

사용자 인터페이스 디자인에 있어서 사용성 개선지수 개발에 관한 연구*

A Study on the Development of Usability Improvement Index in User Interface Design

박남춘**†

Nam-Choon Park**†

서울여자대학교 미술대학 산업디자인학과**

Department of Industrial Design, Seoul Women's University**

Abstract

Because of the development of digital technologies and information-oriented products, the importance of user interface design and usability are increasing in recent year. So we need a systematic and rational way to measure and manage how effective the improvement of user interface is. The purpose of this study is to develop measurable 'Usability Improvement Index' which signify how much usability has improved in user interface design. First, I made the framework of measuring and evaluating usability through analysis of existing usability metrics and previous studies. Then, I suggested the process of calculating 'Usability Improvement Index' based on categorizing characteristics of measurable usability and definition of usability metrics. The process is as in the following : measuring the present level of usability, analysis of problems, making alternative user interface design, measuring the improved level of usability, calculating Usability Improvement Index. For verification, the case study on application of Usability Improvement Index to usability improvement of application software of digital camcorder presented. This study is expected to assist 'goal oriented user interface design' and to manage usability systematically.

Keywords : User Interface Design, Usability, Usability Evaluation, Usability Improvement Index

요약

디지털 기술의 발전과 제품의 정보화로 인해 최근 사용자 인터페이스 디자인의 중요성이 더욱 더 증대되고 있다. 따라서 사용자 인터페이스의 개선이 어느정도 효과가 있는지를 체계적이고 합리적인 방법으로 측정하고 관리하는 것이 필요하다. 본 연구는 사용자 인터페이스 디자인에 있어서 사용성이 얼마나 개선되었는지를 나타내는 측정 가능한 '사용성 개선지수'를 개발하는 것을 목적으로 한다. 먼저 기존의 사용성 평가방법과 선행 연구에 대한 분석을 통해 사용성 개선지수 개발을 위한 사용성 측정 및 평가방법에 대한 틀을 마련하였다. 다음으로는 측정 가능한 사용성의 특성을 분류, 정의하고 이를 바탕으로 사용성 개선지수 산출 프로세스를 제안하였다. 현재의 사용성 수준을 측정하고 문제점을 분석하여 사용성을 개선한 대안을 제시하고, 개선된 사용성 수준을 측정하여 개선전과 개선 후의 사용성 수준을 비교할 수 있게 하였다. 개발된 사용성 개선지수는 실제 제품의 사용성평가와 개선작업을 사례연구로 진행하여 그 적용 가능성을 확인하였다. 본 연구는 목표에 의한 UI디자인을 가능하게 하고 사용성을 체계적이고 합리적인 방법으로 개선하고 관리하는데 도움을 줄 수 있을 것으로 기대한다.

주제어 : 사용자 인터페이스 디자인, 사용성, 사용성 평가, 사용성 개선지수

* 이 논문은 2008년 정부(교육과학기술부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임(KRF-2008-332-G00035)

† 교신저자 : 박남춘(서울여자대학교 미술대학 산업디자인학과)

E-mail : ncpark@swu.ac.kr

TEL : 02-970-5720

FAX : 02-970-5985

1. 서론

1.1. 연구배경 및 목적

사용자 인터페이스 디자인(User Interface Design : 이하 UI디자인이라고 함)에서 사용성(Usability)은 어떤 제품의 UI를 평가하는데 있어서 가장 기본이 되는 제품의 특성으로서, 사용자가 제품을 어떻게 다루는가와 얼마만큼 사용하기 쉬운가를 나타내는 척도가 된다. Nielsen, J. (2003)을 비롯한 많은 사용성 전문가들은 휴리스틱 평가(Heuristic Evaluation) 및 사용성 테스트(Usability Test) 등 사용자 인터페이스를 향상시키기 위한 다양한 방법론을 제시하였고, 여러 가지 후속 연구와 사례를 통해 사용자 인터페이스를 디자인함에 있어서 사용성을 평가하고 개선하는 것이 매우 중요하다는 것이 입증되었다.

그러나 사용자 인터페이스를 개선하기 위한 이러한 방법론적인 노력에 비해 사용성이 어느정도 개선되었는지 그에 따른 UI디자인의 품질은 어떻게 높아졌는지를 체계적이고 합리적인 방법으로 측정하고 지속적으로 관리하는데 있어서는 어려움이 따른다. 이는 UI디자인의 정성적인 특성상 UI디자인의 품질을 총체적으로 측정하거나 평가하기 쉽지 않기 때문이다. 사용자 인터페이스의 개선정도를 평가하기 어려우니, UI디자인의 품질을 개선했다 하더라도 지속적으로 관리하기조차 쉽지 않게 된다.

또한 기업에서의 디자인품질은 이제 제품개발 프로세스에서 최우선적으로 고려해야 하며 지속적으로 관리해야 하는 전략적 경영자원으로 받아들여지고 있다. 그러나 UI디자인은 그 중요성과 파급효과에도 불구하고 아직까지 제품개발 프로세스상에서 전략적 역할이 크지 않다. 이는 개별적인 제품개발 프로세스에서는 그 역할을 기대할 수 있으나, 기업의 경영자원으로서 관리할 수 있는 체계를 가지고 있지 못하기 때문이다. 체계적으로 관리를 하려면 그 수준을 측정할 수 있어야 하는데, 사용성은 유일하게 측정 가능한 성격을 가지고 있다. 따라서 UI디자인을 총체적으로 평가하고 관리하지는 못하더라도 가장 기본적인 특성인 사용성의 현재 수준을 파악하고 목표수준을 설정할 수 있으며, 개선 정도를 평가하여 지속적으로 향상시킬 수 있는 관리체계를 갖출 수 있도록 하는 노력이 필요하다.

