upon the endoprosthesis, including pressure. Actual three-dimensional modeling of the ankle, following computer modeling, allowed for pre-implantation assessment of the endoprosthesis prior to operative intervention. This allowed for further studies to determine the optimal meniscus thickness that will go with the new customized endoprosthesis. This method was also used in calculating which bone density and what amount of bone removal would provide the best outcome for the customized implant and meniscus. It was the hope of this study to provide a more desirable result for total ankle joint replacement that will increase the life of the implant as well as provide patient satisfaction.

1. Introduction

Total ankle arthroplasty has been available since the 1970s [1] and has become a viable treatment for patients with painful tibiotalar arthritis who have not responded well to conservative treatments. It is generally preferred by surgeons who implant artificial joints in the lower extremities that the total ankle replacement (TAR) has been less successful than the replacement of hip and knee. Several factors have been considered to be causative in relation to failure of TAR in comparison to total knee and total hip arthroplasty, which include application of greater weight bearing loads (5 times body weight), a smaller surface area (9-11 cm^2) through which the load is dispersed, and greater degrees of freedom (movement) relative to the knee and hip [2, 3]. Recent advances in ankle implant design and availability have increased its popularity. The three most commonly used ankle endoprostheses currently available include the Alvin Agility Total Ankle (AATA) (Deput Orthopedics, Inc., a Johnson & Johnson company, 9140 K-Guilford Rd Columbia, MD 21046), the Buechel-Pappas (BP) (Endotec, 20Valley Street, South Orange, NJ 07079) device, and the Scandinavian Total Ankle Replacement (STAR), W. Link GMBH. & Co, Barkhausenweg 10, D-22339 Hamburg Germany) also known as Ramses device. Although each of these devices has certain desirable characteristics, they also convey a considerable number of problems that have been noted by other authors [2, 3, 4]; among the three, the AATA has been cited as the most popular of the currently available ankle endoprostheses [1, 4].

Every ankle endoprosthesis is subjected to failure when malalignment and less than optimal fit is present [5]. Less than optimal fit promotes eccentric loading of the tibial tray and progressive loosening, and therefore implant instability. In addition, these factors also promote excessive wear of the ultrahigh molecular weight polyethylene (UHMWPE) meniscus which ultimately will result in implant failure or stress fracture of the supporting bone [5]. Therefore, it is the objective of this study to determine a method to find the optimal solution of implant size and meniscus thickness according to patient specific geometry to improve on surgical outcomes of TAR.

2. Methods

To conduct the study, a CT scan of a 75-kg white female in her mid-fifties was randomly selected from the Ankle & Foot Care Center of Miami. The patient had previously been conservatively treated during a period of 2 years for painful arthritis of the right ankle. All forms of non-invasive treatments have failed to alleviate the pain. Symptoms of pain along with diminished range of motion have subsequently progressed. Findings from bone x-rays of the right ankle showed signs of degenerative joint disease (DJD). Surgical treatment via total ankle arthroplasty was addressed and highly recommended to the patient. With patient's consent, a CT scan of the right ankle was obtained to create a 3D finite element (FE) model. This was done using 180 2-D cross-sectional slices of the CT scan as shown in Fig 1.

The 3D FE model (Fig. 2) of the ankle complex was then created from the 2D images using ANSYS 8.1 (ANSYS Inc., 275 Technology Drive, Canonsburg, PA 15317). The model was subsequently used to help determine the proper sized ankle prosthesis for the patient by using patient specific geometry. Afterwards, the ankle prosthesis was incorporated into the model to analyze the effects of variable meniscus thickness, bone removal, and bone density on contact stresses.

During normal weight bearing, it is understood that the ankle joint supports a load of approximately 5 times the body weight. Therefore, a compressive force along the y-axis of approximately 3,678.75 N was applied constantly to the FE model (implant included) in order to analyze the contact pressure, δ_{yy} of the ankle joint while in neutral position.

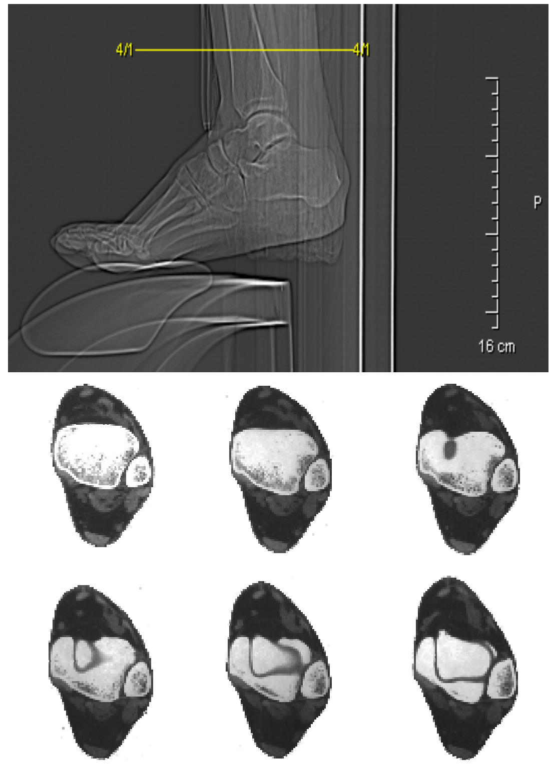
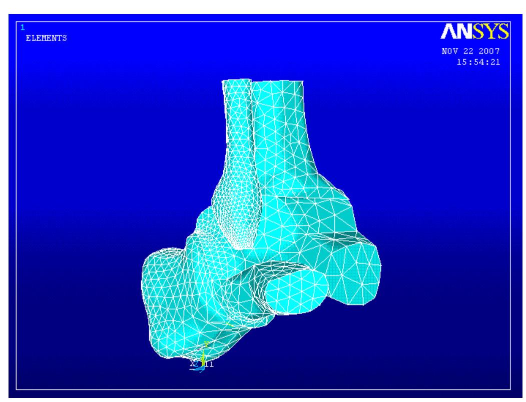
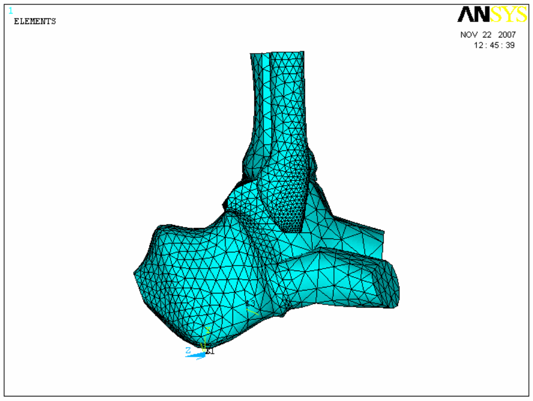


Figure 1. The 2D images generated from the CT scan of the ankle joint.

Contact pressure was examined between the inferior surface of the meniscus and the superior surface of the talar dome. The analysis was simulated in neutral position and in static conditions to eliminate variables such as malalignment and changes in force loads under dynamic conditions.



(a) Anterolateral view



(b) Lateral view

Figure 2. A 3D finite element model of the ankle complex.

Simulations were subsequently carried out for different meniscus thickness. Thickness was increased in 2 mm increments beginning with 2 mm and ending with 12 mm. The thickness was proportional to the thickness of tibial plateau bone removal. Ultimate contact stresses were analyzed and compared. Contact stresses were also determined for variable bone densities and different amounts of tibial bone removal. Data were collected and recorded as shown in Table 1. The results were also graphed as seen in Figs. 3 and 4.

Table 1. Data collected from this study by employing a load of 5 times the body weight of an average person of 75 kg.

(a) Bone mineral density

	Applied Loads (N)	Bone Mineral Density (gm/cc)	Ultimate Contact Stress (MPa)
1	3678.75	0.5	29.698
2	3678.75	0.6	28.753
3	3678.75	0.7	26.995
4	3678.75	0.8	25.771
5	3678.75	0.9	24.927
6	3678.75	1.0	24.027
7	3678.75	1.1	23.386
8	3678.75	1.2	22.733

(b) Meniscus thickness

	Applied Loads (N)	Meniscus Thickness (mm)	Ultimate Contact Stress (MPa)
1	3678.75	2	31.269
2	3678.75	4	28.494
3	3678.75	6	25.394
4	3678.75	8	24.861
5	3678.75	10	24.850
6	3678.75	12	23.889

(c) Tibial bone removal

	Applied Loads (N)	Tibial Bone Removal (mm)	Ultimate Contact Stress (MPa)
1	3678.75	2	32.135
2	3678.75	4	29.037
3	3678.75	6	26.504
4	3678.75	8	25.107
5	3678.75	10	24.105
6	3678.75	12	23.889

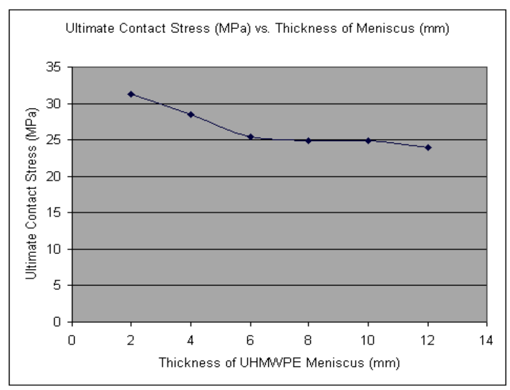


Figure 3. Increasing the thickness of the meniscus, the ankle joint experiences less contact stress.

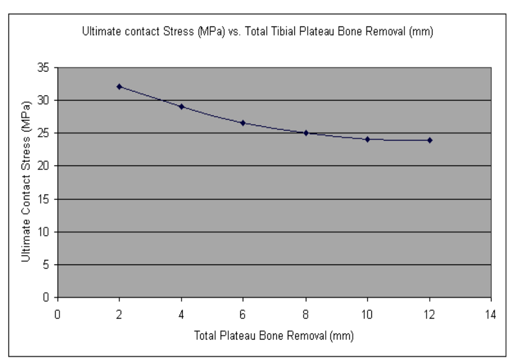


Figure 4. The relationships between ultimate contact pressure and thickness of bone removed from tibial plateau.

3. Results and discussions

It is seen from the simulation results that the contact pressure decreased dramatically when increasing meniscus thickness from 2.0 mm to 6.169 mm and decreased slightly when increasing the thickness of meniscus from 6.170 mm to 12.0 mm as shown in Fig. 3. The optimal thickness came out to be 8.468 mm. These results were found to be comparable to the results of previous studies. In a similar manner, the effect of UHMWPE thickness on contact pressure was described and analyzed by Gur et al [6]. The study was based on the allowable contact stress during compressive loading from 0 to 60 MPa applied to a typical knee design while changing the meniscus thickness from 4 to 24 mm. They concluded that there was an inverse relationship between thickness and contact stress. In addition, the study of Bartel et al [7, 8] also concluded that the stresses were inversely related to the thickness of the polyethylene. A similar study [9] from Orthoteers Company on wear rate of biomaterials determined that thin polyethylene between 6–8 mm in thickness increases fatigue wear and thereafter determined that the thickness level of the meniscus should be between 8–10 mm. So the results from this study are acceptable in terms of the thickness of meniscus for this patient and implant size. It can also be said that the contact stress component is a function of the meniscus thickness. On the other hand, when applying Gur's research to ankle implants, there are other important factors that must be taken into consideration such as differences in body load seen in the ankle joint as compared to the knee and also differences in surface contact area.

This study not only focuses on the effects of meniscus thickness but also on the amount of bone removal. Other research studies [10, 11] showed that the quantity of bone removed around the ankle joints created a statistically significant decrease in bone resistance to compression causing the implant to move further from the subchondral plate. Therefore, resecting more than 4–5 mm of bone from either side of the ankle joint can result in an increased risk of fatigue related bony fractures along the edges. In other words, they suggested that the total bone removed from both sides should be less than 10 mm. When dealing with the ankle implant, in order to be able to remove 4–5 mm from both the medial and lateral malleoli there must be a cut out of approximately 8 to 10 mm from the tibial plateaus; consequently, the tibial component thickness should be equal to the total talar bone removal which is about 2 mm to offset any shortening. Figure 4 shows the relationship between the contact pressure and tibial plateau bone removed. More tibial bone removal (related to increased medial and lateral malleoli removal) increases the area of surface contact which decreases the contact pressure. It is important to point out the region of the graph in Fig. 4 shows that the ultimate compressive stress (contact pressure) slightly decreases in a domain of 8–12 mm of tibial bone removed. This data suggests that the tibial plateau bone removal must be determined from this domain for optimal results. In Fig. 3, the graph not only exhibits the relationship between contact pressure and the thickness of the UHMWPE meniscus (size of the new implant depends on the thickness of the UHMWPE meniscus) but also expresses a stable region of contact pressure with the thickness of meniscus increasing from 7–12 mm. In other words, the optimal size of the meniscus should be within this domain (7–12 mm). Since the thickness of the UHMWPE meniscus should be equal to the tibial bone resected after installation of tibial and talar dome components, so at this common thickness they experience the same ultimate contact stress. The result in Figure 5 shows the optimal solution for this study. The finite element model of the new implant is shown in Fig. 6.

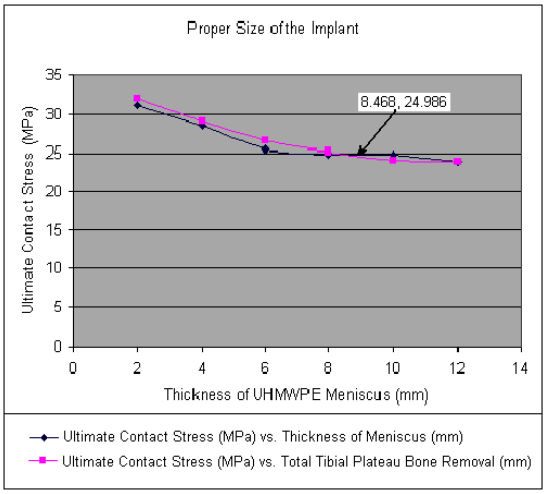
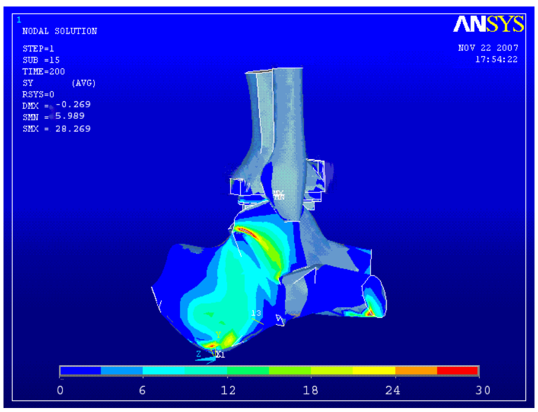
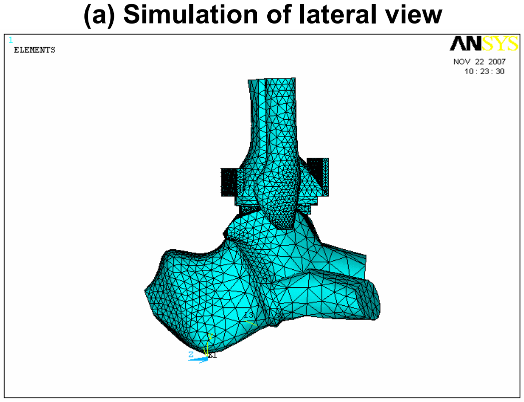


Figure 5. The optimal solution (size of the implant) determined from the intersection of graphs shown in Figs. 3 & 4.



(a) Simulation of lateral view



(b) 3D mesh of lateral view

Figure 6. Finite element model of the new ankle implant

Another factor that has a major effect on the ultimate compressive load (contact pressure) and the overall success of the ankle implant is bone density. The surgeon must account for the patient's bone density in order to prevent fractures of bone or total joint collapse due to the maximum contact pressure of the implant [12]. Osteoporosis patients are not ideal candidates for any of the orthopedic implants because they are not strong enough to withstand the fatigue load especially at the implant and bone interface. According to Table 1, bone density has a major effect on contact stress showing an inverse relationship.

4. Conclusions

The objective of this study was to create a novel method to determine the optimum solution of the proper meniscus thickness and implant size using ankle implant patent # US6863691 for patients with painful ankle arthritis. Previous studies on the effect of thickness of the polyethylene on contact pressure suggested that the thickness of the polyethylene meniscus be between 8–10 mm [7, 8]. In our study, the optimal solution for the thickness of UHMWPE meniscus was 8.468 mm which falls within this range.

From the study a flow chart was created as shown in Fig. 7 which describes the complete step-by-step procedure used to determine the proper sized implant for patient specificity. This process can be applied to other types of replacement surgeries such as hip and knee. With the proper size determined, further analysis can be carried out to obtain a proper meniscus thickness.

An actual customized model was built as shown in Fig 8; the talar component is at the bottom and the UHMWPE meniscus is inserted into the tibial component. This method can be incorporated into joint replacement surgeries to help reduce implant size errors that often occur during surgery. It gives a more customized fit to help preserve stability. To also help optimize surgical results, this study determines the optimal bone density and bone removal.

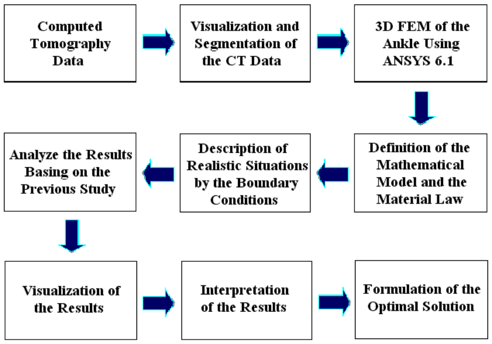


Figure 7. The diagram shows the process starting from 2D CT scans to 3D FE model to analyze the stress in order determine the proper size of the ankle implant

Figure 8. Picture shows the pre-operative implantation of the new ankle joint invented by one of the authors. This demonstrates the preparation for the actual implant.