이에 본 연구는 명백한 목표수준을 정의할 수 있는

측정 가능한 사용성 개선지수를 개발함으로써, 사용성 평가의 결과를 객관적인 지표로 나타낼 수 있도록 하고, 사용성 지수를 활용한 평가와 개선작업을 포함하는 사용성 평가 프로세스를 적용하여 개별적인 태스크의 개선을 비교하는 것이 아니라 제품의 사용자 인터페이스 전반적인 사용성 수준의 개선을 비교할 수 있도록 하는 것을 목적으로 한다.

1.2. 연구의 내용 및 방법

먼저 사용자 인터페이스 디자인에 있어서 사용성 평가 방법과 선행연구에 대한 탐색과 분석을 실시하였다. 이를 통해 사용성 개선지수 개발을 위한 새로운 사용성 측정 및 평가 방법에 대한 틀을 마련하였다. 다음으로는 사용성 개선지수를 개발하는 데 있어 측정 가능한 사용성의 특성을 분류, 정의하고 정량적/정성적, 객관적/주관적 측정 방법에 대한 조사와 분석을 통해 사용성의 측정과 평가에 대한 기본 방향을 설정하였다. 이를 바탕으로 사용성 개선 지수 산출 프로세스를 제안하였으며, 사용성 테스트를 통해 현재의 사용성 수준을 측정하고 문제점을 분석하여 목표수준을 설정한 다음, 사용성 개선을 위한 디자인을 통해 개선안을 만들고 개선 후의 사용성 수준을 측정하였다. 이러한 결과를 사용성 개선지수로 나타내어 개선전과 비교할 수 있도록 하였다. 개발된 사용성 개선지수는 실제 제품의 사용성평가와 개선작업을 사례연구로 진행하여 그 적용 가능성을 확인하였다.

2. 사용성의 측정과 평가

2.1. 수행(Performance)측정

수행측정은 사용성평가에서 가장 많이 사용되는 방법으로, 제품에 대한 효율성을 평가할 수 있는 방법이다. 사용자들이 많은 에러를 범한다는 것은 제품이 개선의 여지가 있다는 것을 말하며, 태스크를 완료하는데 예상보다 시간이 많이 걸린다면 효율성이 개선되어야 하는 것이다. 이렇듯 수행측정은 사용자들이 실제로 제품을 어떻게 사용하는지 알아볼 수 있는 가장 좋은 방법이다(Tom Tullis & Bill Albert, 2008). 일반적으로 과업분석(Task Analysis)을 통해 사용자가 어떠한 정보를 필요로 하고 또한 어떠한 문제점이 발생할지 여지가 있는지를 파악한 후에 특정한 태스크에 대한

측정이 이루어지는데, 사용자들이 실제로 주어진 과업(Task)을 수행하게 된다. 따라서 수행측정은 사용자의 사고과정보다는 실제 행동을 관찰하는데 주안점을 두며 과업 수행시간, 오류의 수 등과 같은 보다 정량적인 측정이 주로 이루어진다(이건표, 1996). 이러한 수행 측정치에 대해서 Laura M. Leventhal과 Julie A. Barnes(2007)는 정확도(accuracy), 에러 수, 태스크 시간, 특정 인터페이스의 빈도 수를 정량적(Quantitative)인 수행측정치로, 조작의 패턴과 에러의 패턴을 정성적(Qualitative) 수행측정치로 정의하였고, Tullis와 Albert(2008)는 태스크 성공(task success), 태스크 시간(time on-task), 에러(errors), 효율성(efficiency), 학습용이성(learnability)의 다섯 가지로 정의하였다. 그 이외에도 다양한 학자들이 수행측정치에 대해서 여러 가지 분류와 정의를 하였지만, 일반적으로 사용자가 과제를 마칠때까지 걸리는 시간과 오류를 범한 회수, 과제를 완료했는지에 대한 성공률의 세가지 공통된 측정치를 제시하고 있다.

그러나 수행 측정만으로는 온전한 사용성평가를 하였다고 판단하기 힘들다. 태스크 성공 또는 완료시간을 측정할 때 데이터 뒤에 놓여있는 이슈들이 간과될 수 있기 때문이다. 수행측정 데이터는 어떤 태스크가 문제가 되는 인터페이스를 가지고 있는지 지적할 수 있지만, 그것이 왜 문제인지, 어떻게 고칠 수 있는지를 더 잘 이해하기 위해서는 대부분 관찰이나 주관적 측정치를 통한 보강이 필요하다.

2.2. 자가기록(Self-Reported) 측정

제품의 사용성에 대한 주관적 측정치로서 자가기록 방식의 측정방법이 많이 개발되어져 왔다. 사용자의 주관적 의견을 통해 제품에 대한 사용자의 이해와 인터랙션에 대한 중요한 정보를 얻을 수 있다. 또한 감성적인 수준에서 데이터는 사용자들이 시스템에 대해서 어떻게 느꼈는지를 말해주고, 사용자의 이러한 반응에 주목하여 사용성 문제를 분석할 수 있다(Tom Tullis & Bill Albert, 2008).

이러한 자가기록 측정은 주로 각 태스크의 마지막이나 전체 사용성 테스트 세션의 마지막에 이루어지며, 상황에 따라서 직접 구두로 데이터를 수집하거나 서면조사 또는 온라인 조사를 통해 데이터를 수집할 수도 있다. 많이 사용되는 고전적인 척도로는 긍정적인 표현에서 부정적인 표현까지 3~7단계의 연속 척도

상의 점수로 나타내는 리커트 척도와 반대되는 한쌍의 형용사 표현을 척도의 각 끝에 두는 의미분별척도법(Semantic Differential : SD법)이 있다.¹⁾

Lewis(1992)는 사용자가 태스크나 시나리오를 완료한 후에 쓰는 사후 시나리오 질문(After-Scenario Questionnaire, ASQ)이라는 평가척도를 개발하였다. 태스크를 쉽게 완료할 수 있었는지, 걸린 시간에 만족하였는지, 지원된 정보에 만족하였는지 등의 질문에 7점 척도로 답하도록 디자인되었는데, 유효성, 효율성, 만족도라는 사용성의 근본적인 영역을 모두 포함하고 있다.

보다 사용자의 주관적인 만족도에 집중한 자가기록 측정방법으로는 QUIS(Questionnaire for User Interface Satisfaction)이 있는데, 사용자에게 대한 전체 반응, 스크린, 용어와 시스템 정보, 학습, 시스템 능력이라는 5가지 카테고리 내에서 27개의 평가항목으로 구성되어 있으며 10점 척도로 이루어져 있다(Chin, Diehl, Norman, 1998).

이러한 자가기록 측정시에는 사용자가 제품을 사용하면서 느낀 주관적인 평가를 모두 개별 평가척도로 점수화할 수는 없기 때문에 주관적인 코멘트를 포함하는 것이 일반적이다. 때로는 이러한 주관적인 코멘트들이 사용성의 개선 방법을 확인하는데 매우 유용하게 쓰일 수 있기 때문이다.

2.3. 행태(Behaviors) 측정

사용성 테스트를 진행하는 동안 참여자들은 태스크를 완료하고 설문지를 채우는 것 외에도 다양한 행동을 하는데, 이러한 행태적인 특성들도 관찰하고 기록할 수 있다. 일반적으로 사용성 테스트 참여자의 외적 행동은 언어적 행동과 비언어적 행동으로 나눌 수 있다(Tom Tullis & Bill Albert, 2008). 언어적 행동은 참여자들이 실제로 소리내어 말하는 것(Think Aloud)이며 비언어적 행동은 이를 제외한 것들을 포함한다. 두 가지 모두 인터페이스에서 사용자에게 문제를 일으키거나 만족스러운 측면을 확인하는데 유용하다.

언어적 행동은 참여자가 제품을 사용하는 동안 감성과 멘탈 상태에 대한 가치있는 정보를 제공한다. 참여자들이 테스트 도중 하는 말중에 어떤 코멘트는 부

1) 미국의 심리학자 Charles Egerton Osgood이 창안한 방법으로, 특정의 개념이나 기호가 지니는 정동적 의미를 객관적이고 정량적으로 측정하는 방법

정적이고 어떤 코멘트는 긍정적이며, 또는 중립적이거나 해석하기 어려운 것도 있다. 이러한 언어적 행동과 관련된 의미있는 측정치는 부정적인 코멘트와 긍정적인 코멘트의 비율이다. UI디자인 개선 후에 얼마나 부정적인 코멘트가 감소하였는가를 판단한다거나 다른 제품간의 비교를 할 때 유용하게 사용할 수 있다.

비언어적 행동은 제품에 대한 참여자의 경험을 잘 드러낼 수 있다. 얼굴표정이나 신체행동 또는 시선의 움직임 등을 통해서 참여자의 습관이나 제품에 대한 반응을 확인할 수 있는 것이다. 일반적으로 사용성 테스트시에 참여자의 얼굴을 비디오로 기록하는데, 이는 참여자가 태스크를 수행할 때 참여자의 얼굴 표정에서 드러나는 정보를 확인하고 분석하기 위함이다. 또한 최근에는 아이트래킹(Eye Tracking)이 사용성 테스트에 많이 사용되고 있는데, 아이트래킹은 사용성 테스트에서 참여자들이 인터페이스의 특정 요소를 바라보았는지, 시선의 이동은 어떻게 움직였는지, 어떤 영역에 시선이 많이 고정되었는지 등을 분석하여 인터페이스의 효율성을 측정하는데 도움을 줄 수 있다 (Lynne Cooke, 2005).

2.4. 측정치의 결합

사용성 테스트를 마치고 나면 사용성에 대한 총체적인 평가를 내려야 한다. 제품이 잘 동작했는지, 과거의 사용성 테스트와 비교해서 개선이 되었는지 등을 판단하려면 사용성 테스트를 통해서 추출한 다양한 측정치를 단일 사용성 점수로 결합해야 하는데, 서로 다른 척도의 점수를 적절히 결합하는 것은 쉽지 않다. 일반적으로 목표 달성치에 의한 결합과 비율에 의한 결합, Z-점수를 이용한 결합 세 가지 방법이 가능하다.

먼저 목표 달성치에 의한 결합은 목표 달성치에서 각 데이터 지점을 비교하고, 목표를 달성한 참여자들의 비율에 기반하여 하나의 단일 수치로 나타내는 것이다. 각 개별 참여자들이 태스크 완료와 태스크 시간 등의 측정치에 대하여 각각 목표 달성치를 정해놓고 테스트를 진행한 뒤에 정해진 목표를 달성했는지를 퍼센트로 나타낸다. 이 방법은 목표에 기반하여 수행 결과를 비교할 수 있고, 비교적 계산하기 쉽다는 장점이 있으나, 실제적인 사용성의 정도를 대표한다기 보다 사용성 스펙(Usability Specification)에 가깝다고 할

수 있다(Mary Beth Rosson & John M. Carroll, 2002).