To optimize the results of this method, more data collection is needed. In future investigations, Mimics software will be used to convert 2D images of the CT scans to a 3D FE model for a more accurate FE model. The studies will also consider other lower extremity disorders such as valgus deformities, varus deformities, cavus foot, and pes planus.

5. References

[1] S. F. Corti et al, "Complications of Total Ankle Replacement", 1(391), 2001, pp. 105-114, October.

[2] M. M. Felman, "The Total Ankle Joint Replacement", Study 98-99.

[3] H. M. Felman, "This Surgical Procedure Became Available to Americans in October of 1998", wysiyg://result-doc-frame.content, 2000.

[4] L. Claes, O. Eberhard and G. Bauer, "Engineering Considerations of a New Ankle Joint", http://www.cmmaccess.com, January 1999.

[5] S. M. Raikin and M. S. Myerson, "Complications of Total Ankle Replacement", pub/pressItem73.html. http://www.drmyerson.com/research

[6] H. Gur, H. Aktiengesellschaff and V. Kunstoffe, 6230 Frankfurt am Main 80, 1982, p. 22.

[7] D. L. Bartel et al, "The Effect of Conformity and Plastic Thickness on Contact Stresses in Metal Backed Plastic Implants", Journal of Biomechanical Engineering, 107, 1985, p. 193-199.

[8] D. L. Bartel, V. L. Bicknell and T. M. Wright, "The Effect of Conformity, Thickness, and Material on Stresses in Ultra-High Molecular Weight Components for Total Joint Replacement", Clinical Orthopedics and Related Research, 68-A (7), 1986, p. 1041-1051.

[9] Implants & Materials in Orthopedics, http://www.orthoteers.co.uk/Nrujpij33Lm/Orthbiomat.htm.

[10] G. K. Gitken et al, "Indentation Stiffness the Cancellous Bone in the Distal Human Tibia", Clin Ortho 201, 1985, 264-270.

[11] I. Hvid, O. Rasmusseu et al, "Trabecular Bone Strength Profile Ankle Joint", Clin Ortho 199, 1985, pp. 306-312.

[12] C. Mautalen et al, "Ultrasound and Dual X-ray Absorptiometry Densitometry in Women with Hip Fracture," Calcif Tissue Int., 57:3, Sep. 1995, pp. 165-168.

Genetic Algorithms to Minimize Makespan of a Capacitated Batch Processing Machine

Purushothaman Damodaran
Florida International University
damodarp@fiu.edu

Siddharth Mestry
Florida International University
smest001@fiu.edu

Murali Krishnamurthi
Northern Illinois University, DeKalb
mkrishna@niu.edu

Abstract

Printed Circuit Boards (PCBs) are an integral part of electronics products. After assembly, the PCBs are typically subjected to stress test to detect early failures. The stress testing chambers (or batch processing machines) can process several PCBs (or jobs) simultaneously. These machines are expensive and a bottleneck. Schedulers are challenged with two interdependent decisions, namely how to group jobs to form batches and how to schedule these batches to maximize the machine utilization. The job processing times and their sizes are known. The machine can process a batch of jobs simultaneously as long as the total size and the total number of jobs in a batch do not exceed its capacity. The batch processing time is equal to the longest processing job in a batch. The machine utilization can be improved by minimizing the makespan or the completion time of the last batch. The problem under study is NP-hard. Consequently, a Genetic Algorithm (GA) is proposed to minimize the makespan. The effectiveness of the proposed GA approach is evaluated by comparing its solution with the current practice at an electronics manufacturing facility and a Simulated Annealing approach.

1. Introduction

This research is motivated by a practical application observed at an electronics manufacturing facility. This facility is responsible to assemble and test PCBs for various applications. The three major process steps to assemble a PCB are solder paste deposition, component placement, and soldering. Each assembly line is typically dedicated to a product family. The machines required to carry out the assembly is typically arranged in a flow shop layout. After assembly, the PCBs are subjected to stress test in a chamber to detect any early failures.

There are different kinds of stress tests, namely thermal cycling, vibration – to name a few. The stress testing chambers can process several jobs in a batch simultaneously. PCBs from different assembly lines are lined up for testing in front of the chamber. The chamber can process a batch of PCBs as long as the total size and the total number of PCBs in the batch do not exceed its capacity. Each PCB is held in the chamber using a fixture. Consequently, the total number of PCBs processed in a batch cannot exceed the number of fixtures available. The sizes of the PCBs also differ depending upon the application in which they are used (e.g. smaller PCBs for cell phones and larger PCBs for servers). The chamber has limited space to hold the PCBs; hence the total size of all the PCBs in a batch should not exceed the available space.

The stress testing chambers are expensive and a bottleneck. Consequently, the schedulers strive to maximize its utilization. The utilization can be maximized by minimizing the completion time of the last batch processed (i.e. minimizing makespan) [1]. The schedulers are faced with two important, challenging and interdependent decisions: (1) how to group the jobs into batches, and (2) how to schedule these batches on the chambers such that the makespan is minimized. The current practice is to form the batches using a first-in first-out (FIFO) policy. The jobs are grouped into a batch based on when they joined the queue in front of the chamber. Clearly, this approach was not satisfactory to both the scheduler and the plant manager as the critical resource was not utilized to the fullest extent. Many batches were not filled to capacity and the makespan was also longer. Consequently, the primary objective of this research endeavor was to aid the scheduler to form and schedule the batches such that the makespan of the chamber is minimized. Hereafter the stress testing chamber and the PCBs are refereed to as a batch processing machine and jobs respectively.

2. Problem Description

The processing time (p_j) and the size (s_j) of the jobs are known. The customer specifies the minimum time (or processing time) for which each job should be tested. Consequently, when jobs with different processing times are grouped together, the batch processing time (P^b) is defined by the longest processing job. The capacity of the machine, both in terms of the size and the number of jobs it can handle is known. The objective is to minimize the makespan (C_{max}).

A special case of the problem under study can be shown to be the bin packing problem, which is a well known NP-hard problem. Consequently, the problem under study is also NP-hard. Given *n* items and their sizes the objective of the bin packing problem is to assign these items to bins such that the bin capacity is not exceeded, while minimizing the number of bins used. When the job processing times are constant, the scheduler has to group the jobs into batches such that the machine (or bin) capacity is not exceeded, while minimizing the number of bins (or batches) formed. The batches formed are equivalent to bins used. By minimizing the number of batches formed, the total time required to process all the jobs can be minimized. Since the problem under study is NP-hard, solving the problem to optimality may require prohibitively long CPU times. Consequently, a Genetic Algorithm (GA) is proposed. The performance (solution quality) of GA is evaluated by solving several problem instances and comparing the results with the current practice (FIFO) and a Simulated Annealing (SA) approach.

The remainder of the paper is organized as follows: Section 3 presents the literature reviewed on scheduling batch processing machines and GA. Section 4 discusses the GA approach and section 5 presents the experimental study conducted to evaluate the proposed approach. Section 6 concludes the paper with remarks for future research.

3. Literature review

Uzsoy [2] proposed heuristics to schedule a batch processing machine when the job sizes are non-identical. The objective was to minimize the makespan (C_{max}). Uzsoy and Yang [3] proposed branch-and-bound (B&B) and some heuristics to minimize the total weighted completion time. Dupont and Dhaenens-Flipo [4] proposed a B&B approach to minimize the C_{max} for a single batch-processing machine. Chang et al. [5] proposed SA to minimize the C_{max} of batch processing machines in parallel. Melouk et al. [6] proposed SA to minimize the C_{max} of a single batch processing machine. All the above references considered only one constraint; the total size of all the jobs in a batch cannot exceed the machine capacity.

Other researchers considered only the maximum number of jobs in a batch as a constraint. Lee et al. [7] presented dynamic programming algorithms to minimize several performance measures on a single batch processing machine. Ghazvini and Dupont [8] developed heuristics to minimize mean flow time on a single batch processing machine. Sung and Choung [9] analyzed the static case (all the jobs are available at time 0) and the dynamic case with different job-release times. For a detailed review on scheduling batch machines, refer to [10].

It can be observed that batch processing machines are typically scheduled using a heuristic, dynamic programming or B&B – especially to minimize the C_{max}. Although heuristics do not guarantee an optimal solution, they are preferred for two reasons viz., scheduling batch-processing machine problems are NP-hard and enumeration techniques and dynamic programming require prohibitively long run times. In this research, a GA approach is proposed to schedule a batch processing machine that is capacitated by both the number of jobs and the total size of the jobs in a batch.

Genetic algorithm tries to mimic the genetic behavior of a species. GA was first introduced by Holland in 1970. Since then it has been used to solve a variety of combinatorial optimization problems. Sevaux and Peres [11] minimized the weighted number of late jobs on a batch processing machine by employing GA. Koksalan and Keha [12] employed GA to minimize the flow time and the number of tardy jobs, and to minimize flow time and maximum earliness of a single machine. Wang and Uzsoy [13] employed GA for minimizing the maximum lateness of a batch processing machine in the presence of dynamic job arrivals. Li et al. [14] employed GA to address earliness and tardiness in production scheduling and planning. For an overview of successful applications of GA in business and industry, refer to Reeves [15].

Damodaran et al. [16] proposed a Simulated Annealing (SA) approach to solve the problem under study. SA is a stochastic local search approach which endeavors to overcome local optimality by accepting inferior solution with some probability. For a detailed discussion on the SA approach proposed for the problem under study, refer to [16]. GA is a population based approach as opposed to SA (population size = 1). Consequently, GA may outperform SA and hence may help to improve the machine utilization. The data set generated and used in [16] is used in this paper to evaluate the performance of the GA approach.

4. Solution approach

Genetic Algorithms are a particular class of evolutionary algorithms that use techniques inspired by evolutionary biology. In GA, the solutions are coded as strings called *chromosomes*. An initial population (solutions) is randomly generated. Offspring are generated by applying operators such as *crossover* and *mutation* to the parents. Crossover operators create new offspring solutions by exchanging parts of the selected parent solutions and mutation operators maintain the population diversity by slight perturbations of selected solutions [17]. In each generation, the fitness of every individual in the population is evaluated, multiple individuals are stochastically selected from the current population (based on their fitness), and modified (recombined and possibly mutated) to form a new population. The new population is then used in the next iteration of the algorithm. For the problem under study, the measure of fitness is C_{max}. The objective is to minimize C_{max}; hence, the chromosomes with smaller fitness measure are given an opportunity to form new chromosomes.

Typically in a single machine scheduling problem, when GA is applied, each chromosome is a permutation of the jobs. A similar representation is adopted in this study. Given a job permutation, the jobs can be grouped to form batches using the bin packing heuristics (e.g., First-Fit (FF) and Best-Fit (BF)). The BF heuristic can be described as follows: jobs from an ordered list are placed in the batch with least residual capacity; whenever a job does not fit in any one of the existing batches, a new batch is opened. The procedure is repeated until all the jobs are assigned to a batch. While assigning jobs to a batch, the machine capacity is not violated. In FF heuristic, jobs from the ordered list are assigned to the first feasible batch.

After the batches are formed, the batches can be sequenced in any order. For a single (discrete or batch processing) machine problem the C_{max} is equal to the sum of the processing times and does not dependent upon the sequence in which the jobs are processed [1].

The proposed GA approach for the problem under study is given below:

1. Create initial population of chromosomes. Set generation = 0;
2. Form batches using FF or BF heuristic. Determine the batch processing times, and compute C_{max} for each chromosome;
3. Select parents to form offspring using crossover and mutation;

4. Compare and replace parents with the offspring to form next generation; and.
5. Repeat steps 2- 4 till generation = N.

Several researchers [11, 13-14] proposed an initial population with *n* or 2*n* chromosomes, where *n* is the number of jobs. Large populations may require long computational time whereas small populations run the risk of failing to cover the solution space adequately. Initial experiments on large problem instances, with 50 jobs, indicated that a constant population size of 10 would be adequate for instances studied in this research. For nine chromosomes the jobs were randomly sequenced and the bin packing heuristic was applied to form the batches. For one chromosome the jobs were arranged in non increasing order of their processing times before applying the bin packing heuristic.

The parents are selected by employing a tournament-based selection for the crossover operation. A randomly selected crossover point divides the chromosome of each parent into two parts (i.e. one-point crossover). Crossover is applied to the job sequence of the parents (see Figure 1). After crossover the bin packing heuristic is applied to the resulting job sequence to form batches. Each offspring is formed by first copying the left part of the corresponding parent's chromosome as it is. The remaining part of the offspring's chromosome is obtained by rearranging the jobs on the right side of the corresponding parents' chromosome as per the sequence in which they appear on the other parent chosen for the crossover. The crossover process results in the generation of two offspring. In order to maintain a diverse population, the offspring is mutated with a probability of 0.01. The mutation point is randomly chosen and the jobs on either side of the mutation point are swapped. The GA was terminated after allowing it to run for 1000 generations.

The C_{max} obtained from the GA approach were compared to the current practice at an electronics manufacturing facility and a SA approach. The GA approach was implemented in MatLab 6.5.

M/c Cap. (S) = 10 units; Max. jobs/batch (D) = 3

Jobs	1	2	3	4	5	6	7	8	9	10
p_j	9	7	9	6	6	7	3	8	4	1
s_j	4	6	2	3	2	8	5	2	7	5

Parent Chromosomes (Before Crossover):

Crossover Point (between the 5th and 6th positions)

Parent 1	4	6	8	3	1	5	2	10	7	9
Parent 2	3	10	7	4	2	6	8	5	9	1

Parent 1

Batch No.	Jobs	P^b
1	4, 8, 3	9
2	6, 5	7
3	1, 2	9
4	10, 7	3
5	9	4
	C_{max}	32

Parent 2

Batch No.	Jobs	P^b
1	3, 10, 4	9
2	7, 8, 5	8
3	2, 1	9
4	6	7
5	9	4
	C_{max}	37

Offspring Chromosomes (After crossover):

Offspring 1	4	6	8	3	1	10	7	2	5	9
Offspring 2	3	10	7	4	2	6	8	1	5	9

Offspring 1

Batch No.	Jobs	P^b
1	4, 8, 3	9
2	6, 5	7
3	1, 10	9
4	7	3
5	2	7
6	9	4
	C_{max}	39

Offspring 2

Batch No.	Jobs	P^b
1	3, 10, 4	9
2	7, 8, 5	8
3	2, 1	9
4	6	7
5	9	4
	C_{max}	37

Figure 1. Crossover Mechanism

5. Experimentation

Due to confidentiality agreements the actual data used for evaluating the GA cannot be disclosed. However, the results presented in this section are similar to what was observed when the actual data was used.

Table 1 presents the parameters used to generate the problem instances. The machine capacity or batch size (*S*) was fixed at 20. Each instance was replicated six times with different random number streams. Each replicate consists of 24 problem instances. Altogether the total number of experiments conducted was 144 (6 replicates x 24 instances / replicate).

All the experiments were run on a Pentium 4, 1.89 GHz computer with 256 MB RAM. The current practice (hereafter referred to as CP) observed was simulated by creating 20 random

Table 1: Factors and Levels

Factors	Levels
No. of jobs (J)	10, 20, 50
Processing time (p_j)	Discrete Uniform [1,10], Discrete uniform [1,20]
Job size (s_j)	Discrete Uniform [1,20], Discrete uniform [2,6]
Max. no. of jobs in a batch (D)	3, 5

job sequences and grouping the jobs using a FIFO policy. The best C_{max} from CP was compared to GA and SA. GA was run for 1000 generations and results were recorded at intervals of 50 generations.

Figure 2 presents the percentage improvement in C_{max} obtained from GA when compared to CP and SA. Instances numbered 1-8 consist of 10 jobs. Instances numbered 9-16 and 17-24 consist of 20 and 50 jobs respectively. On twelve out of 144 instances, both GA and CP reported same C_{max}. For the remaining instances GA outperformed CP.