비율에 의한 결합은 측정 가능한 목표가 없을 때 다른 척도로 된 점수들을 비율로 전환하여 이들의 평균값을 계산하는 것이다. 태스크 완료율이나 주관적인 평가는 최대값이 정해져 있으므로 쉽게 비율로 나타낼 수 있고, 태스크 수행시간같은 경우는 가장 빠르게 진행한 참여자의 측정치를 최상으로 기준으로 삼고 비율로 나타낼 수는 있다. 그러나 비율에 의한 결합은 전반적인 질적 레벨(Quality Level)을 표현할 수는 있지만, 태스크별 비교나 다른 제품과의 비교는 어렵다는 단점이 있다.

다른 척도가 서로 결합될 수 있도록 점수를 변환하는 또 다른 방법으로 z-점수를 사용할 수 있다. 이것은 통계적 개념에서 얼마나 균일한 산포를 가지고 있는가와 산포의 중심이 기준에 도달했는가를 판단할 수 있도록 해준다. 즉, 측정치가 정규분포를 따른다고 할 때, 규격 중심으로부터 규격 한계까지의 거리가 표준편차의 몇 배나 될 것인가를 나타내는 척도가 Z-점수이다. 측정한 데이터들을 평균이 0이고 표준편차가 1인 표준정규분포로 치환하여 표현하고 표준화 Z값을 구하면 다양한 형태의 측정치들을 동일한 기준으로 표현할 수 있다.

다음은 측정치를 Z-점수로 변환하는 공식이다.

$$Z = (x - \mu) / \delta$$

x = 변형시키고자 하는 측정치

μ = 측정치들의 분포에 대한 평균

δ = 측정치의 분포에 대한 표준편차

Z-점수는 각각의 척도가 다른 사용성 측정치를 표준화된 정규분포의 확률밀도 값으로 변환하여 동일한 기준으로 비교가 가능하다. 다만 Z-점수의 전체 평균이 전체 사용성 점수를 대변하는 것은 아니다. Z-점수는 동일한 사용성 테스트에서 다른 상황이나 다른 그룹의 참여자 데이터의 비교 등 한 집합 내에서 다른 데이터들을 비교할 때 유용하게 사용할 수 있다.

Sauro & Kindlund(2005)는 비율에 의한 결합과 Z-점수에 의한 결합을 바탕으로 식스시그마2) 방법을 적용한 SUM(Single Usability Metric)을 개발하였는데, 태스크 성공률, 태스크 수행 시간, 에러 수, 만족도의 4가

2) 1987년 모토롤라에서 시작된 대표적인 품질개선활동. 현재는 제조분야 뿐만 아니라 기업경영 전 분야에서 고객만족을 위한 혁신활동으로 인식되고 있음.

지 측정치를 비율과 Z-점수의 변환을 통해서 단일점수(시그마 수준)로 표준화함으로써, 태스크별로 전반적인 사용성 정도를 표현하였다. SUM에서 각 측정치별로 시그마수준(Z-점수)을 산출하는 방식은 표 1과 같다.

표 1. SUM의 각 측정치 산출방식

구분	산출방식
성공률	Quality Level = 1 - 총 실패수/총 기회수
태스크 시간	$Z = (x - \mu) / \delta$
에러 비율	Quality Level = 총 에러수/총 기회수
만족도	$Z = (x - \mu) / \delta$
시그마수준 $Z_{st} = Z_t + Z_{shift}(1.5)^3$	

본 연구에서는 SUM의 기본적인 사용성 측정(Usability Metrics) 포맷을 바탕으로 사용성 개선을 위한 데이터 유형별 측정방법과 개선지수를 개발하도록 한다.

3. 사용성 개선지수 개발

3.1. 사용성 개선 프로세스

SUM은 사용성 측정치를 계산하는 단계에서만 식스시그마 방법을 적용하였는데, 식스시그마 방법론은 측정 뿐 아니라 측정치의 개선을 위한 기본적인 프로세스를 제시하고 있다. 일반적으로 단순개선을 위해서는 DMAIC(Define, Measure, Analyze, Improve, Control)를, 대안 개발과 프로세스 개선이 필요할 때는 DMADV(Define, Measure, Analyze, Design, Verify)를 쓰고 있는데, 사용성 개선은 새로운 UI디자인이라는 대안을 제시하여야 하므로 DMADV 프로세스를 적용하는 것이 효과적이다(박남춘, 2006).

사용자 인터페이스 디자인의 사용성 개선을 위한 DMADV프로세스를 단계별로 정의하면 표 2와 같이 정리할 수 있는데, 본 연구에서는 이러한 DMADV프

3) 시그마수준을 의미하는 공정능력은 균일한 산포를 가질 수 있는 공정 고유의 능력을 말하는데, 단기공정능력과 장기공정능력으로 나눌 수 있다. 단기공정능력(Z_{ST})은 공정에 외부적인 영향이 없다고 판단되는 짧은 기간의 공정능력을 의미하며, 장기공정능력(Z_{LT})은 공정에 외부적인 영향이 있다고 판단되는 충분히 긴 기간의 일상적인 공정능력을 의미한다. 공정평균은 경험상 장기적으로 규격중심으로부터 약 1.5시그마(Z_{Shift}) 정도 움직이는 것으로 알려져 있다. 그래서 일반적으로 시그마수준을 이야기할 때는 장기공정 능력에 1.5를 더한 단기공정능력을 산출한다.

로세스를 따라 Measure단계에서 사용성 테스트를 진행하여 현재의 사용성 수준을 파악하고 UI디자인 개선이 끝난 후인 Verify단계에서 다시 한번 사용성 테스트를 진행하여 사용성 개선 정도를 비교하였다.

표 2. UI디자인의 사용성 개선을 위한 DMADV 프로세스

단계	정의
Define	기본적인 조사와 사용자에 대한 분석을 통해 UI 디자인의 사용성 개선에 대한 목표를 정한다.
Measure	사용성 측정치(Usability Metrics)를 선정하고 현 수준을 파악한다.
Analyze	사용성 측정치에 영향을 주는 요소들을 수집하고 가설검정을 통하여 핵심적인 영향요소를 선정한다.
Design	선정된 영향요소를 바탕으로 대안을 만들고 사용성을 개선할 수 있는 UI디자인의 최적안을 개발한다.
Verify	시뮬레이션을 통해 개선된 UI디자인이 사용성 측면에서 얼마나 개선되었는지를 테스트한다.

3.2. 사용성 개선지수 산출을 위한 사용성 측정방법(Usability Metrics)

2.4장에서 설명한 SUM은 개별적인 사용성 측정치를 효과적으로 결합하고 정량적인 사용성 분석에 도움을 줄 수 있는 방법이다. 그러나 다음과 같은 한계점도 가지고 있다.

첫째, 성공률은 이진성공에 대한 기준(어디까지를 성공으로 볼 것인가)과 신뢰구간을 포함해야 하는데, SUM에서는 단순한 백분율로 비율계산한 값을 시그마 수준으로 나타내었다. 단순한 성공률을 정량적으로 계산하여 사용성 측정치를 결합하기 보다는 어떤 요인이 성공여부에 영향을 주었는지를 분석하는 정성적 접근이 필요하다.

둘째, SUM에서는 모든 데이터를 이산형과 연속형 데이터로 치환하여 계산하였는데, 이는 산포에 따라서 그 차이가 현저히 클 수 있으며 만족도는 계량형 데이터가 아니므로 다른 방법의 지수 산출이 필요하다. 결점형 데이터인 에러 수 또한 비율로 치환하여 전체기회 중 결점 수를 기준으로 Z-점수를 산출하였는데, 엄격히 말하면, 결점형 데이터는 이산분포를 따르지 않으므로 계산식이 달라져야 한다.

셋째, 현재 상태의 사용성을 단일지수로 표현함으로써 동일한 사용성 테스트 세션에서의 사용성은 비교했지만, 사용성의 개선정도에 대한 비교는 빠져있

다. 사용성을 단일지수로 나타낸다는 것은 현재의 수준과 개선된 수준차이를 산출할 수 있으며, 더 나아가 다른 제품간의 사용성 수준을 비교할 수 있는 가능성이 있다.

표 3. 사용성 테스트에서 사용되는 측정치와 데이터 유형

데이터 유형	단위	측정치
Continuous (계량형)	시간	태스크를 마칠 때까지 걸리는 시간 메뉴 간 이동하는데 걸리는 시간 특정 정보를 찾는 데 걸리는 시간 에러로부터 복귀하는데 걸리는 시간
Defect (결점형)	회수	메뉴를 잘못 선택한 회수 잘못된 기능을 선택한 회수 도움말 화면을 반복적으로 본 회수 메뉴얼을 참조한 회수 기타 에러의 회수
Defective (불량형)	비율	태스크 성공률 만족도에 대한 자가기록 점수

일반적으로 사용성 테스트에서 사용되는 측정치를 데이터 유형으로 분류하면 표 34)과 같이 구분할 수 있다. 이는 데이터를 연속형 데이터와 이산형 데이터 2가지로 구분했던 SUM보다 세분화된 분류로서, 계량형 데이터나 비율데이터로 치환하지 않고 산출할 수 있는 측정치들로 분류하였다.

본 연구에서는 SUM을 기본 모델로 하여 측정치를 단일지수로 결합하되, 위에서 언급한 한계점을 보완하여 사용성 측정치를 ‘수행시간’, ‘에러 수’, ‘만족도’ 세 가지로 재정의하고 측정기준과 사용성 개선지수 산출방식을 개발하였다.

‘수행시간’은 계량형 데이터로 정규분포를 따르므로 표준 정규분포의 확률분포를 기준으로 확률 P(d)를 구하여 Z-점수를 산출하고, 계수형 데이터인 ‘에러 수’는 뵈와중(Poisson)분포를 따르므로, DPU값을 기준으로 수율을 계산하여 P(d)를 도출한 다음 이를 통해 Z-점수를 산출하였다. ‘만족도’는 점수 백분율을 그대로 P(d)로 잡는 SUM과는 달리, DPO값을 구하여 확률 변수에 대한 Z-점수를 계산하도록 하였다.

3.3. 사용성 개선지수 개발

사용성 개선을 위해서는 개선하고자 하는 대상의

4) Laura M. Leventhal & Julie A. Barnes(2007)이 제시한 수행 측정치와 주관적 측정치를 기준으로 식스시그마 방법의 관점으로 정리하고 분류하였음.

명확한 정의와 그 측정치의 현재 수준을 파악하는 것이 가장 중요하다. 개선할 대상과 목표가 명확해야만 얼마나 개선되었는지를 파악할 수 있고, 지속적으로 관리가 가능하기 때문이다. 본 연구에서는 개선할 대상인 ‘수행시간’, ‘에러 수’, ‘만족도’ 세 가지의 사용성 측정치(Usability Metrics)에 대하여 표 4와 같이 시그마 수준 산출 절차를 개발하였다. 또한 태스크별로 개별적인 사용성 개선지수 SUI Index(Separate Usability Improvement Index)를 산출하고, 이를 토대로 전체적인 사용성 개선지수 TUI Index(Total Usability Improvement Index)를 산출하도록 하였다.