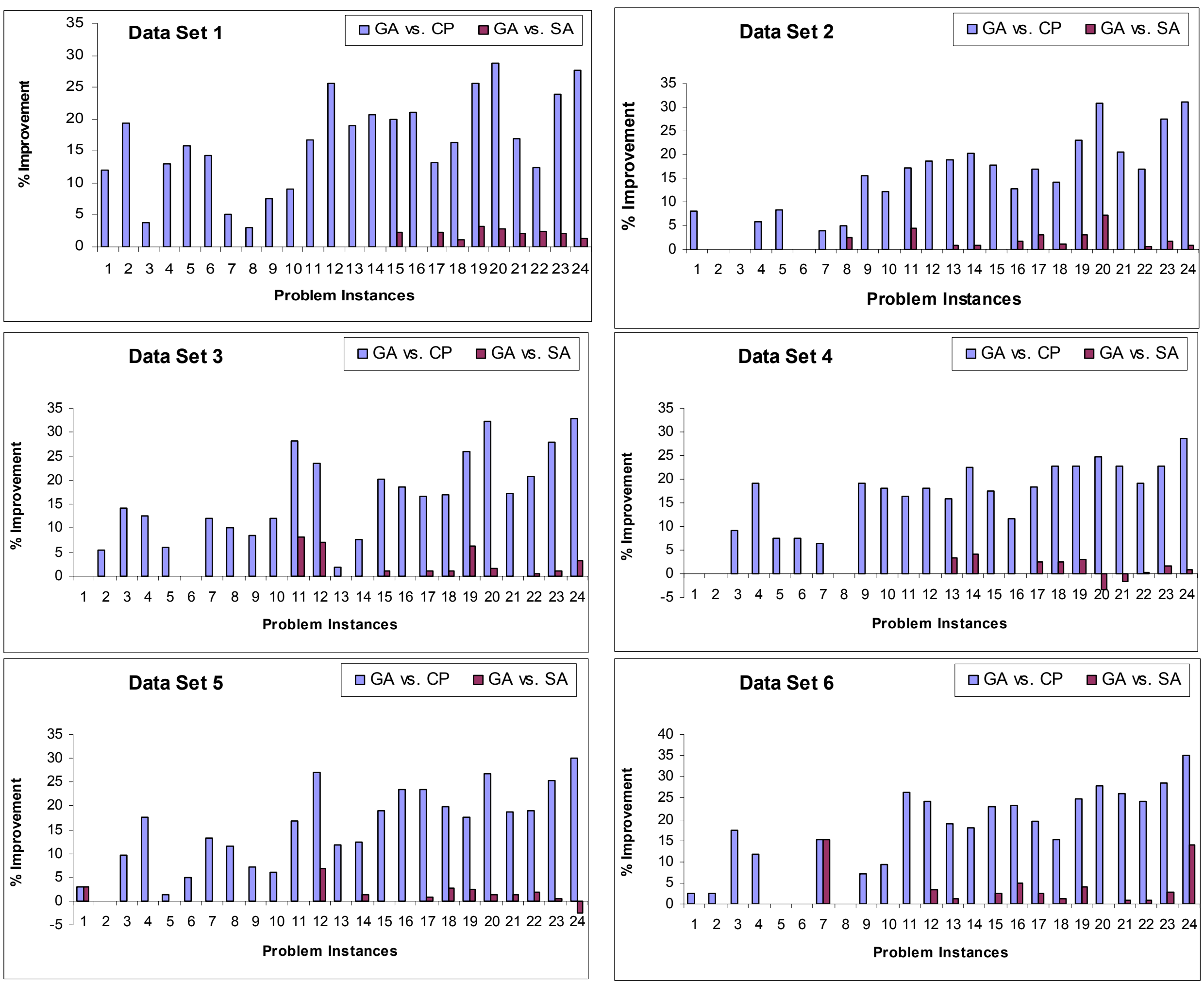


Figure 2. Percentage Improvement in Makespan

GA outperformed SA on several 20 and 50 job problem instances. GA solution was as good as or better than SA on 141 instances. GA was better than SA on 61 out of 144 instances. There were three instances (data set 4 and 5 with 50 job instances) for which SA solution was better than GA. The solution obtained from GA was also compared to a commercial solver used to solve the mixed-integer programming model proposed in [16]. GA outperformed the commercial solver on all 20 and 50 job problem instances. On an average GA was 11.41% better than the solver on 50 job problem instances. The commercial solver reported same results as obtained from GA and SA for 10 job problem instances. However, the commercial solver required longer run time.

Table 2 presents the average percentage improvement in solution when GA is compared with CP and SA. The percentage improvement is higher as the number of jobs increase – especially GA vs. CP.

Table 2. Average Improvement (in %)

Data Set	10 Jobs		20 Jobs		50 Jobs	
	GA vs. CP	GA vs. SA	GA vs. CP	GA vs. SA	GA vs. CP	GA vs. SA
1	10.81	0.00	17.42	0.28	20.63	2.18
2	3.92	0.32	16.70	1.00	22.56	2.16
3	7.55	0.00	15.10	2.08	23.82	1.96
4	6.21	0.00	17.41	0.94	22.70	0.75
5	7.60	0.37	15.47	1.05	22.54	1.08
6	6.13	1.89	18.79	1.52	25.16	3.25

6. Conclusions

This paper presents a GA approach to minimize the makespan of a batch processing machine with two constraints (i.e., the total batch size and the number of jobs in a batch). Experiments were conducted to evaluate the performance (solution quality) of the GA approach with the current practice at an electronics manufacturing facility, SA approach, and a commercial mixed-integer solver. On 10 jobs problem instances, GA, SA and the commercial solver reported same results. GA outperformed CP and the commercial solver on almost all the problem instances. It performed well when compared to SA on several 20 and 50 job problem instances. The proposed solution approach is currently used by an electronics manufacturing facility to schedule their test chambers.

7. References

[1] Pinedo M, Scheduling: theory, algorithms, and systems, Englewood Cliffs, NJ, Prentice-Hall, 2002.

[2] Uzsoy R., "Scheduling a single batch processing machine with nonidentical job sizes," International Journal of Prod Research, 1994, 32(7), 1615–35.

[3] Uzsoy R, Yang Y., "Minimizing total weighted completion time on a single batch processing machine," Production and Operations Management, 1997, 6(1), 57–73.

[4] Dupont L, Dhanence-Flipo C., "Minimizing the makespan on a batch machine with non-identical job sizes: an exact procedure," Computers & Operations Research, 2002, 29, 807–19.

[5] Chang PY, Damodaran P, Melouk S., "Minimizing makespan on parallel batch-processing machines," International Journal of Production Research, 2004, 42(19), 4211–20.

[6] Melouk S, Damodaran P, Chang PY, "Minimizing makespan for a single machine batch processing with non-identical job sizes using simulated annealing," International Journal of Production Economics, 2004, 87(2),141–7.

[7] Lee CY, Uzsoy R, Martin-Vega LA, "Efficient algorithms for scheduling semiconductor burn-in operations," Operations Research, 1992, 40(4), 764-775.

[8] Ghazvini FJ, Dupont L., "Minimizing mean flow time criteria on a single batch processing machine with non-identical jobs sizes," International Journal of Production Economics, 1998, 55, 273–80.

[9] Sung CS, Choung YI, "Minimizing makespan on a single burn-in oven in semiconductor manufacturing," European Journal of Operational Research, 2000, 120, 559–74.

[10] Potts CN, Kovalyov MY, "Scheduling with batching: a review," European Journal of Operational Research, 2000, 120, 228–49.

[11] Sevaux, M., Peres, S.D., "Genetic algorithms to minimize the weighted number of late jobs on a single machine," European Journal of Operational Research, 2003, 151, 296–306.

[12] Koksalan, M., Keha, A.B., "Using genetic algorithms for single machine bi-criteria scheduling problems," European Journal of Operational Research, 2003, 145, 543–556.

[13] Wang, C.S., Uzsoy, R., "A genetic algorithm to minimize maximum lateness on a batch processing machine," Computers and Operations Research, 2002, 29, 1621–1640.

[14] Li, Y., Ip, W.H., Wang, D.W., "Genetic algorithm approach to earliness and tardiness scheduling and planning problem," Intl. Journal of Production Economics, 1998, 54, 55–76.

[15] Reeves, C., "Genetic algorithms for the Operations Researcher," INFORMS Journal on Computing, 1997, 9, 231–250.

[16] Damodaran P., Srihari K., Lam S.S., "Scheduling a capacitated batch-processing machine to minimize makespan," Robotics and Computer-Integrated Manufacturing, 2006, 23, 208-216.

[17] Nearchou, A.C., "The effect of various operators on the genetic search for large scheduling problems," International Journal of Production Economics, 2004, 88, 191-203.

The Journal of Management and Engineering Integration
Vol. 1, No. 2, December, 2008

Six Sigma Scorecard: A Valid Methodology to Align Strategy and Performance Improvements to Satisfy Customers

Adriana Rodriguez
University of Central Florida, ab.rod7@gmail.com
Sandra Furterer
Wal-Mart Corporation, furterer@sbcglobal.net

Abstract

The alignment among strategy, performance, and customer is considered the key element to bring value to any organization. A case studied in a service organization, provided the opportunity to assess the effects that the proposed Six Sigma Scorecard (SSS) methodology had on productivity indicators and on the level of customer satisfaction (measured by cycle time, line capacity, number of complaints).

The research methodology used a Case Study/Action Research which described their merging process in three main places: Prioritization and selection of the Six Sigma Projects, Complement business Opportunities and Strategic Priorities and Relate Six Sigma and BSC measures.

Following Yin and Eisenhardt, during phase I, II and III of the Case Study/Action Research, the researcher followed the proposal theory and used triangulation in order to keep the quality of the research and pursuit validity in the research process.

The SSS methodology had proven successful in increasing the performance of the auto credit process measured by cycle time, process capacity, number of value added activities, and percentages of BSC targets reached, during the same period of time. There was evidence of alignment among strategy, improvement performance and customer satisfaction.

The SSS approach had proved to be flexible and adaptable to organizational needs, and strengthens the advantages of the BSC and Six Sigma approaches. The validation of the SSS methodology offers a new roadmap that can be implemented by Engineering Managers in any Organization to solve the strategy, performance and outputs misalignments.

Introduction

This paper describes the development, implementation, and effectiveness of a combination of two recognized management methodologies and tools – Six Sigma and Balanced Scorecard (BSC) – to align strategy and performance improvement in order to translate it to customer satisfaction.

Six Sigma is a quality management philosophy and methodology that focuses on reducing variation, measuring defects, and improving the quality of products, processes, and services. (Evans and Lindsay 2005; Hayes 2006).

On the other hand, the Balanced Scorecard is a strategic management tool and performance measurement system designed to directly translate the organizations' strategies into action-oriented plans. Intensive research has been conducted to assess the benefits associated with the applications of the BSC and Six Sigma in industry (Davis 2000; Sierra 2003; Andersen Henrik V 2004; Phillips 2004; Shu-Hsin Huang 2004; Paladino 2005).

However, the changing environments, forces, and threats that organizations are facing in this area have introduced some failures and called for new researches. (Gupta 2004) (Hayes 2006). (Jiju and Banuelas 2002; Jiju 2004; Pfeifer, Reissiger et al. 2004; Hayes 2006).

New gaps in the research are apparent and new approaches are needed in order to adapt and integrate the new tools available (Forrest 2003; Kubiak 2003; Andersen Henrik V 2004; Pfeifer Tilo 2004; Schultz 2006).

Proposed Six Sigma Scorecard Methodology (SSS)

The proposed methodology was built upon the success and failure stories of the Balanced Scorecard and Six Sigma implementation methodology. Merging these two methodologies, Balanced Scorecard and Six Sigma, promises an increase in productivity and customer satisfaction indicators of any organization.

The implementation of the proposed methodology starts with an assembled Balanced Scorecard and is structured on three primary merging points, which are represented in Exhibit 1.

First Merging Point: Selection of the Improvement Initiatives Projects

The objective of this step is to shift the organization's strategy to the tactical level. This step requires a deep understanding of the organization's background. Important elements to be considered are: organization's vision, mission and objectives, organizational culture and leadership, and internal and external drivers. The Balanced Scorecard, which is prepared by the executive and managers of the organization, needs to be studied in order to get the most out of it. This step is an evaluation process, when it is possible and desirable for the information to go back and forth between all levels of the organization.

Once the BSC is understood, the researcher requests a list of the principal improvement initiatives from each one of the business units involved on the SSS improve methodology.

The initiatives that have obtained more points will be counted as Six Sigma projects. Depending on the number of initiatives and the business units involved in the processes to be improved, a number of Six Sigma teams will be formed. The team must include the sponsor and project champion from the executive level of the organization, at least two Six Sigma specialists, and organization personnel involved in the day-to-day activities of the process selected for improvement.

The next steps follow the activities that need to be performed during the define phase of any

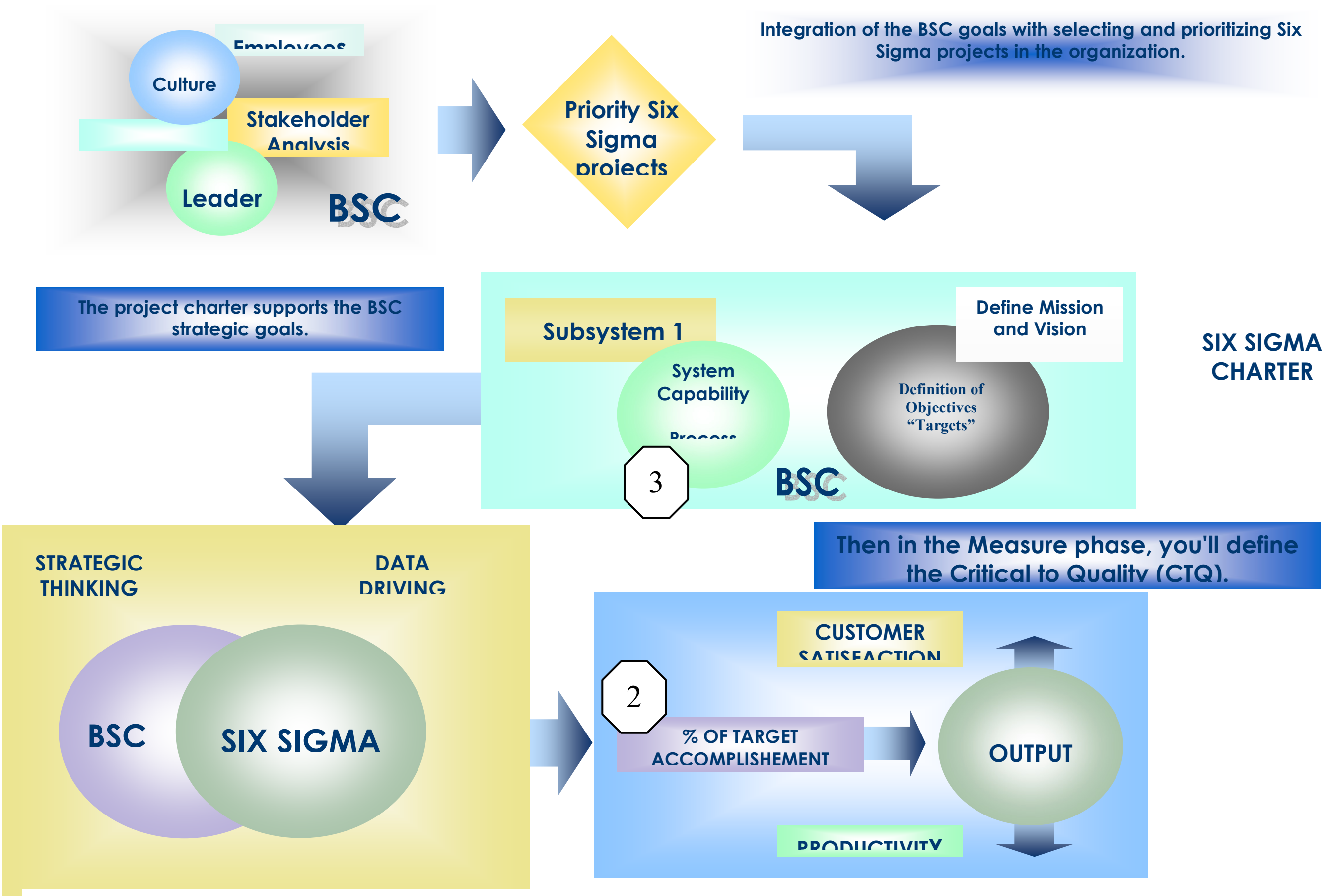


Exhibit 1: Merging Points.

Six Sigma project. The SSS team will start defining the project goals and objectives in order to prepare a charter, which represents the commitment of the organization and people involved to the improvement project. At this moment, the use of some quality and management tools, such as SIPOC diagram, shareholder analysis, stakeholder commitment, and flow charts, play an important role because they help the team to visualize the scope of the project as well as barriers and opportunities and make the appropriate planning.

Second Merging Point: Complement Business Opportunities and Strategic Priorities

The identification of the project's objectives is an integral part of the define phase for any Six Sigma project. The purpose of this second merging point assures that all Six Sigma project objectives target, directly or indirectly, the BSC objectives. In order to achieve that, a Matching Matrix needs to be created, with the BSC objectives in the first row and the Six Sigma project objectives in the first column.

Ideally, all of the SSS project objectives should match the BSC objectives. However, this is not always possible, and if not, the analysis and evaluation of both objectives must adhere to a strategy that avoids conflict and maximizes benefits. Flexibility and adaptation are critical elements for the implementation of the methodology. Deep understandings of the organization's background and situations that may affect the success of the project need to be studied in order to negotiate the most favorable context for the project. Adaptation to the real business world and the flow of information are essential to sustain the project during the measure and analyze phases of the Six Sigma Project.

Third Merging Point: Relate Six Sigma Indicators and BSC Measures

The third merging point occurs during the improvement phase of the SSS project. This point is the result of the two previous phases—measure and analyze. In the measure phase, the Critical to Quality (CTQ) characteristics are defined.

The team must identify and define the customer of the SSS project and list the characteristics of the services that are important for them or CTQ. The CTQ and the strategic objectives of the organization are analyzed by the use of tools such as the House of Quality, which is a six sigma tool, or by the use of a matching matrix. The purpose of this evaluation process is to balance the CTQ objectives and the Strategic objectives expressed on the BSC.

During the analyze phase, the root causes are identified as the factors (independent variables) that cause the problems. Finally, during the improvement phase, once the recommendations are implemented, such as changes on the productivity and customer satisfaction indicators, represented as cycle time, number of errors, and customer satisfaction index, should be compared to the changes on the BSC indicators and productivity indicators before the SSS implementation.

Research Design

The research design to be applied in the study of the development and application of the Six Sigma Scorecard on Institution Z was a convergence of the case study and action research methodologies.

Three conditions of this investigation point to the use of a case study as a research strategy. These conditions were the essence of the research questions, the fact that the theory refers to future conditions and lacks control over events. In addition, the exploratory nature of this case study implies a deep interaction between researchers, the elements of the phenomenon to be studied, and its embedded system. This constant feedback that places the researcher and phenomenon in a learning cycle is known as Action Research (Barton Cunningham 1993; Yin 1994; Yin 2003).