표 4. 데이터 유형에 따른 시그마 수준 산출 공식

데이터 유형	데이터 분포	시그마 수준 산출 절차
수행시간 (계량형)	정규분포 (Normal)	$P(d) = P(d)_{\text{최대}} + P(d)_{\text{최소}}$ $Z\text{-Value} = Z \text{ of } P(d)$
에러 수 (결점형)	뵈와중분포 (Poisson)	$Yield = e^{-DPU}$, $P(d) = 1 - Yield$ $Z\text{-Value} = Z \text{ of } P(d)$
만족도 (불량형)	이항분포 (Binomial)	$P(d) = DPO$ $= \text{현재값} - \text{목표치} / \text{목표치}$ $Z\text{-Value} = Z \text{ of } P(d)$

시그마수준 $Z_{st} = Z_{lt} + Z_{shift}(1.5)$

※P(d)=결함이 발생할 확률, Z-Value=Zt, DPU=Defect/Unit, DPO=Defect/Opportunity, Yield=수율

사용성 개선절차는 현재 사용성 수준 측정, 목표치 설정, 개선 후 사용성 수준 측정의 순서로 진행되며, 사용성 개선지수 SUI Index와 TUI Index 산출식은 다음과 같다.

$$SUI\ Index = \text{개선후 } Z_{st} - \text{개선전 } Z_{st}$$

$$TUI\ Index = \text{개선후 } Z_{st} \text{ of } \sqrt[3]{RTY} - \text{개선전 } Z_{st} \text{ of } \sqrt[3]{RTY}$$

($\sqrt[3]{RTY}$ = 표준화 수율, RTY=누적수율)⁵⁾

4. 사례 적용

4.1. 개요

사용성 개선지수의 적용을 위해서 S사의 포켓캠코

5) 전체공정의 품질수준을 표현하는 누적수율(Rolled Throughout Yield : RTY)은 개별 수율의 곱이다. 그러나 사용성 측정치와 태스크별 수준이 각각 다르므로 기하평균 개념의 표준화 수율로 전체 시그마 수준을 산출하는 것이 적절하다.

더 제품 번들소프트웨어인 인텔리 스튜디오라는 PC 네트워킹 프로그램을 선정하여 사용성 테스트를 실시하고, 문제점을 분석하여 UI를 개선한 후에 다시 사용성 테스트를 실시하였다. 개선전과 개선후의 테스트 환경을 동일하게 유지하기 위하여 시료는 플래시 시뮬레이션으로 제작하였고 사용자 관찰실에서 동일한 조건으로 진행하였다. 태스크는 사전 분석에서 잠재적인 사용성 문제가 있다고 판단된 ‘태스크 1 : 네트워킹 연결 및 프로그램 실행’, ‘태스크 2 : 동영상 공유 및 업로드’, ‘태스크 3 : 동영상 삭제 및 복귀’의 3가지 태스크를 선정하였고, 참여자는 개선전 10명, 개선후 10명 총 20명으로, 숙련도에 따른 비율은 초심자 4명, 중간사용자 3명, 고급 사용자 3명으로 동일하게 유지하였다.

사용성 개선지수 산출을 위해서 식스시그마 프로젝트에서 주로 사용하는 통계프로그램 Minitab16을 사용하여 기본적인 통계값과 시그마수준을 계산하였다.

4.2. 개선전의 사용성 측정치

사용성 테스트 결과 각 태스크별로 사용성 측정치는 표 5와 같고, 이 측정치를 Minitab16을 활용하여 시그마 수준을 산출하였다. ‘수행시간’은 계량형 데이터이므로 정규분포를 따르는 산포의 공정능력분석인 Capability Analysis(Normal)을 실시하였고, ‘에러 수’는 결점형 데이터이므로 Poisson분포의 DPU계산을 통하여 시그마 수준을 구하고, ‘만족도’는 이항분포를 따르므로 누적확률 DPO에 대한 Z-점수를 계산하여 각각의 시그마 수준을 산출하였다.

표 5. 개선 전 각 Task별 측정치

구분	참여자	참여자									
		A	B	C	D	E	F	G	H	I	J
수행시간 (초)	Task1	47	20	25	40	27	18	35	27	33	45
	Task2	69	48	51	112	48	109	60	50	68	75
	Task3	49	66	102	70	32	22	40	37	55	38
에러 수 (회)	Task1	10	2	3	1	1	2	4	2	2	5
	Task2	1	1	8	6	1	5	2	4	3	7
	Task3	2	7	3	7	3	3	0	0	3	1
만족도(10점척도)		6.6	6.0	5.8	6.2	7.2	6.5	6.3	7.0	6.1	6.0

그림 1은 태스크 1의 ‘수행시간’에 대한 공정능력 분석을 Minitab16의 Quality Tools를 활용하여 수행한 결과의 예로, Z Bench값이 -1.17이므로 시그마 수준

Zst = -1.17 + 1.5 = 0.33σ가 된다.

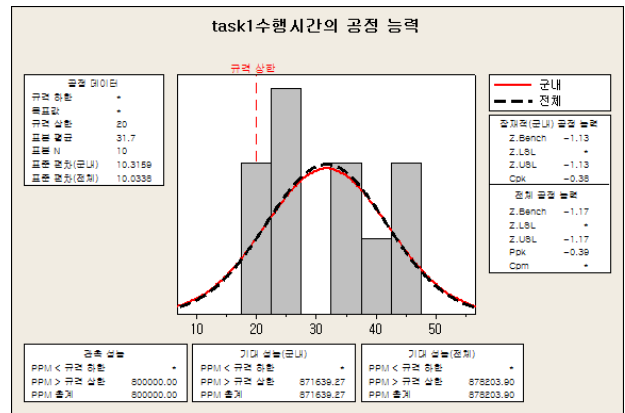


그림 1. Task1의 수행 시간에 대한 공정능력 분석 예

그림 2는 태스크 1의 ‘에러 수’에 대한 공정능력 분석을 Minitab16의 Capability Analysis(Poisson)를 활용하여 수행한 결과의 예로, DPU값이 3.8이므로 수율 RTY = e^{-DPU} = 0.0407622이 되고, 이를 표준정규분포로 치환하여 Z-점수를 구하면 -1.74가 나온다. 따라서 시그마 수준 Zst = -1.74 + 1.5 = -0.24σ가 된다.

만족도는 불량률과 같은 이항분포 데이터라고 할 수 있다. 그러나 사용성 테스트와 같이 측정치를 다수 확보할 수 없을때에는 상대적 지수를 계산하는 것이 타당하다. 본 연구에서는 |현재의 만족도 평균 - 만족도의 목표치| / 만족도의 목표치를 DPO값으로 계산하여 지수를 산출하였다. 따라서 개선 전의 만족도를 위의 산출방식에 따라 계산하면 DPO = |6.37-8| / 8 = 0.20375가 되고, Z bench값은 0.83이 된다. 그러므로 만족도의 현재수준 Zst=0.83+1.5=2.33σ가 된다.

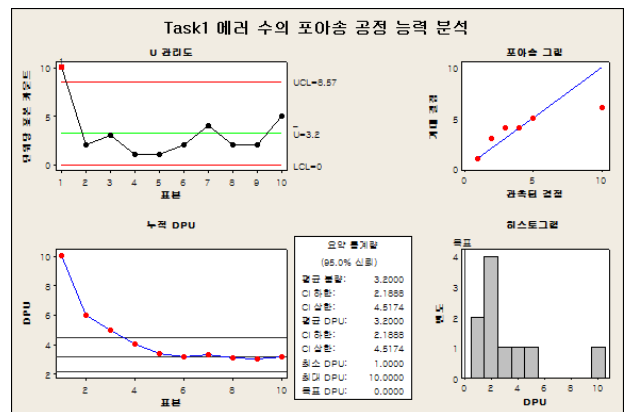


그림 2. Task1의 에러 수에 대한 공정능력 분석 예

이러한 방법을 통하여 산출한 시그마 수준에 따라

서 측정치의 현재 수준과 목표수준을 각 태스크별로 정리하면 표 6과 같다. 목표수준은 현 수준의 효율을 20% 향상시킨다고 가정하여 시그마 수준을 명시하였다.

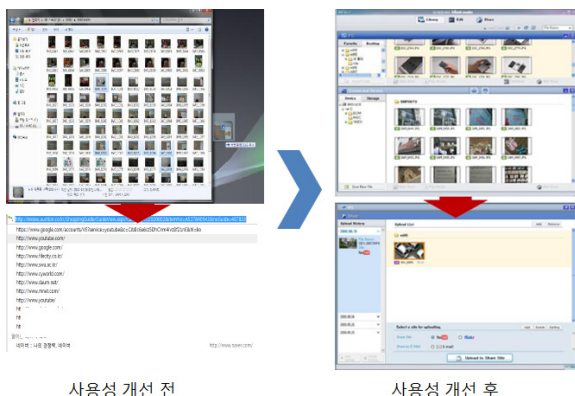
표 6. 개선 전 각 Task별 측정치의 시그마 수준

구분	데이터 유형	규격	현수준	목표수준	
수행시간	계량형	Task1	초	0.33 δ	1.03 δ
		Task2		0.29 δ	1.22 δ
		Task3		1.02 δ	1.63 δ
에러 수	결점형	Task1	회	-0.24 δ	0.8 δ
		Task2		-1.47 δ	-0.82 δ
		Task3		0.34 δ	0.74 δ
만족도	불량형	점수	2.33 δ	2.7 δ	

4.3. UI디자인 개선

사용성 테스트를 통한 분석 결과, 네트워크 연결 경로의 단순화와 전체 화면구조의 레이아웃 변화, 그리고 인터랙션 방식과 메뉴이동 경로의 개선 등이 필요하다고 판단하였고, 다양한 UI디자인 대안 검토를 통해 그림 3과 같은 개선된 UI디자인 결과물을 제시하였다.

예를 들면 ‘태스크 2 : 동영상 공유 및 업로드’의 경우, 기존의 많은 파일들을 하나씩 확인해가며 선택하고 다운받아야 하는 불편함이 따랐으나, 개선된 UI에서는 올리고 싶은 동영상을 촬영직후 본체에서 미리 공유설정이 가능하도록 해 자동으로 로그인창이 생성되도록 하였다. 또한 동영상 선택 삭제가 가능하도록 변경하고, 삭제화면에서 메인화면으로 복귀하는 단계를 줄였다. 개선된 UI디자인 대안은 사용성테스트를 할 수 있도록 플래시 시뮬레이션으로 제작하였다.



사용성 개선 전

사용성 개선 후

그림 3. 사용성 개선전과 개선후의 UI디자인 예

4.4. 개선 후 사용성 측정치

사용성 개선 후에 동일한 조건에서 동일한 3가지 태스크에 대하여 또 다른 10명의 참여자를 대상으로 사용성 테스트를 실시하였다. 표 7은 사용성 개선 후의 사용성 테스트 결과 각 태스크별 측정치를 정리한 것이다.

표 7. 개선 후 각 Task별 측정치

구분	참여자	참여자									
		A	B	C	D	E	F	G	H	I	J
수행시간 (초)	Task1	33	20	20	25	25	15	20	18	20	25
	Task2	37	44	38	43	35	50	40	42	35	42
	Task3	44	42	48	38	30	25	38	35	30	40
에러 수 (회)	Task1	1	0	1	0	0	0	1	0	0	2
	Task2	1	1	3	1	1	0	1	1	0	2
	Task3	0	1	2	1	0	1	0	0	0	1
만족도(10점척도)		7.8	7.0	6.6	7.1	7.6	6.8	6.9	7.9	7.1	7.8

위의 측정치만 보더라도 개선 전보다 각 태스크별로 수행시간과 에러 수, 만족도가 상당 부분 개선되었음을 알 수 있지만 어느 정도 개선되었는지를 정확하게 표현하기 위해 개선 전과 마찬가지로 태스크별로 데이터 유형에 따라 시그마 수준을 산출하였다. 표 8에서 개선 전과 개선 후의 각 태스크별 측정치의 시그마 수준을 비교해 보면, 거의 모든 측정치에서 시그마 수준이 향상되었음을 알 수 있다.