The research design was conducted in three major phases which are discussed as follows.

Phase I: Model and Concepts Design.

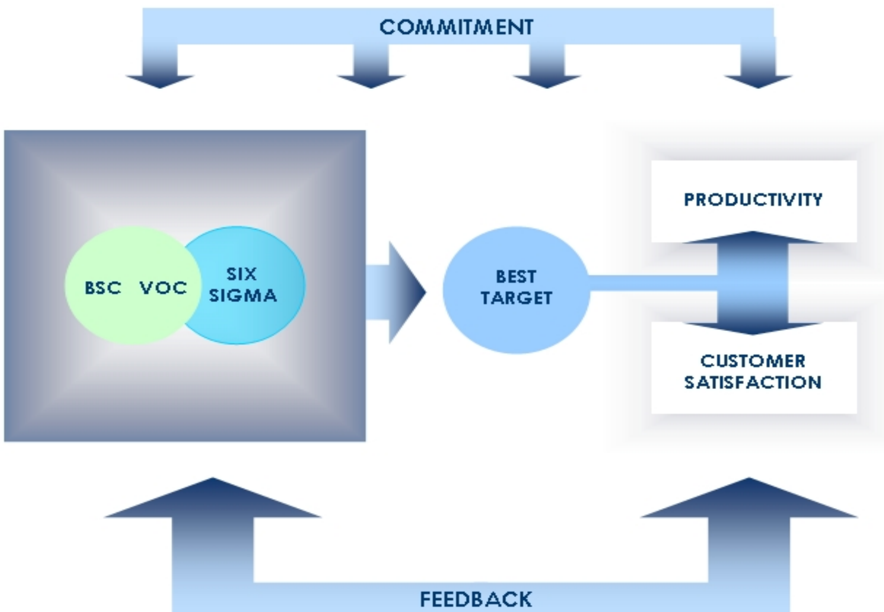
In Phase I, this research combined elements from the conceptual domain and the methodological domain applied in a substantive level to pursue the validity of the investigation.

The ideas that give meaning to the implementation of the SSS methodology were expressed by the conceptual model.

The value of the model will be assessed by the effects of the methodology on the productivity indicators, and the level of customer satisfaction. Exhibit 2 shows the SSS model where the merging methodology target productivity and customer satisfaction indicators.

Proposed Model: Six Sigma Scorecard (SSS).

Exhibit 2: Six Sigma Scorecard



The independent variables include the strategy-performance-customer link. The model explains that there was a need to integrate the organization's strategy, organization's performance, and organization's customer satisfaction index by merging the BSC and Six Sigma methodologies in three main places.

The dependent variables include productivity indicators and customer satisfaction indicators, measured by productivity metrics such as cycle time, production rates, production efficiencies, rework percentages, and so on.

The development of a formal case study protocol provides the reliability that is required of all research (Yin 1994; Tellis 1997). Although an identical replication was not possible, the use of the case study protocol as a guide during the research process provided evidence that a similar set of findings can be reproduced when the same pathway and the same set of elements, relations, and embedding system from each domain are used again.(Brinberg and Joseph E. 1985; Yin 1994).

Six Sigma Scorecard Validation

The proposed SSS methodology was implemented into the auto credit project. The framework described the three merging points identified in the methodology, as well as activities and tools that were needed to conduct the project. Numerous insights were generated from the SSS implementation. Four critical factors were determined to be keys to the success of the project:

Flexibility: The project life cycle was not considered a series of rigorous steps; instead, it was necessary go back to the measure and define phases from the analyze phase in order to facilitate the interaction between project resources and Institution Z resources.
Executive and Management Commitment: During this project, the executive and management commitment to the SSS project was the main guarantee that all steps in the process would be completed.
Internet Communication: During the Case Study/Action Although this kind of communication restrained the direct interaction with front-desk employees as well as limited interaction with middle-level managers, the researcher's knowledge in technology education and the advantages of the Internet in terms of managing time and distance allowed for continuous feedback, support, and the development of a secure record of data during the project.
Cultural knowledge for negotiation/barriers: In an environment where political issues drastically change the economic and external conditions that organizations face, it was necessary to use negotiation strategies in order to allow the project to run to its end.

The SSS improvement recommendations were implemented in two principal steps. The first recommended change was the migration to an information system, called ABANKS. The migration took place in October 2006, and the new process flow was implemented starting in January 2007. These implemented recommendations produced a significant increase in the average number of auto credit applications processed, from 67 loans during the first half of 2006 to 132 loans by March 2007. After the application of the SSS methodology, the auto credit process reached a 40% improvement rate on the approval cycle time by the end of May and expected an 80% improvement rate in the approval cycle time by October 2007. In addition to the productivity changes, major changes took place in order to eliminate rework and have savings that affected all of the organization such as the financial indicators of their BSC. The BSC targets show more than 70% reached in most indicators by the second half of 2006. The improvement in financial indicators used by Institution Z, called GIC (acronym in Spanish which meant Credit Intermediation Level), which measured the participation of the credit portfolio, served as an important sign of the level of alignment of the SSS target and the BSC indicators. At the beginning of the project, the GIC rate was low, below 2%; during March, there was an increase to 2.64%, and by August, it was at 3.51%. By the end of this research project, this BSC productivity indicator increased by around approximately 80%. The auto credit percentage rate in the GIC increased from 2.8% to 5.87% at the end of the project, which contributed to elevate Institution Z GIC to 40% as it was mandate by the BSC.

Conclusions

The literature does not provide a framework for implementing the proposed merger of BSC and Six Sigma methodologies herein named Six Sigma Scorecard, but there was potential for the SSS methodology. From the validated framework of the Six Sigma Scorecard methodology, this research was able to show that:

- Managers can move from the strategy to an improvement performance program that adds value for the customer.
- Managers will be able to evaluate performance and quality programs from an heuristic perspective, and this evaluation tool is important.
- There is evidence of alignment between strategy and improvement performance and customers.
- The use of the Six Sigma Scorecard improved performance levels.
- The use of SSS strengthens both the BSC and Six Sigma methodologies.
- SSS methodology was validated for a services organization.

The application of a Case Study/Action Research allowed an understanding of the phenomenon under study and its surrounding system. The multiple sources of evidence allowed hypothesizing the causal links of the Six Sigma Scorecard methodology and illustrated the basics of the three merging points between the two parent methodologies, which described the proposed "SSS" methodology.

References

[1] Andersen Henrik V, L. G. (2004). "Effective quality management through third-generation balanced scorecard." International Journal of Productivity and Performance Management **53**(7): 634-645.
[2] Barton Cunningham, J. (1993). Action Research and Organizational Development. Westport, CT, Praeger Publishers.
[3] Brinberg, D. and M. Joseph E. (1985). Validity and the Research Process. Newbury Park, California, Sage Publications, INC.
[4] Davis, S. B. (2000). An Investigation of the Development, Implementation, and Effectiveness of the Balanced Scorecard: A Field Study. Accountancy. Tuscaloosa, The University of Alabama**:** 96.
[5] Evans, J. and W. M. Lindsay (2005). The Management and Control of Quality. Ohio, Thomson South-Western.
[6] Forrest (2003). "Beyond six sigma lean:Methods to ensure your organization's health." ASQ **57**: 215.
[7] Gupta, P. (2004). Six Sigma Business Scorecard: Ensuring Performance for Profit. New York, McGraw-Hill Professional.
[8] Hayes, B. J. (2006). "Six Sigma Critical Success Factors." Retrieved Jan/23/2006, 2006, from www.isixsigma.com/library/content/c020415a.asp.
[9] Jiju, A. (2004). Some Pros and Cons of Six Sigma. The TQM Magazine. **16:** 303-306.
[10] Jiju, A. and R. Banuelas (2002). "Key Ingredients of Effective Implementation of Six Sigma Program." Measuring Business Excellence **6**(4): 20-27.
[11] Kubiak, T. (2003). "An Integrated Approach System." Quality Progress **36**(7): 41-45.
[12] Paladino, R. E. (2005). "Balanced forecasts drive value." strategic fianalcial **86**(7): 36.
[13] Pfeifer, T., W. Reissiger, et al. (2004). Integrating Six Sigma with Quality Management System. The TQM Magazine. **16:** 241-249.
[14] Pfeifer Tilo, R. W., Canales Claudia (2004). "Integrating six sigma with quality management system." The TQM Magazine **16**(4): 241-249.
[15] Phillips, J. K. (2004). "An Application of the Balanced Scorecard to Public Transit System Performance Assessment." Transportation Journal **43**(1): 26.
[16] Schultz, B. (2006). "Merging the Six Sigma and the Balance Scorecard." GE Retrieved 01/20/2006, 2006, from http://healthcare.isixsigma.com/library/content/c031028a.asp.
[17] Shu-Hsin Huang, P.-L. C., Ming-Chin Yang, Wen-Yin Chang, How-Jenn Lee (2004). "Using a Balance Scorecard to improve the Performance of an Emergency Department." Nursing Economics **22**(3): 140-146.
[18] Sierra, J. (2003). Quality Constraint Approach: A Six Sigma/Throughput Approach in Manufacturing to Achieve Company Success and Continuous Improvement. Industrial Engineering & Management System. Orlando, University of Central Florida.
[19] Tellis, W. (1997). "Applications of a Case Study Methodology." The Qualitative Report Retrieved 5/8/2006, 2006, from www.nova.edu/sss/QR/QR3-3/tellis2.html.
[20] Yin, R. K. (1994). Case Study Research. Design and Methods. California, Sage Publications Inc.
[21] Yin, R. K. (2003). Applications of Case Study Research. California, Sage Publications Inc.

The Journal of Management and Engineering Integration
Vol. 1, No. 2, December, 2008

A 3D Clustering Algorithm to Model Clutter in Electronic Geospatial Displays

Maura C. Lohrenz and Marlin L. Gendron
mlohrenz@nrlssc.navy.mil, mgendron@nrlssc.navy.mil
Naval Research Laboratory, Stennis Space Center

Abstract

Complex multivariate geospatial displays (e.g., electronic charts) can present unique challenges for users if the display becomes too cluttered. Clutter is exacerbated as more features are displayed. This paper presents a model of clutter, called the Color-Clustering Clutter (C3) model, as a function of two components: color density (or uniformity) and saliency. The C3 model is based on observations that geospatial displays with more uniform coloring appear less cluttered, regardless of the saliency of displayed features, while displays with less uniform coloring (i.e., greater color entropy) appear more cluttered when feature saliency is high, but not when saliency is low. We model color density by first clustering all pixels of an image in three dimensions – geospatial (x, y) location and color. The clustering method uses geospatial bitmaps to group pixels that are close in both geospatial proximity and color, bounds the clusters into polygons, and calculates density as the number of clustered pixels divided by the area of the polygon. We define *local* color density as the density of clustered pixels for a specific feature and *global* color density as the weighted average of densities for all features in the display. We estimate local and global saliency as a weighted average of color differences among adjacent clusters for a specific feature (local) or the entire display (global). Finally, we model local or global C3 clutter as a function of local or global color density and saliency.

1. Introduction

The Navy is implementing electronic charts throughout the surface fleet and has used moving-map displays in fighter aircraft for over a decade. As new and diverse sources of information become available, users are tempted to display everything of interest. High-resolution satellite, radar and other imagery can be overlaid with mission-specific routes and targets, intelligence data, etc. The resultant clutter in these displays introduces human factors issues concerning the ability of users to access, interpret and effectively use the displayed information. Previous studies have associated user performance with display complexity; e.g., clutter on an aircraft map display has been found to disrupt a pilot's visual attention, resulting in greater uncertainty concerning target locations [1],[2]. When a moving-map is scrolling at even a moderate rate of speed, such as on a large-scale cockpit display in a fighter jet, the detrimental effects of clutter are exacerbated.

While researchers have demonstrated a link between user performance and the presence of clutter (such as overcrowding of otherwise important information as well as unwanted data or noise), we still lack a reliable, quantifiable clutter metric that can be empirically tied to performance. This paper presents a model of clutter that can be empirically tested against both subjective evaluations of display clutter and user performance in a typical targeting task.

2. Our model of visual clutter

We theorize that our perception of clutter and our ability to locate items in complex geospatial displays are influenced by two main attributes: "saliency" and "color density", described below. Our C3 clutter model includes global and local components. Global clutter refers to the overall impression of clutter in a display, while local clutter refers to the amount of clutter immediately surrounding some feature of interest. The following sections discuss how we define and calculate the major components of our C3 model (saliency and color density) for both global and local clutter estimates.

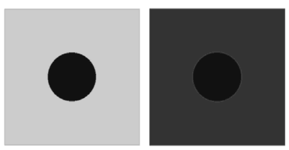


Figure 1. Examples of relatively high saliency (left) and low saliency (right).

2.1. Saliency

Saliency refers to how clearly one color or feature "pops out" from surrounding features in an image (e.g., figure 1). In general, highly salient features are more easily detected than features with low saliency. Here, we define a "feature" as any displayed geospatial item categorized as one of four primary types: points with associated symbols (e.g., airport symbols, city centers, etc.), lines (roads, rivers, etc.), areas (forested regions, metropolitan areas, etc.), or labels (country names, elevation labels, etc.).

Previous studies [3], [4] on clutter in complex displays focus primarily on the contribution of saliency to clutter. For our C3 clutter model, we estimate saliency as a weighted average of the color differences between adjacent features. For example, when one feature is completely surrounded by another (e.g., the black dot in the light gray field in figure 1), the saliency of the enclosed feature is the distance between these colors (light gray and black) in a perceptually uniform color space. Smaller color distances result in lower saliency: the dot in the darker gray field (figure 1) is less salient because the two colors are more similar and the distance between them is shorter.

We refer to the saliency of a single feature in an image as local saliency. We describe global saliency as the overall saliency of an image, calculated as the weighted average of all the local saliencies for all features in the image:

Saliency $= \Sigma\,(\mathbf{C_e}\,\mathbf{L_e}) / \Sigma\,(\mathbf{L_e})$, where

$\mathbf{C_e}$ = the color distance between adjacent features along their shared edge, **e**

$\mathbf{L_e}$ = the length of edge **e**

The equation for local saliency would only include edges between the feature of interest and each adjacent feature, while global saliency would include all edges bordering all features in the image.

2.2. Color density

Color density refers to how densely similarly-colored pixels are packed within the image, and can be thought of as the inverse of color variability or entropy. In general, we theorize that lower color density equates to higher clutter, when saliency is moderate-to-high. When saliency is very low, color density has little impact on clutter.

To calculate color density, we adapted an algorithm [5, 6] that clusters items lying within a preset distance from each other. The points are represented in a "geospatial bitmap" (figure 2a), such that each point to be clustered is represented by a "set" bit (bit value=1) and all other bits in the bitmap are "cleared" (0). The algorithm expands each set bit into a shape of predetermined dimensions. For example, in figure 2b, a set bit is expanded into a 3x3 bit square. All expanded bits that touch or overlap each other are clustered together via the formation of new bitmaps (figure 2d). The algorithm produces vertices for a bounding polygon, the density of which is calculated as the number of clustered points divided by the area of the bounding polygon (figure 2d).

For this project, we adapted the algorithm to cluster pixels in three-dimensional (3D) space, in which the third dimension is color. A separate clustering is performed for each color in the image; the density of each color is derived from all the densities of similarly-colored pixels within the resulting clusters. After clustering all pixels in the image into bounded polygons for a given "seed color" $\mathbf{s}$, a cluster density $\mathbf{D_P}$ is calculated for each cluster polygon $\mathbf{p}$:

$\mathbf{D_P} = \Sigma_{\mathbf{c}=1 \text{ to } \mathbf{n}} [(1 - \mathbf{E_C}/\mathbf{M})\mathbf{N_c}] / \mathbf{A_P}$, where

$\mathbf{n}$ = Number of different colors that clustered with pixels of color $\mathbf{s}$ in cluster polygon $\mathbf{p}$

$\mathbf{E_C}$ = Euclidean distance between colors $\mathbf{c}$ and $\mathbf{s}$ in the chosen color space

$\mathbf{M}$ = Maximum possible distance between any two colors in the color space

$\mathbf{N_C}$ = Number of pixels of color $\mathbf{c}$ in the cluster polygon

$\mathbf{A_P}$ = Area of cluster polygon $\mathbf{p}$

The color of each pixel in the cluster will be within a color distance of $\mathbf{z}$ from all immediately surrounding pixels in the cluster, starting with pixels of color $\mathbf{s}$. The color distance $\mathbf{z}$ is chosen to be a "just noticeable" difference in color. In other words, the cluster will "chain" similarly-colored pixels together to form the cluster, starting with each pixel of color $\mathbf{s}$ and subsequently including all other pixels within a geospatial distance of $\mathbf{x}$, $\mathbf{y}$ and a color distance of $\mathbf{z}$. If $\mathbf{z} = 0$, then $\mathbf{D_P} = \mathbf{N_S} / \mathbf{A_P}$ and only pixels of exactly color $\mathbf{s}$ are included.

The algorithm includes a decay function, such that the clustering of similarly colored pixels ends after a certain point. For example, in figure 3a, assume the seed color is white, each color is one pixel, and all color differences between adjacent, progressively darker pixels are <12, shown in table 1. Without a decay function, all the pixels in this figure would be clustered together if the user specified the clustering distances $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ to be {1, 1, 12}, even though the final pixel to be included (black) would be very different from the seed color (white).

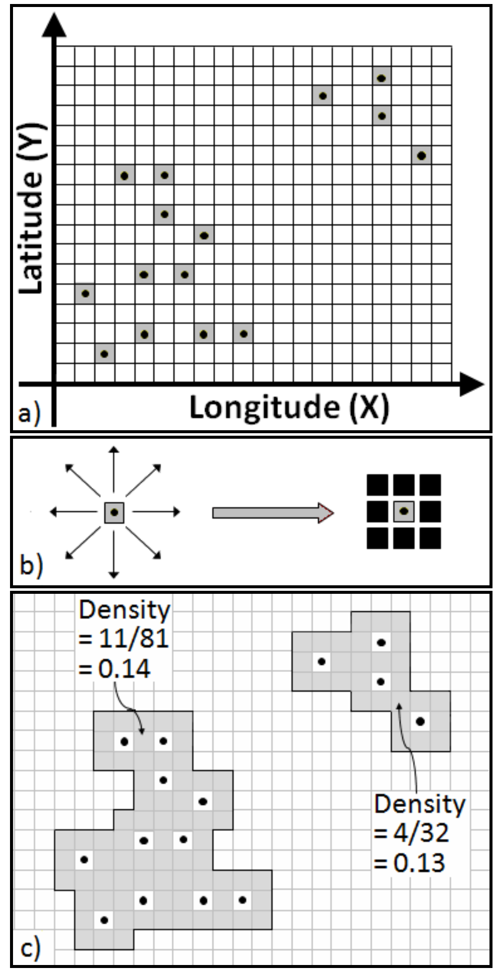


Figure 2. Using geospatial bitmaps to cluster points: a) representing points in a bitmap; b) expanding each set bit; c) forming clusters, calculating density.

The decay function limits the clustering of pixels with colors that progressively change along a subtle color gradient: a new pixel will be clustered with the previously clustered pixel only if the two pixels are geospatially close (within a distance of $\{\mathbf{x}, \mathbf{y}\}$) <u>and</u> if the difference between the new pixel's color and the previously clustered pixel's color is less than or equal to $\mathbf{z'}$:

$\mathbf{z'} = \text{round}(\mathbf{z}e^{-0.45\mathbf{r}})$ where

$\mathbf{z}$ = color distance specified for clustering ($\mathbf{z}$ = 12 in the above example)

$\mathbf{r}$ = color distance between the previously clustered pixel and the seed color.

For example, in figure 3a, if clustering starts with the white pixel (#0 in table 1), it then includes the lightest gray pixel (#1), since it is within the specified clustering distances of {**x**=1, **y**=1, **z**=12} from pixel #0. The next-darkest pixel (#2) is included because the color distance between it and pixel #1 is **z′**=2. Pixel #3 is not included, because the color distance between it and pixel #2 is 4, greater than the new **z′**=1.

Table 1. Color-clustering algorithm for figure 3. Descriptions for variables (x, y, c, r, and z') are provided in the text.

Pix #	Color	x	y	c	r	z'	Include if c ≤ z'
0	white	0	0	0			Y
1	lt. gray	0	1	4	0	12	Y
2	…	1	0	2	4	2	Y
3	…	0	1	4	6	1	N
4	…	0	1	Clustering stops after pixel #2 (for first cluster)			
5	…	1	0				
6	…	1	0				
7	dk. gray	0	1				
8	black	0	1				

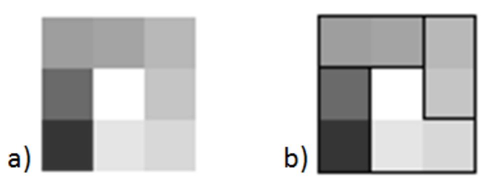


Figure 3. a) Original and b) clustered image.

At this point, the algorithm selects a new seed color from which to start clustering. Previously clustered pixels are not included in future clusters. Based on the distances in the above example, the final clustering (figure 3b) consists of four clusters with the following pixels: (0,1,2); (3,4); (5,6); (7,8).

2.3. Local and global clutter

We describe clutter as local (target-specific) or global (target-independent) and comprised of local or global saliency and density. Local and global saliency was defined in section 2.1. Local color density ($\mathbf{D_S}$) represents how much an individual seed color **s** contributes to the overall color density of the image. $\mathbf{D_S}$ is computed as the weighted average (by area) of the densities for all pixels clustered around color **s**:

$\mathbf{D_S} = \Sigma_{\mathbf{p}=1 \text{ to } \mathbf{n}} (\mathbf{D_P} \mathbf{A_P}) / \mathbf{A_S}$ where:
n = Number of clusters for seed color **s**
$\mathbf{D_P}$ = Density of cluster polygon **p**
$\mathbf{A_S}$ = Sum of areas of all clusters for color **s**

Global density ($\mathbf{D_i}$) estimates the overall color density for the entire displayed image **i** and is computed as the weighted average of local densities for all colors in the image:

$\mathbf{D_i} = \Sigma_{\mathbf{s}=1 \text{ to } \mathbf{n}} (\mathbf{D_S} \mathbf{A_S}) / \mathbf{A_i}$ where:
n = Number of different seed colors **s** for the entire displayed image
$\mathbf{D_S}$ = Local density for color **s** (above)
$\mathbf{A_i}$ = Sum of all $\mathbf{A_S}$'s for image **i**

When calculating global density for an image, we start by clustering around the most prevalent color, followed by the next most prevalent color, etc. To ensure that each pixel in the image is used only once during density calculations, any pixels that clustered around more prevalent colors are ignored during subsequent clustering sequences.

Figure 4 shows part of an aeronautical sectional chart of Iowa City, IO (left) and its associated color density plot (right). In the density plot, lighter areas represent higher color density (implying lower clutter), and darker areas represent lower color density (higher clutter). The original color figures were converted to grayscale for publication.

Figure 5 presents enlargements of the boxed feature (a numeral 6) in figure 4 to illustrate the results of our color clustering. The seed color is a medium blue (reproduced as medium gray). All pixels within preset distances from a seed pixel in both geospatial (here x=y=1) and color (z=12) coordinates cluster together. The resulting cluster is shown at right, with seed pixels depicted as squares, and similarly-colored pixels that clustered with the seeds as smaller circles. The density of this cluster (as defined in section 2.2) equals the weighted number of pixels included in the cluster (28.4) divided by the area of the cluster (69) = 41%.

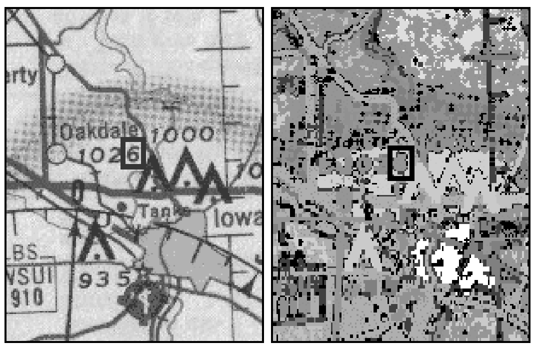


Figure 4. Sectional chart (left) and its color density plot (right). Lighter areas in the color density plot denote higher density (suggesting lower clutter).

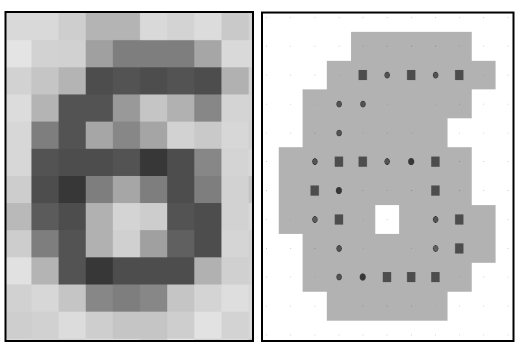


Figure 5. Results of 3D clustering around medium-blue "seed" colored pixels in the selected feature (zoomed-in from figure 4).

2.4. Interaction of saliency with density

As mentioned earlier, we theorize that both saliency and color density influence our perception of clutter in a graphic display. Very low color density suggests very high clutter when the features are highly salient (e.g., figure 1, left), but not if saliency is very low. Our C3 model of clutter is as follows:

$$15(1-\mathbf{D_I})*\exp[-6.3\exp(-\mathbf{S_I}/10)]-0.0002$$

$\mathbf{D_I}$ = Global density of image (section 2.3)

$\mathbf{S_I}$ = Global saliency of image (section 2.1)

The correction factor of -0.0002 ensures that the clutter measure for a blank screen is 0. The scaling factor of 15 is user-defined, to set the highest possible C3 clutter rating.

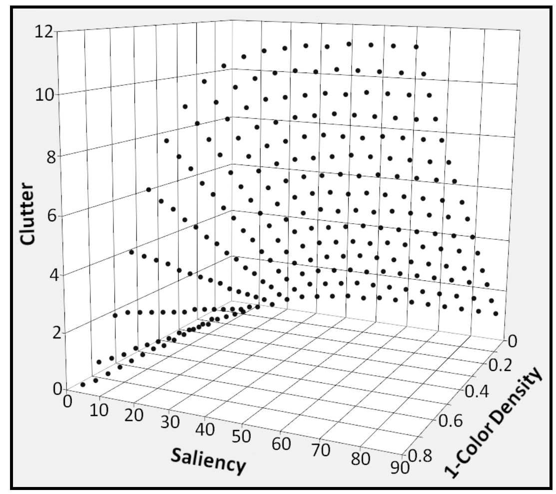


Figure 6. Clutter perception model as a function of saliency and color density.

Figure 6 illustrates how our model of perceived clutter varies as a function of these two variables. The impact of saliency on the model increases exponentially. E.g., for a saliency of 5 (lowest line of points), the greatest possible clutter rating is less than 0.5, regardless of color density. For saliencies greater than approximately 80 in our model, the resulting clutter ratings are influenced solely by color density.

2.5. Color difference formula

An appropriate color difference formula is crucial to our calculations of density and saliency. An ideal color space would be "perceptually uniform" such that a given distance between two colors anywhere in the space is perceived as a similar difference in color. Unfortunately, no color space can perfectly model human color perception. Figure 7 presents four color pairs, each of which is 10 units apart in the Commission International de L'Eclairage (CIE) L*a*b* color space, commonly accepted as more perceptually uniform than other color spaces. The bottom color in each pair is the same. If L*a*b* were perceptually uniform, each pair of colors would appear similarly different, but this is clearly not the case.

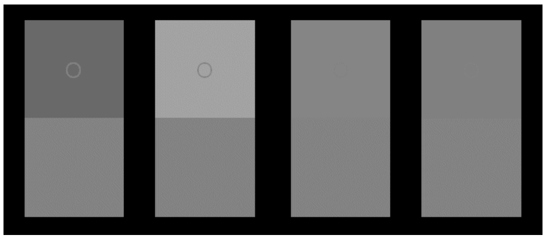


Figure 7. Four pairs of colors, each of which is 10 units apart in CIE L*a*b* color space.

A far more complex color difference formula, CIE d00, is based on the L*a*b* color space and intended to improve color difference calculations [7]. However, there are still limits to how well d00 mimics human color perception [8].

3. Discussion

We tested our C3 model with several color distance formulas, including d00, and compared resulting clutter estimates against subjective clutter ratings and target search times. A complete description of our experiments are pending publication [9], but in summary we found that in one experiment, C3 accounted for 58% of the variance in subjective clutter ratings for a wide variety of graphic displays ($r=0.76$, $p<0.0001$) and an astounding 74% of variance in ratings for a set of aeronautical chart displays ($r=0.86$, $p<0.0001$). In a target search experiment using the aeronautical charts, there was a main effect on target search time for C3 ($p<0.001$). These results suggest C3 is a good predictor of both subjective clutter perception and target search performance.

4. Acknowledgments

This work was sponsored under Program Element 002435N by the Naval Research Laboratory (NRL) 6.2 Base Program. We thank Geary Layne, Stephanie Myrick and Todd Lovitt for their valued assistance.

5. References

[1] V. Schons and C. Wickens, "Visual separation and information access in aircraft display layout", U. of Illinois Institute of Aviation Technical Report ARL-93-7/NASA-A3I-93-1, 1993.

[2] C. Wickens and C. Carswell, "The proximity compatibility principle: its psychological foundation and relevance to display design", Human Factors, 37(3), 1995, pp. 473–494.

[3] R. Rosenholtz, Y. Li, J. Mansfield, and Z. Jin, "Feature congestion: measure of display clutter", Proceedings of the Conference on Human Factors in Computing Systems, Portland OR, April 2005.

[4] M. Zuschlag, "Quantification of visual clutter using computational model of human perception", Proceedings of the HPSAA Conference, Daytona Beach, FL, March 2004.

[5] M. Gendron, M., G. Layne, M. Lohrenz, "A system, method and apparatus for clustering features", Navy patent 84,921, NRL, 2005.

[6] M. Gendron and M. Lohrenz, "A 2-D clustering algorithm to aid mine-warfare analysts in determining navigable lanes over cluttered seafloors", J. of Management and Engineering Integration, Winter 2009 (in press).

[7] Commission International de L'Eclairage (CIE), "Improvement to industrial colour-difference evaluation." CIE Publication No. 142-2001, Central Bureau of the CIE, Vienna, 2001.

[8] Sharma, G. (2003) "Digital Color Imaging Handbook." CRC Press. 797 pp.

[9] Lohrenz, M., J.G. Trafton, M. Beck, and M. Gendron. "A model of clutter for complex displays," submitted to *Human Factors.*

Success Factors for Change Management: The Roles of Leadership and Knowledge

Richard W. Monroe
Coastal Carolina University
RMonroe@coastal.edu

Leslie R. Pagliari
East Carolina University
PagliariL@ecu.edu

Abstract

Change is difficult to manage and successes are not as abundant as failures. This paper explores success factors for change efforts and develops a linkage between the group of success factors, change management and knowledge management. Change management leadership needs to incorporate explicit knowledge about these success factors.

1. Introduction

"The rate of change is not going to slow down anytime soon. If anything, competition in most industries will probably speed up even more in the next few decades."

— John P. Kotter
Leading Change [13]

"The information revolution is sweeping through our economy. No company can escape its effects. Dramatic reductions in the cost of obtaining, processing, and transmitting information are changing the way we do business" [15]. So began the article, "How information gives you competitive advantage," by Porter and Millar in Harvard Business Review in 1985. Today, as a result of the information revolution, knowledge management is a major issue for most companies attempting to deal with and utilize the vast amounts of information that they possess and process.

Change management must be an explicit consideration in conjunction with information technology and knowledge management. "After all, it is through projects and initiatives, however disjointed, that most significant change gets created" [7]. Some examples of how major change occurs and where change management is needed are: first, a major change initiative such as a Lean implementation or a Six Sigma initiative and second, a major information technology initiative. Table 1 lists some of these examples. In any of these situations, managing change and managing the knowledge associated with change are critical to ongoing success. Knowledge is created in these projects/initiatives and new knowledge is transferred over the course of projects and initiatives.

Table 1: Example Change Initiatives

Organizational Change Initiatives	Information Technology Change Initiatives
Operational Excellence	ERP implementation
Lean Enterprise	New Hardware/Software applications
Six Sigma	Software upgrades by vendor
Business process reengineering	In-house system/software revisions

With each of these specific initiatives there are documented and proven success factors. If we compare them there are a few similarities but the majority of the items differ from one program to the next. We will look in greater detail at several sets of success factors, make comparisons between the sets and also draw out the success factors that relate directly to change management and knowledge management.

"The brutal fact is that 70% of all change initiatives fail" [3][4]. This shocking statistic is an indication that there is a desperate need for better Change Management, better Knowledge Management, better Leadership and better use of known success factors.

2. Motivation for Change

We start with the motivation for change to better understand the circumstances surrounding change initiatives. Beer and Nohria [3] suggest that the motivation for change follows one of two theories, Theory E or Theory O. Theory E is based on economic value and is primarily focused on maximizing economic value for the shareholders of the company. Scott Paper is the prime example used by Beer and Nohria [3] to demonstrate Theory E in action. According to Al Dunlap who assumed the CEO position at Scott Paper, more than one objective (or more than one constituency) distracts management and prevents the company from achieving the only goal that is important [3, p.6]. After taking over the company in a troubled state, Dunlap turned around the financial performance with a return of 200% to shareholders within two years and his final act was to sell the company to one of its long time competitors, Kimberly Clark [3].

Theory O describes the motivation that is concerned with organizational capabilities over the long term [3]. Champion International is the example of choice for Theory O. CEO Andrew Sigler began a massive change initiative in 1981 which was aimed at making core cultural and behavioral changes in the organization including management, workers and the unions representing those workers [3, p.12]. It seems apparent that this approach is more concerned with long term survival of the firm rather than a short term financial payoff. The new values came to be called the Champion Way and included employee involvement in the efforts to improve the company along with a strong commitment to supporting the communities where Champion was located [3].

For the purposes of this paper we subscribe to the views of Theory O as being more consistent with the types of Change Initiatives that are of interest to management and engineering in our domain . While Theory E is dominant in the domain of holding companies and turnaround specialists, we cannot discount its importance as a starting approach which may be followed by the more positive approaches of Theory O.

3. Leadership and Knowledge

Knowledge about Change and Change Management is essential for successful change initiatives. Where does this knowledge reside? CEOs do not necessarily possess specific knowledge to lead major change initiatives [4]. A contributing factor seems to be executive turnover. CEO tenure decreased to 5.5 years during the 1990s compared to average tenure of 10.5 years when calculated in 1990 [4]. Schonberger [19] echoed a similar sentiment recently when he suggested that part of the problem (with the lack of sustained lean initiatives) was really an education problem. With the turnover among executives the successor does not possess the knowledge and does not hold the same commitment to the initiative of his predecessor [19].

That's not to say that individual executives are not knowledgeable about lean, six sigma or other change initiatives. "It seems clear that knowledge about best practice does not equal successful adoption. That is because a comprehensive, widely accepted theory of organizational change does not exist today. Knowledge is fragmented and piecemeal. Academics and consultants of different persuasions give dramatically different advice about how organizational change can be brought about. All this can be traced to the paucity of good research" [4, p.430].

Major consulting firms do possess the knowledge and the resources to mount significant change campaigns. Reverting to Theory E for a moment, for "rapid and extraordinary improvements in economic value, companies often hire large consulting firms in multimillion-dollar engagements to bring in the motivation and knowledge employees are thought to lack. In 1997 AT&T was reported to have spent $200 million on consultants. Armed with the latest analytic methods and a large cadre of very smart people, consulting firms often drive change by infusing the company with new ways to look at their business and new methods to manage it. Consulting firms help new CEOs who are uncertain of employee support and capability, and they offer important political cover to existing CEOs under fire from financial markets for the underperformance of their companies" [3, pp. 9-10].

Companies operating under Theory O may also resort to consultants to move the change process along quickly. The key with Theory O is transferring the knowledge to employees within the company in order to accomplish the desired long term results. This type of company cannot afford to allow the knowledge to walk away when the consultants leave.

Beer and Nohria [3] also suggest that using a combination of Theory E and Theory O may be the best approach for accomplishing change. It is also important to do them exactly in that sequence. Theory E is "hard" and Theory O is "soft". Clearly the switch from a "hard" approach to a "soft" approach is an easier transition. The move from "soft" to "hard" may be impossible without losing much of the good faith that has been created with employees. These observations suggest that executives do have active knowledge about Theory E and Theory O and that they are knowledgeable about success factors for change. Given the 70% failure rate mentioned earlier [3][4], there appears to be a lack of knowledge about the success factors for change. With this latter viewpoint in mind, in the following section we offer several examples of success factors associated with a range of change initiatives.

4. Leadership and Change

Peter Senge envisions a "community of leaders" which is comprised of leaders at many different levels of a "knowledge-creating" organization [20]. This idea is the ultimate in employee involvement and empowerment where every employee with leadership aspirations or ability steps up to the challenge of "spreading and fostering commitment to new ideas and practices" [20]. Specifically, the "community" includes "local line leaders, … executive leaders, … and internal networkers, …" because "… these three types of leaders absolutely rely on one another. None alone can create an environment that ensures continual innovation and diffusion of knowledge" [20].

To take Senge's community idea further and to counter the emphasis on top-down leadership we take another perspective from Bennis:

"In a society as complex and technologically sophisticated as ours, the most urgent projects require the coordinated contributions of many talented people working together. Whether the task is building a global business or discovering the mysteries of the human brain, it doesn't happen at the top; top-down leadership can't hope to accomplish it, however gifted the person at the top is. There are simply too many problems to be identified and solved, too many connections to be made" [5, p.114-115].

Bennis [2000] goes on to decry the celebrity of individuals and how, as a society, we heap praise and recognition on individuals rather than groups or teams. This leads to what should be obvious but still resonates as a powerful and surprising statement. "If there is one generalization we can make about leadership and change it is this: No change can occur without willing and committed followers" [5].

Still another take on leadership comes from Argyris [1] who questions the actions of many "change leaders" and even challenges the claims of highly regarded Stephen Covey. From his perspective, Argyris sees a major problem with the automatic responses that many "change leaders" may express which are really defensive in nature and actually very deceitful in many instances. A principled leader should communicate in an "open, forthright, authentic" manner [1]. If leaders are in fact resorting to a defensive posture or being deceitful then they are not "walking their talk" [1].

"Organizational defensive routines are antilearning and overprotective. These self-reinforcing antilearning routines are so powerful that they flourish even though they violate the basis of managerial stewardship and even though management education and human resources practices demand that they be reduced, if not eliminated.

Mixed messages are classic examples of organizational defensive routines. The theory in use is:

- State a message that is mixed (inconsistent).

- Act as if this is not so.
- Make the above undiscussable.
- Make the undiscussability also undiscussable" [1, p.421-422].

Argyris provides examples where change leaders, Covey and various change professionals in his workshops all exhibit "defensive routines. Either they would deny that they were producing them, or they would reflect and say they were unaware of them" [1].

"It is the reasoning of line management and change professionals that is soft. It is time that we educate ourselves not to use this defensive logic. If we do not, I predict that genuine changes in leadership, learning, change, and commitment will be limited to routine ones, even though they will be sold as producing double-loop change. At best the advice will have a short life; it will exist as a fad only [1, p.423].

From this discussion of Senge, Bennis and Argyris, we synthesize the following recommendation: change leaders are needed from all levels of the organizational 'community'; these leaders need to appeal to the organizational followers and nurture their attitudes to create a willingness and strong commitment for change efforts; and the leaders must foster the desired attitudes in the followers by being genuine, authentic stewards who are concerned for the benefit of the organization and the organization's stakeholders which naturally includes concern for the employees or followers. This combines the thoughts of the three authors and provides a summary recommendation for leaders of change initiatives.

Developing the followers is not an easy task so one last suggestion for appealing to followers comes from Martin [14]. "Change agents have four levers to pull: aspirations, insight, incentives, and learning capacity. …A robust model of change utilizes all four levers in coordination" [14]. Different people are motivated by different things. Martin's suggestion reinforces that idea and offers alternatives for leaders to develop the willingness and commitment among the followers which will be needed for successful change.

5. Knowledge about Successful Change

We offer "success factors" for change initiatives as a potential focus area in Knowledge Management. Explicit knowledge about these success factors both individually and collectively has the potential to crystallize change concepts and lead to greater success with change initiatives.

The CSSBB Primer [6] devotes several pages to the topic of Knowledge Management. This highlights the importance of KM as a crucial element within major change initiatives such as a Six Sigma program.

Kotnour [9][6] offers this definition of Knowledge Management – "The set of active processes that support a firm in creating, assimilating, disseminating, and applying its knowledge." Theory O is an opportunity to utilize the company's knowledge and the following statement makes a strong statement about the connection between KM and the organizational capabilities of Theory O. "Knowledge management is a process that a firm can use to build their capabilities in maintaining and improving performance" [9][6].

If we look at change on a grander scale. Kotter and Cohen [12] offer "the eight steps for successful Large-Scale Change" which include:

- Increase urgency
- Build the guiding team
- Get the vision right
- Communicate for buy-in
- Empower action
- Create short-term wins
- Don't let up
- Make change stick

These represent the steps that should lead to success but are not "success factors" per se.

Davenport, DeLong and Beers [7] take a different approach by suggesting that change can be accomplished through "knowledge management projects." Their research investigated 31 different knowledge management projects in 23 different companies. Ultimately they offered the following qualifying statements: "None of these projects is optimal. Some beg the question of whether it is really 'knowledge' that is being managed, and many are quite limited in their impacts. Very few contribute to the much touted goal of 'organizational transformation' " [7, p. 2]. It should be noted that the projects in the referenced research were software, hardware, and I.T. systems projects.

In the context of "knowledge projects" the authors offer a list of eight success factors which they identified in their research:

- Link to economic performance or industry value
- Technical and organizational infrastructure
- Standard, flexible knowledge structures
- A knowledge-friendly culture
- Clarity of purpose and language
- Different motivational practices
- Multiple Channels for knowledge transfer
- Senior management appreciation and support [7][8].

We also offer other published versions of success factors associated with other change initiatives in Table 2 (see Appendix at the end of the paper). These success factors are taken from Lean, Six Sigma and Business Process Engineering which represent some of the significant change initiative programs.

6. Role of Management and Engineering

"Your success in life isn't based on your ability to simply change. It is based on your ability to change faster than your competition, customers and business."

— Mark Sanborn [18]

Management and engineering play key leadership roles in the functional activities which are the focus of many change efforts. Today's business environment is characterized by a number of challenges for managers including "greater diversity; greater synchronization requirements; greater time pacing requirements; faster decision making, learning, and innovation; faster newness and obsolescence of knowledge; more frequent environmental discontinuities; faster industry life-cycles; greater risk of competency traps and faster newness and obsolescence of organizations. The challenges are being driven by the increased globalization of the knowledge economy and the increasing complexity of the systems" [11][2]. At the risk of oversimplification, the long list of challenges can be characterized as environmental change and the pace of change.

The Engineering Management field is one example where the interests of managers and engineers intersect or are closely aligned. Kotnour and Farr [11] also offer an initial model for the Engineering Management Body of Knowledge. The model consists of three levels: Life-Cycle Issues, Core Processes and Enabling Process and Tools (or Core Disciplines). The detailed elements include:

- Life-Cycle Issues: New Product/Technology Development; Value Chain Management; Production; Technology Marketing

- Core Processes: Strategic Management; Project/Program Management; Systems Engineering; Knowledge Management; Change Management

- Enabling Processes and Tools: Organizational and Workspace Design; Economics of Engineering; Quantitative Methods and Models; Quality Management; Developing Engineering Management Professionals [11].

Change and the need for Knowledge Management to assist Change are a common thread through all of the Life-Cycle Issues and have been explicitly stated as Core Processes.

We offer our own list of success factors which are synthesized from the other lists plus our own interpretation. This proposed list of Enabling Conditions has been assembled based on our judgment as to the most critical factors from the accumulation of the other lists.

The factors listed in Table 3 represent the collection taken from many different sources. "Top management support and active participation" is common across the majority of the references. "Make the Hard Decisions first" is our take on the recommendation from Beer and Nohria [3] to employ Theory E first. "Have an explicit Change Management Plan" is advocated by those most directly involved in change management rather than other initiatives [16]. "Have a deployment plan" is taken from Goldstein's recommendations [9]. "Have an explicit plan to transition to Theory O approaches" requires careful planning and precise timing to put into action at the appropriate stage of the change initiative as suggested by Beer and Nohria [3]. "Management by fact" is fundamental to Six Sigma [6] and is an excellent way to bring more objectivity into decision making. "Incorporate knowledge management techniques to institutionalize changes" [7][6] speaks to the issue of capturing knowledge and keeping it within the company especially in those instances where consultants have been a part of the change efforts. "Leverage technology to fuel and propel innovation and further change" [17] is a recommended approach to make the maximum use of the technologies that likely represent recent and substantial investments. This group represents the primary factors that we have synthesized from the literature that we feel offer the greatest potential to impact change initiative success.

Table 3 Enabling Conditions

Top Management Support and Active Participation
Make the Hard Decisions first
Have an explicit Change Management Plan
Have a deployment plan
Have an explicit plan to transition to Theory O approaches
Adhere to "management by fact"
Incorporate knowledge management techniques to institutionalize changes
Leverage technology to fuel and propel innovation and further change

The sampling of success factors adds an additional level that we label "Enabling Conditions." We suggest that these belong in an intermediate level between the Enabling Processes and Tools and the Core Processes. They form a direct linkage with Knowledge Management and Change Management.

Exhibit 1 (see Appendix at the end of the paper) is a modification to the model for the Engineering Management (EM) Body of Knowledge as proposed by Kotnour and Farr [11]. This particular model was chosen as a main intersection between management and engineering decisions.

As depicted in Exhibit 1, the "Enabling Conditions" consist of the success factors which we have discussed and synthesized to a smaller set of factors. These factors serve to improve and enhance the effectiveness of the "Enabling Processes and Tools" from the original model by Kotnour and Farr [11].

7. Conclusion

The salient issue in change management is achieving greater successes and fewer failures. To that end we have accumulated a broad range of success factors associated with a variety of change initiatives.

The main objective of the paper has been to document the success factors from a wide range of initiatives and publicize them in order to create a greater awareness that these factors need to be considered in conjunction with any type of change initiative.

Knowledge management is also tightly woven with change management and we assert that the success factors contained within this paper represent a directly linked set of "enabling conditions." These "enabling conditions" are an intermediate set of enablers that directly impact the 'Enabling Processes and Tools' which are needed to drive the 'Core Processes.' We offer a modified model of the Engineering Management Body of Knowledge in order to incorporate the success factors that make up the Enabling Conditions which we believe have the potential to fuel greater success with Change Initiatives.

Beyond the enabling conditions that we suggest there are also important considerations related to leadership. Primary among those considerations is: who will be the leaders of change and how will followers be motivated? In that regard, we offer a set of prescriptive recommendations for leaders to consider as they undertake change initiatives.

We propose that greater success with change initiatives will be the result if the enabling conditions are followed and put into action. The required leadership for change will be enhanced first by the enabling conditions and secondly by changing the leadership approach as we have prescribed.

8. Future Research

Future research might focus on finding companies that have made use of these success factors to alter the course of their change initiatives in a positive way. Documenting the real world experiences of such companies would provide supporting evidence for the effectiveness of these factors.

The topic here also offers the potential for Action Research if the researcher is involved with an industry partner and is assisting with any of the leading change programs where the success factors could be employed. That type of opportunity would provide an excellent case study and potentially a rigorous test of the success factors. Other related research topics are expected to emerge over time and as more researchers become involved in this research stream.

9. References

[1] Argyris, Chris. The relevance of actionable knowledge for breaking the code. In *Breaking the Code of Change*, edited by Michael Beer and Nitin Nohria, Harvard Business School Press: Boston, 2000, pp. 415-427.

[2] Barkema, H.G., J.A. Baum and E.A. Mannix, "Management Challenges in a New Time," *Academy of Management Journal,* 45:5, 2002, pp. 916-930.

[3] Beer, Michael and Nohria, Nitin, "Cracking the code of change," *Harvard Business Review*, May-June, 2000, pp. 133-141.

[4] Beer, Michael. Research that will break the code of change. In *Breaking the Code of Change*, edited by Michael Beer and Nitin Nohria, Harvard Business School Press: Boston, 2000, pp. 429-446.

[5] Bennis, Warren. Leadership of change. In *Breaking the Code of Change*, edited by Michael Beer and Nitin Nohria, Harvard Business School Press: Boston, 2000, pp.113-121 .

[6] Certified Six Sigma Black Belt Primer. Quality Council of Indiana, 2001.

[7] Davenport, Thomas, DeLong, David W. and Beers, Michael C. Building Successful Knowledge Management Projects. Center for Business Innovation Working Paper, Ernst & Young LLP, 1997.

[8] Davenport, Thomas, DeLong, David W. and Beers, Michael C. "Successful Knowledge Management Projects." *Sloan Management Review*, Volume 39, Issue 2, pp. 43-57, Winter 1998.

[9] Goldstein, Mark. Six Sigma Success Factors. *Six Sigma Forum Magazine*, Nov. 2001.

[10] Kotnour, Timothy. "Knowledge Management." In R.C. Dorg's, *The Technology Management Handbook.* Boca Raton, FL: CRC Press, 1999.

[11] Kotnour, Timothy and Farr, John V., "Engineering Management: Past, Present, and Future," *Engineering Management Journal*, 17:1, March, 2005, pp. 15-26.

[12] Kotter, John P. and Cohen, Dan S. *The Heart of Change*, Harvard Business School Press: Boston, 2002.

[13] Kotter, John P. *Leading Change*, Harvard Business School Press: Boston, 1996. Quote Source: <http://www.leadershipnow.com/changequotes.html>

[14] Martin, Roger. Breaking the code of change: Observations and critique. In *Breaking the Code of Change*, edited by Michael Beer and Nitin Nohria, Harvard Business School Press: Boston, 2000, pp. 449-471.

[15] Porter, M. and Millar, V. How information gives you competitive advantage. *Harvard Business Review*, 63, 4, 149-160, 1985

[16] Reengineering Success Factors: <(http://www.prosci.com/factors.htm)>

[17] Roadmap to Lean Success, Aberdeen Group, 2006.

[18] Sanborn, Mark. Quote Source: <http://www.leadershipnow.com/changequotes.html>

[19] Schonberger, Richard. Luncheon Speech based on *Let's Fix It*, ICAM Conference, Norfolk, Virgina, July 19, 2006.

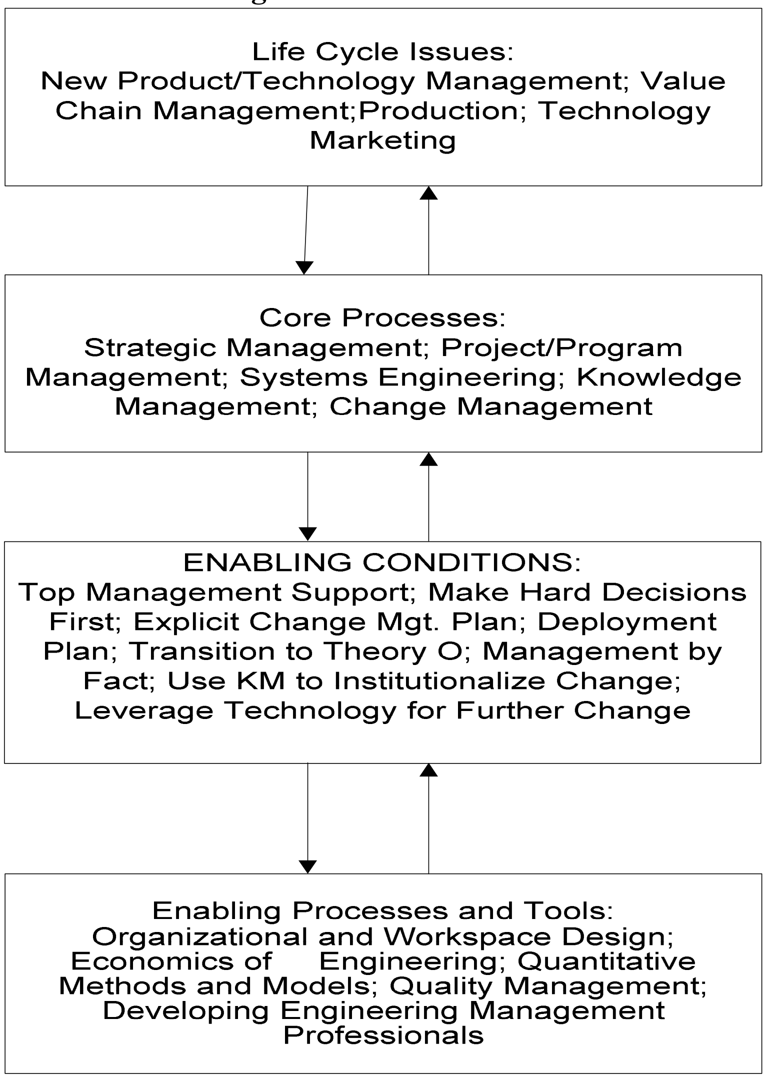
[20] Senge, Peter with Peter F. Drucker, Esther Dyson, Charles Handy, and Paul Saffo. Looking ahead: Implications of the present. *Harvard Business Review*, 75: 5, Sept-Oct, 1997, pp. 18-26.

APPENDIX

Table 2 Comparison of Change Initiative Success Factors

Reengineering Success Factors	Six Sigma Success Factors	Lean Success Factors
1. Top Management Sponsorship 2. Strategic Alignment 3. Compelling Business Case for Change 4. Proven Methodology 5. Effective Change Management 6. Line Ownership (accountability) 7. Reengineering Team Composition (both process experts and process novices) Source: [16]	1. Deployment Plan 2. Active Participation of Senior Executives 3. Project Reviews 4. Technical Support (Master BBs) 5. Full time vs. Part time resources 6. Training 7. Communications 8. Project Selection 9. Project Tracking 10. Incentive program 11. Safe environment 12. Supplier plan 13. Customer "WOWS" Source: [9]	1. Senior Management Involvement & support across the enterprise 2. Multiple value stream mappings across enterprise entities as a primary stepping stone toward properly defined measurement and control. 3. Maintaining a consistent and pervasive 'eye' on waste elimination and other non-value added activities that must be measured and chiseled away as the heart of continuous improvement. 4. Lean consultancy [and additional steps for 'Best in Class' lean companies] 5. Refined granularity of and accountability for defined measurements and controls. 6. Scale and coordinate Lean programs 7. Drive continuous improvement 8. Leverage technology to fuel innovation (as data is unlocked, put critical information in the hands of decision makers) Source: [17]

Exhibit 1 Enabling Conditions and E.M. Model



(adapted and expanded from Kotnour and Farr [11])

The Journal of Management and Engineering Integration
Vol. 1, No. 2, December, 2008

Development of a Quality System for Customer Satisfaction in a Financial Institution

R. Radharamanan
Mercer University
Macon, GA 31207
radharaman_r@mercer.edu

Angela P. Ansuj
Federal University of Santa Maria
Santa Maria (RS), Brazil
angelaansuj@yahoo.com

Abstract

In this paper, the factors that influence the functioning of financial institutions to properly satisfy the requirements of the customers for their diversified financial needs are highlighted and analyzed. One of the key issues identified from the customers was the excessive waiting time in the queue before they were served by the tellers. A multi-server system was analyzed to illustrate the importance of minimizing the waiting time in the queue. After estimating the arrival and service patterns, simulations were carried out to determine the average time spent by the customers in the system and in the queue.

1. Introduction

At present, the enterprises are forced to develop periodically new products/services to maintain their forefront of their market segment due to global competition. To accomplish this, the management system should return to quality in service, introduce innovations, and develop necessary human resources [1, 2].

Service quality is defined as providing internal and external customers with what they want. To accomplish this, a rigorous system of quality indicators must be developed [3]. Quality indicators are quantitative or qualitative measures associated with the level of service provided. Measurements such as customer and employee satisfaction, customer retention, employ turnover, order processing time, and on-time delivery should be considered when developing quality indicators [4].

The quality characteristics of the service organizations are intangible and different from that occur in manufacturing industries. Figure 1 shows the importance of the client within the service organization placing it in the central circle of the service triangle. The three critical elements service strategy, system, and person must act jointly for maintaining a minimum level of service quality [5].

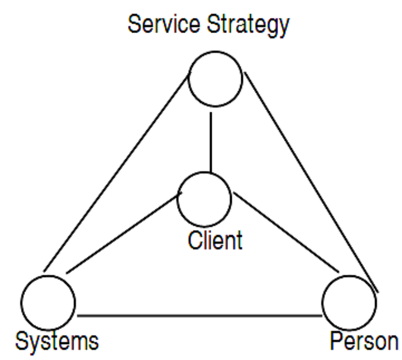


Figure 1. Service triangle

Managing finance for quality is another major issue in manufacturing and service organizations. Corporate finance must be integrated into the organization's total quality management efforts [6].

Most of the service companies fall into Type I or Type II organizations. In Type I, the management's objective is to increase sales volume and move on to the next customer as quickly as possible. This organization needs internal and external rework and several of its customers remain dissatisfied. In Type II, the resources are expanded on activities that are not adding to customer service, bureaucracy prevails, and the customers will probably take their business elsewhere, without announcing their departure. Most of the services in Type I and Type II organizations contain waste embedded in them such as waste of human effort, waste of machine time, and non-productive use of accompanying burden [4].

Type III organization use Total Quality Management (TQM) principles to meet their customer requirements by continuously improving and refining their operations. The ingredients that can make type III organizations successful are: policy deployment, employee empowerment, team work, training, communication, continuous improvement, and supplier performance [7].

The use of industrial models in service organizations leads to degradation in quality of service a company can provide. It sets in motion failures that are uniformly bad for customers, employees, shareholders, and the country. Some of its symptoms are: customer disaffection, high employee turnover, flat or falling sales, and little or no growth in productivity [8].

In this paper, some of the quality problems associated with financial organizations are highlighted and discussed. As a case study, a multi-server system was analyzed to determine the expected waiting time of the customers in a Brazilian financial institution. Also, simulations were carried out to determine the average time spent by the customers in the queue as well as in the system, after estimating the arrival and service patterns.

2. Quality in financial organizations

The financial organizations acknowledge that the management knows less about their customers than the management in any other business. At present, the coordination of pulling together the accounts of any one customer – his/her checking account, savings accounts, fiduciary accounts, trust accounts, loans, etc. - is greatly simplified by modern data processing machines. But this is far short of knowing the needs of the customers, and to what extent the financial institution fails to meet these needs. Why does a customer of the bank get a loan elsewhere for the purchase of an automobile or to buy a home or to remodel his present home? Neither the facts nor the reasons are on the record. Some kind of customer research may answer this question.

The financial institutions indicate that from 40% to 60% of any staff in such institutions is involved in verifying the work of other staff members. Statistical methods can help to reduce the frequency of mistakes. Program of improvement can be instituted at any institution regardless of size, for they can be tailored to meet specific needs and designed to grow in new areas of application. A human operator, a machine or a system, is monitored over a period of time to determine its process capability. The process capability of an operation can usually be determined in about three months. If the process capability is not within the acceptable limits, then the management must do something about the process/system. Quality can not exceed the capability of the system. Quality cannot be inspected into a product or service but it must be built into it [9].

Some of the areas that need attention for continual studies in financial institutions are: rejection rate of checks processed through high speed sorters, maintenance and downtime of machines, cost of handling exceptions, vendor performance measures such as periodic review of checks from outside printers for errors, time elapsed between receipt of request from customer and action taken, number of customers waiting in the line, distribution of time for transaction by tellers, turnover rate of tellers, error rate of tellers, number of checks or commercial notes returned because of errors in completion, downtime of computers, error rate associated with money transfer, number of past due accounts, number of new accounts opened, number of calls made to solicit new accounts, and number of classified loans.

3. Queuing models

In general, the financial institutions operate as single queue multi-server systems. Simulation is often used as a device to model the queuing system as well as to verify analytical solutions.

3.1. Single queue multi-server system

For the single queue multi-server system, considering that the arrivals follow Poisson distribution with mean arrival rate, λ and the service follows exponential distribution with mean service rate, μ with s number of servers in the system, the steady state requirements are [10]:

Probability that the system is empty:

$$P_0 = 1/[\sum_{n=0}^{s-1}\{(\lambda/\mu)^n/n!\} + \{(\lambda/\mu)^s/s!\}(1/1-(\lambda/s\mu)] \quad (1)$$

Expected number of people in the queue:

$$L_Q = P_0[(\lambda/\mu)^{s+1}/\{(s-1)!(s-\lambda/\mu)^2\}] \quad (2)$$

Expected number of people in the system:

$$L = L_Q + \lambda/\mu \quad (3)$$

Expected waiting time in the queue:

$$W_Q = L_Q/\lambda \quad (4)$$

Expected waiting time in the system:

$$W = W_Q + 1/\mu \quad (5)$$

Equations (1) through (5) are used to determine the steady state parameters for a single queue multi-server system.

3.2. Simulation

Simulation is intuitively appealing to a client because it mimics what happens in a real system or what is perceived for a system that is in the design stage. The output data from a simulation should directly correspond to the outputs that could be recorded from the real system. Additionally, it is possible to develop a simulation model of a system without dubious assumptions [11]. In this paper, simulation is used as a device to model a queuing system of a major financial institution as well as to verify analytical solutions.

4. Case study

The Brazilian financial institutions are either Type I or Type II. The use of quality indicators is not common in these organizations. Total quality management tools are not being used to improve the quality of the services provided to customers. Their major concern is the customers' dissatisfaction on waiting time in the queue before the services are provided by the tellers. Scheduling of tellers during peak and non-peak hours is important to minimize the number of people in the queue and reduce the waiting time of the customers in the queue.

A small branch of a major financial institution (Caixa Economica Federal) in Brazil was observed for quality using TQM concepts and in particular for its customer satisfaction. Most of the days, during peak hours, a long queue was observed inside the bank, sometimes even outside the building for paying bills, deposits, and withdrawals. The study showed that the distribution of arrivals followed Poisson distribution with arrival rate λ = 2 per minute. That is, on the average, a person arrived every 30 seconds. The mean time taken to serve a customer by a teller was 3 minutes. The customers formed a single queue and were served by the first available teller. The problem was solved using a single queue multi-server system. The steady state parameters were determined using equations (1) through (5).

Another branch of the same financial institution was observed for its general and business accounts. It showed interest in simulating the arrival and service patterns of its clients. It was known from the past data that 1/3 of its clients had business account and 2/3 of them had general account. The bank had four tellers, two of them specialized in business account and the other two specialized in general account. The clients arrived at the bank at a rate of one every 3 ± 1 minute. Clients randomly chose between the two tellers available for each type of account. Business accounts took 15 ± 10 minutes to complete the service. The general accounts normally took 6 ± 5 minutes. For the purpose of simulation, it was assumed that the customer chose a line without regard to its length and did not change lines. On the average about 60 transactions were handled on a given day. The available data provided the arrival and service distributions of both general and business customers. An appropriate simulation model was designed and run for the queue system consisting of general accounts and business accounts customers.

5. Results and discussions

The results obtained are presented in two sub-headings: (i) single queue multi-server system and (ii) simulation model. In both cases the expected waiting time of the customers in the queue as well as in the system were determined and analyzed. Using the simulation model, the utilization of the tellers serving different types of customers was determined and analyzed.

5.1. Single queue multi-server model

Under the steady state conditions: $\lambda/\mu = 2/(1/3) = 6$. This indicates that a minimum of seven tellers are required during the peak hours to server the customers ($\lambda < s\mu$). The steady state parameters obtained for the single queue multi-server system considering 7, 8, or 9 servers are presented in Table 1. It is clear from Table 1, that increasing the number of servers from 7 to 8 significantly reduces the average waiting time of the customers in the queue (from 1.8415 min. to 0.5355 min.). Also, it is possible to reduce the average service time by properly training the new tellers and improving their efficiency. When the queue builds up quickly due to fluctuations in the arrival rate on certain days in a month, it is recommended to increase the number of servers to 8 or even to 9 thus reducing significantly the waiting time of the customers in the queue during peak hours.

Table 1. Single queue multi-server system

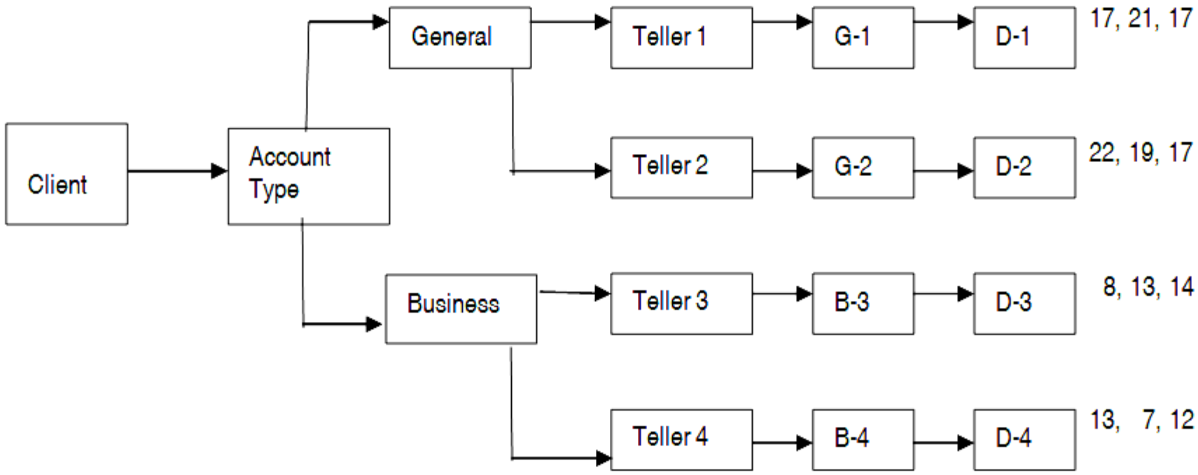
Steady State Parameters	Number of Servers		
	s = 7	s = 8	s = 9
$P_o \times 10^{-3}$	1.5788	2.1428	2.3523
L_q	3.6830	1.0709	0.3930
L	9.6830	7.0709	6.3920
W_q (in min)	1.8415	0.5355	0.1960
W (in min)	4.8415	3.5355	3.1960

5.2. Simulation model

Simulations were carried out for a three day period using the above data. The block diagram for the simulation model is shown in Figure 2. The results obtained are presented in Table 2. It is seen from the results that the average utilization for the general accounts tellers 1 and 2 are 0.51 and 0.53 respectively indicating that they have almost 50% ideal time on a given day whereas the average utilization of the business accounts tellers 3 and 4 are 0.91 and 0.94 respectively indicating that they have less than 10% ideal time on a given day.

The average waiting time of a general account customer in the queue is 9.4 minutes whereas that of a business customer is 29.81 minutes. The average waiting time of a general account customer in the system is 16.51 minutes and that of a business account customer is 46.38 minutes.

The average service time of the general account customer is 7.11 minutes which is within 6 ± 5 minutes. Similarly the average service time of the business account customer is 16.57 minutes which is again within 15 ± 10 minutes.



G-1, G-2: General accounts tellers 1 and 2 respectively;
B-3, B-4: Business accounts tellers 3 and 4 respectively;
D-1 through D-4: Dispose.

Figure. 2. Block diagram for the simulation model

Table 2. Single queue multi-server system

Teller	Customer Served: Days			3-Day Average	Utilization	[1/μ]	W_Q	W
	1	2	3					
G-1	17	21	17	18.33	0.51	7.11	9.40	16.51
G-2	22	19	17	19.33	0.53			
B-3	8	13	14	11.67	0.91	16.57	29.81	46.38
B-4	13	7	12	10.67	0.94			

$1/\mu$ = Average service time in minutes; W_Q = Average waiting time in the queue in minutes;
W = Average time spent by the customer in the system in minutes.

To minimize the waiting time of the business accounts customers, as well as to increase the utilization of the general accounts tellers, it is recommended to cross train the general accounts tellers in business accounts and allow them to attend the business accounts customers as and when they are idle. This also will provide additional idle time for the business accounts tellers.

6. Conclusions

The single queue multi-server system and the simulation model presented, analyzed, and discussed in this paper clearly indicate that one of the major problems in the Brazilian financial institutions is excessive waiting time of both general and business customers.

The multi-server model showed that there is a need for a minimum of 7 tellers during the peak hours on a given day. If the queue length increases due to fluctuations in the arrival rate, or if the institution wants to reduce the waiting time of the customers, use of 8 or 9 tellers during the peak hours is recommended.

The simulation model showed under utilization of the general accounts tellers. It is recommended that the general accounts tellers are also trained to deal with the business accounts. This will help the institution to make use of the general accounts tellers more effectively and to reduce the burden on the business accounts tellers as well as reduce the waiting time in the queue of the business accounts customers.

7. References

[1] J. M. Juran, Planning for Quality, A Division of Macmillan, Inc., New York, 1988.

[2] J. M. Juran and A. B. Godfrey, Juran's Quality Handbook, Mc-Graw Hill, 5th Edition Section 33, 1999.

[3] J. Lobes, Encantanto o Cliente, Externo e Interno, Editora Humburg, São Paulo, Brazil, 1993.

[4] B. M. Sheridan, "Changing Service Quality in America", Quality Progress, December 1993, pp. 97-99.

[5] K. Albercht, Revolução nos Serviços, Editora Pioneira, São Paulo, Brazil, 1992.

[6] J. A. F. Stoner and F. M. Werner, Managing Finance for Quality: Bottom-Line Results from Top-Level Commitment, ASQC, Quality Press Publications, 1994.

[7] R. Hestand, "Measuring the Level of Service Quality", Quality Progress, September 1991, p. 57.

[8] L. A. Schlesinger and J. L. Heskett, "The Service Driven Service Company", Harvard Business Review, September-October 1991, p. 71.

[9] E. W. Deming, Out of the Crisis, MIT Center for Advanced Engineering Study, Cambridge, MA, 1986.

[10] F. J. Gould, G. D. Eppen and C. P. Schmidt, Introductory Management Science, Prentice Hall, 3rd Edition, Englewood Cliffs, New Jersey, 1991.

[11] J. Banks, J. S. Carson II, B. L. Nelson and D. M. Nicol, Discrete-Event System Simulation, Prentice Hall, 3rd Edition, 2001.

The Journal of Management and Engineering Integration
Vol. 1, No. 2, December, 2008

A Case Study on Results-Oriented Graphic User Interface Design for Handheld Devices

Christian Sonnenberg
Harris Corporation
csonnenb@harris.com

S. Ann Becker
Florida Institute of Technology
abecker@fit.edu

Abstract

The user interface design standard, *What You See is What You Get* (WYSIWYG), has weaknesses when applied to today's emerging technologies. Results-Oriented design is increasingly used to address these weaknesses in the development of more usable software applications. However, little research has been conducted on Results-Oriented design for software applications targeting handheld devices. This paper initiates research in this area by studying the usability of Results-Oriented design for a PocketPC software application. The results of this initial case study are summarized and future research opportunities identified.

1. Introduction

For many years, human-computer interfaces (HCI's) have been designed using the generally accepted *What You See is What You Get* (WYSIWYG) approach. WYSIWYG is a design approach that provides direct function mapping to design elements in a software application. It has been the standard for designing software applications for many years [1].

The "Results-Oriented" approach, an HCI design spearheaded by Microsoft Corporation, has emerged as a viable alternative to WYSIWYG. It focuses on providing the user preconfigured functionality supported by more intuitive HCI's. Microsoft's Vista software product utilizes the "Results-Oriented" approach. This supports the industry claim that Results-Oriented design is becoming a new standard for HCI development.

The handheld device, inclusive of a PocketPC, has been largely ignored in terms of human-computer interfaces that meet the needs of end users while maximizing portability. Touch screen capability, tiny screen size, and miniaturized screen objects reflect the added complexity of designing handheld devices that are usable in a range of physical environments. Since the introduction of handheld devices known as Personal Digital Assistants (PDAs), WYSIWYG has been the standard design approach to software application development. As such, many of the HCI components associated with PDAs have the same look and feel as personal computers HCI components.

To date, the authors are unaware of any studies related to WYSIWYG-designed handheld applications specifically targeting PocketPC devices. Yet, there are practitioner-known drawbacks with WYSIWYG applications for the PocketPC. This study focuses on the application of the Results-Oriented approach to the design of a Pocket PC interface. The Results-Oriented approach holds promise in the design of more intuitive, simple, and easy-to-use applications. The study uses Microsoft Office Word 2003 and 2007 applications to identify user preferences in WSIWYG interface designs when compared to Results-Oriented interface designs. Initial results are presented along with future research directions.

2. Background

WYSIWYG takes a bottom up approach when it comes to developing the application feature set. At its core, WYSIWYG provides simple commands that must be combined into features to achieve the final product [2]. With increasing numbers of commands, users are easily overwhelmed and unclear as to what combinations create the desired results.

Some alternatives to WYSIWYG design have been used over the years and are still used today. Structural Oriented Design [3], in the form of markup language editors, has been around for years even before the creation of WYSIWYG. These editors provide users precise and exact control over what output should be. However, this approach negates the primary concern of WYSIWYG, to allow users the ability to see just what they are doing. This approach is often referred to as, *What You See is What You Ask For* (WYSIWYA).

Editors, such as LaTex, continue using markup language and have maintained popularity among research and technical fields. HyperText Markup Language (HTML) editors also implement markup language interfaces, as precise control over underlying document structure is necessary for Web page creation. Nevertheless, the knowledge required to operate such editors is often too great for average users when compared to WYSIWYG [4].

The Results-Oriented approach to design focuses on providing live feedback and template structures to the user. Referred to as, *What You Get Is What You See* (WYGIWYS), the paradigm emphasizes fast results in fewer steps performed using an intuitive and simple design approach [2]. Results-Oriented has as goals a reduction in time in achieving desired results and a minimization in time required to find functional features. Results-Oriented design focuses on providing an intuitive navigational schema utilizing both contextual menus and workflow organization.

Unlike WYSIWYG, WYSWYA, and popular editors, the Results-Oriented paradigm is based on the idea of visual representation of desired output before it becomes permanent. Menus in Results-Oriented design, for example, contain "live preview" features that show changes in real-time to a document before final changes are made.

2.1. PocketPC design

WYSIWYG is considered by many developers as the only option for mobile platforms. Handheld devices in the past have had limited processing power. Until recently, handheld devices had restricted functionality thus requiring a simplified HCI. As such, WYSIWYG has been the ideal approach to HCI design for handheld devices [5].

The PocketPC increasingly has expanded functionality that goes beyond note-taking and scheduling capabilities. Today, there is a plethora of software applications that run on a PocketPC ranging from simple games to sophisticated tax software and full-featured word processing and graphics applications. WYSIWYG design drawbacks become more apparent as application complexity increases.

Microsoft's Pocket Word is a WYSIWYG application that forces the user to create a document from scratch. The user must set margins, choose formatting options, add footer and header data, and perform other preliminary steps. The constraints of screen, keyboard and object sizes, compound the issue of features being nested deep in menu structures making them difficult to find. The design interface of the handheld device also relies heavily on pop-up dialogs largely due to the lack of screen space. Pop-up dialogs often lead to more pop-ups dialogs with a "next" or "continue" button. The user may become lost in the complexity of the navigational schema of chained dialogs.

Adding complexity to the discussion, in the comparison of WYSIWYG and Results-Oriented design, are the physical constraints of a PocketPC. The standard screen size measures 240 by 320 pixels. The display is flat and two-dimensional with low resolution graphics. A stylus pen is used to manipulate objects and a tiny, internal keyboard measures 240 by 80 pixels. Screen protectors typically cover the display area thus increasing the glare of surrounding light [6]. Memory is limited to roughly 100 megabytes of hard drive memory and 32 to 256 megabytes of RAM. Backlighting, a light source for the screen display, shuts off frequently to conserve battery power resulting in a constant dimming and lighting of the screen.

3. Case study

The overall objective of this study is to initiate research in the comparison of HCI design approaches for usable PocketPC applications. For the purpose of this study, Microsoft Word products are used to compare WYSIWIG and Results-Oriented design approaches. Microsoft's Pocket Word was used as the WYSIWYG software application. To date, there is no Pocket Word 2007 version available as a PocketPC software application. Hence, a prototype version was designed using Microsoft Embedded C++.

According to Jakob Nielsen, five subjects represent an adequate test space (refer to [7] for an overview of this justification). As such, five subjects were recruited for the case study. The subject pool consisted of experienced male computer users ranging in age range from 23 to 30 years. Each was familiar with the use of a handheld device, though none had prior experience using a word processor on a handheld device. Each was familiar with the use of a stylus pen and the use of the internal keyboard for text entry.

The usability attributes associated with usability testing of WYSIWIG and Results-Oriented software applications included: Learnability, Repeatability, Navigational Complexity, Error Rate, and User Satisfaction.

- *Learnability* is the amount of time it takes to complete a task with no prior knowledge of the application. This attribute was measured through user scenarios mimicking real-life situations.
- *Repeatability* is the ability to perform identical or related operations quickly. It was measured in the time taken to repeat similar consecutive tasks.
- *Navigational complexity* is the overall number of paths a user must take to reach a destination. It was measured in the number of clicks to accomplish a task.
- *Error rate* is the number of times a user strays from a directed path unintentionally. It was measured as the numbers of incorrect steps tracked by observation.
- *User satisfaction* is a qualitative metric that determines a user's satisfaction with the application. Users answered questions for this study related to each of the two Word applications.

Users had no prior experience with either Pocket Word or the Results-Oriented Word thus promoting a sound comparison. The applications first used were alternated between Pocket Word and Results-Oriented Word in order to average the effects of prior knowledge of one application before testing the other.

3.1. Learnability

Table 1 shows the results of the learnability component of the usability test. For Part 1 in Table 1, whereby a simple function was performed, Results-Oriented outperformed WYSWIG with an average of 48.2 seconds required for task completion compared to an average of 112 seconds for Pocket Word task completion Subject data points for task completion varied more greatly for Pocket Word. This is reflected in its higher standard deviation when compared to Results-Oriented Word.

For Part 2, the subject was required to perform several additional functions and multiple repetitive tasks. The Part 2 task completion time for the Results-Oriented Word was on average 275.8 seconds compared to Pocket Word on average being 495.6 seconds.

Table 1. Learnability (time/secs)

	Pocket Word		Results-Oriented	
	Part 1	**Part 2**	**Part 1**	**Part 2**
User 1	94	527	45	289
User 2	119	439	45	268
User 3	105	465	48	260
User 4	132	554	56	299
User 5	110	493	47	263
Average	112	495.6	48.2	275.8
Std Dev	14.37	46.21	4.54	17.22

3.2. Error rate

Errors were defined as actions the subject performed that are not associated with the test script. Table 2 shows the Results-Oriented Word to have on average, a smaller number of errors per operation and a smaller standard deviation of errors made than Pocket Word. User 1, having the highest number of errors using Pocket Word, cited problems with dialog boxes. In fact, most of the errors in Pocket Word involved closing dialog boxes. Subjects using both Word versions closed the application in error by inadvertently tapping on the exit button feature.

Table 2. Error rates

	Pocket Word	Results-Oriented Word
User 1	5	2
User 2	2	1
User 3	3	0
User 4	4	3
User 5	2	1
Average	3.2	1.4
Std Dev	1.30	1.14

3.3. User satisfaction

The results of the user satisfaction test component were gathered through questionnaires provided at the end of scripted tasks. All subjects note that they preferred Results-Oriented Word Processor to Pocket Word, citing "ease of use" and "immediate feedback" among their reasons. Several subjects noted, however, that drawbacks of the Pocket PC platform as a whole limited the usefulness of both applications.

Responses to, "The application was easy to learn", resulted in Results-Oriented Word having a much higher satisfaction rating than Pocket Word. These results correspond with results from the learnability test component. Responses to, "It was easy to find buttons I needed", Results-Oriented Word had significantly higher ratings than Pocket Word. Subjects responded to "Organization of information is clear" with higher satisfaction ratings for Results-Oriented Word than Pocket Word.

A productivity statement produced similar satisfaction results, as subjects felt that both applications "did what was needed to be done in a reasonable amount of time". The "ease of use" and "comfort" statements did not produce significant differences between the two Word applications. There is an indication that while Results-Oriented Word provided a lower learning curve, the subjects' familiarity with the WYSIWYG paradigm may have had an influence on user perception.

3.4. Repeatability

Each subject was asked to perform the same series of steps iteratively as part of the repeatability test component. Time was measured for each step during the use of each Word application. An iteration consisted of two similar tasks, thus providing a repeating sequence of operations. A shorter time to completion meant that a task was more readily repeatable.

The results showed that the Results-Oriented Word task had a much faster performance rate on subsequent iterations. Both applications required roughly the same time for the initial setup; however, additional operations using Results-Oriented Word took almost half the amount of time as when using Pocket Word. The average time to complete a series of steps for Pocket Word was 14.8 seconds; whereas, the average time for Results-Oriented Word was 6 seconds. Figure 1 shows the performance differences, measured in seconds, for each of the Word applications.

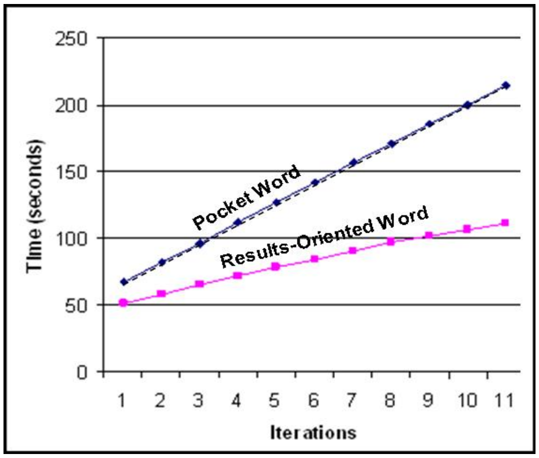


Figure 1. Repeatability test results

3.5. Navigational complexity

Navigational complexity was measured by the number of clicks for each step in the completion of a task. Results-Oriented Word maintained roughly half as many required clicks than that of Pocket Word. This may be attributed to Results-Oriented Word having all functionality within a one-click reach; whereas, Pocket Word requires multiple clicks to find, open, and close a dialog box.

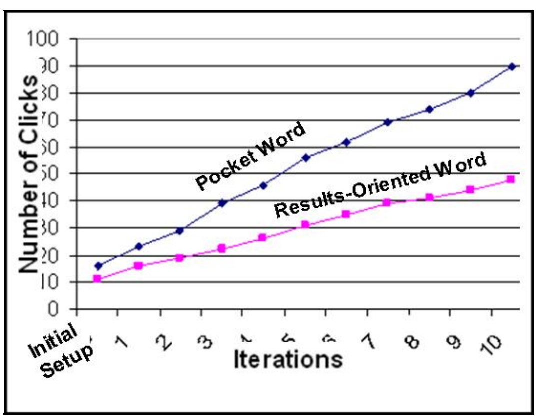


Figure 2.Navigational Complexity

The average number of clicks for each of ten Pocket Word iterations was 7.4 clicks; whereas, the average number for Results-Oriented Word was 3.7 clicks. The graph in Figure 2 further illustrates the disparity between the applications. These findings are similar to the results obtained for the repeatability metric.

4. Data analysis

Though the study results are preliminary, observations are made to identify future research directions. There are significant opportunities for expanding this research taking into account: larger sample size; subject diversity in terms of age, gender, and handheld experience; and environmental factors such as light levels.

It appeared that Results-Oriented Word had a much shorter learning curve Pocket Word in the successful completion of tasks. Subjects knowledgeable with the WYSI-WYG paradigm often performed tasks more efficiently when using Results-Oriented Word. Further research is needed to assess these usability findings taking into account task complexity.

The study results showed that subjects found buttons faster, spent less time hunting and clicking, and navigated more intuitively using Results-Oriented Word than Pocket Word. In fact, task completion time for Pocket Word was nearly double that of Results-Oriented Word. This may be somewhat attributed to Pocket Word using nested dialogs and pop-ups, which take time to load and close. Results-Oriented Word never redirects from the main dialog saving several seconds in traversal time. Additional research is needed to determine the impact of Results-Oriented features on performance time associated with user completion of tasks (e.g., hunting for a feature) and application processing (e.g., closing a dialog box).

Further research is needed to address perceived drawbacks to using the Results-Oriented design approach. It was observed that a major drawback to using Results-Oriented Word was the computational power required to produce "live previews" (e.g., real-time simulation of how a user selection is displayed). In addition, all actions are taken from the same page, thus the main window will occasionally flicker or lag behind user operations. The time spent waiting for an update may significantly offset user performance gains on simpler navigational schemas and intuitive feature use. Further research is needed to study these trade-offs from a usability perspective.

5. Conclusion

The results of this preliminary study indicate that a Results-Oriented approach holds much promise for more usable handheld interfaces. Though further research is needed, this study suggests that Results-Oriented design may provide significant efficiencies in the use of high-productivity applications such as spreadsheets, mail clients, picture manipulation, and task managers. The study also suggests that for simple, single-function applications, the efficiencies from a Results-Oriented approach may be negligible. This too requires further research in terms of application complexity and usability factors.

Due to the high profile nature of Microsoft products, it is a likely that Results-Oriented design will grow in popularity. It is becoming widely accepted that Results-Oriented design offers a valid alternative to the WYSIWYG standard. As such, further research on its use particularly in emerging technologies is warranted to promote usability by a broad user group.

The handheld device in the past has been often overlooked in usability testing associated with complex software applications. There has been research conducted on usability aspects of mobile devices (e.g., Brewster's 2002 [8] research screen space and Copas 2004 [9] work on data input), but little has been done in studying the use of more sophisticated applications. There are extensive research opportunities in this field, as both mobile and handheld devices become increasingly ubiquitous in the virtual world of global technology.

6. References

[1] Ahmed, H. (2006). What is WYSIWYG?. Retrieved March 9, 2008, from http://hossamahmed.wordpress.com/2006/09/13/what-is-wysiwyg/.

[2] Nielsen, J. (2005). R.I.P. WYSIWYG. Jakob Nielsen's Alertbox. Retrieved March 8, 2008, from www.useit.com/ alertbox/wysiwyg.html.

[3] Tidwell, J. (2005). *Designing Interfaces.* O'Reilly Publishing, Sebastopol, CA.

[4] Pemberton, S. (2000). So much for WYSIWYG, *Reflections*, *7*(5), Association of Computing Machinery, pp. 60 –ff.

[5] Stone, B. R. (2001). Handheld communication interfaces: design, development, and process. Ohio State University. Retrieved March 8, 2008, http://www.idsa.org/webmodules/articles/articlefiles/Brian_Stone.pdf.

[6] Weiss, S. (2002). *Handheld Usability*. John Wiley & Sons, Inc., New York, NY.

[7] .Nielsen, J. (2000). Why you only need to test With 5 users. Jakob Nielsen's Alertbox, October. Retrieved March 8, 2008, from http://www.useit.com/alertbox/20000319.html.

[8] Brewster, S. (2002). Overcoming the lack of screen space on mobile computers. *Personal and Ubiquitous Computing*, *6*(3), pp. 188-205.

[9] Copas, G. (2004). Comparing data input methods on handheld computers. *Usability News 6.2,* Wichita State University. Retrieved on March 9, 2008 from http://psychology.wichita.edu/surl/usabilitynews/62/HHC.htm .