표 8. 개선 전과 후의 각 Task별 측정치의 시그마 수준

구분	데이터 유형	규격	개선 전	개선 후	SUI Index	
수행시간	계량형	Task1	초	0.33 δ	1.08 δ	0.75 δ
		Task2		0.29 δ	1.37 δ	1.08 δ
		Task3		1.02 δ	1.92 δ	0.9 δ
에러 수	결점형	Task1	회	-0.24 δ	1.77 δ	2.01 δ
		Task2		-1.47 δ	1.07 δ	2.54 δ
		Task3		0.34 δ	1.62 δ	1.28 δ
만족도	불량형	점수	2.33 δ	2.82 δ	0.49 δ	

4.5. 사용성 개선지수 산출

사용성 개선지수 SUI Index = 개선후 Zst - 개선전 Zst이므로 각 태스크별로 개선지수를 산출하면 표 8과 같다. 전반적으로 사용성이 개선되었지만, 그 중에서

도 태스크 2의 ‘수행시간’과 ‘에러 수’의 SUI Index가 높았다. 개선 후의 UI디자인이 태스크 2의 사용성 개선에 영향을 많이 주었다는 것을 알 수 있다.

전체적인 사용성 개선 지수 TUI Index를 계산식에 의해서 산출하면 다음과 같다.

개선 전의 $\sqrt{RTY} = 0.0769787$ 이므로 $Zst = -1.42 + 1.5 = 0.08\sigma$ 이 되고, 개선 후의 $\sqrt{RTY} = 0.518382$ 이므로 $Zst = 0.05 + 1.5 = 1.55\sigma$ 가 된다.

따라서 $TUI\ Index = \text{개선후 } Zst\ of\ \sqrt{RTY} - \text{개선전 } Zst\ of\ \sqrt{RTY} = 1.55\sigma - 0.08\sigma = 1.47\sigma$ 가 된다. 즉, 전체적인 사용성 수준은 0.08σ 에서 1.55σ 로 1.47σ 만큼 향상되었음을 알 수 있다.

5. 결론 및 금후 연구과제

본 연구에서 제시한 사용성 개선지수는 명백한 목표수준을 정의할 수 있는 측정 가능한 특징을 가지고 있다. 따라서 사용성 평가의 결과를 객관적인 지표로 나타낼 수 있고, 사용성 지수(시그마 수준)를 활용한 평가와 개선작업을 포함하는 사용성 평가 프로세스를 적용하면 개별적인 태스크의 개선 비교(SUI Index)는 물론 제품의 사용자 인터페이스 전반적인 사용성 수준의 개선(TUI Index)을 비교할 수 있도록 하였다.

Molich(2004)는 사용성 테스트의 효과는 어떤 태스크를 선정하고 어떤 방법론을 적용하고 어떤 전문가가 테스트에 참가하는가에 달려있다고 하였다. 본 연구에서 제시한 사용성 개선지수 또한 위의 조건에 따라 그 신뢰도가 달라질 수 있다고 생각한다. 다만, 그럼에도 불구하고 사용성 평가에 있어서 표준화된 일정한 척도로 사용성의 개선정도를 나타낸다는 것 자체만으로도 강력한 틀이 될 수 있을 것이다. 본 연구의 기대효과를 정리하면 다음과 같다.

첫째, 사용자 인터페이스 디자인 연구에 있어서 구현과 대안제시 과정을 보다 체계화할 수 있을 것이다. 주로 전문가의 경험과 종합적인 문제해결 능력이 중시되는 정보구조 설계(Information Architecture)와 GUI(Graphic User Interface)디자인에 있어서도 사용성 개선을 위해서는 명확한 목표를 설정하고 그에 맞는 인터페이스의 대안을 설계하고 제시하도록 하고 이를 평가할 수 있게 됨으로써 합리적이고 체계적인 방법론이 개발되어질 수 있을 것이다.

둘째, 목표에 의한 사용자 인터페이스 디자인이 가능해 진다. 단순히 사용하기 편리한 UI 또는 사용자가

만족할 만한 UI와 같은 추상적인 목표가 아니라 사용성 수준으로 명확히 목표를 설정하고 그 목표를 달성하기 위하여 합리적인 방법으로 대안을 창출할 수 있다.

셋째, 제품의 품질 특성으로서 사용성은 제품의 기능성과 달리 사용자가 제품을 어떻게 다루는가를 나타내므로 동일한 기준에 의해서 관리하기가 쉽지 않았다. 하지만 사용성 개선 지수를 개발함으로써 보다 체계적인 방법으로 현수준과 개선후의 수준을 비교할 수 있으며 데이터의 지속적인 관측 및 예측이 가능해져 제품 유형별, 사용자별로 체계적인 관리를 할 수 있을 것이다.

넷째, UI디자인 프로세스에서 사용성의 지표를 관리 가능하도록 함으로써, 전략적인 투자와 관심이 필요한 경영자원으로서의 UI디자인의 역할을 명확히 할 수 있을 것이다. 이러한 노력은 추상적으로 그 중요성 정도만 인식되던 UI디자인과 사용성 관련 분야의 위상을 더욱 높일 수 있을 것이다.

본 연구에서는 사용성 개선지수 개발을 위하여 측정 가능한 사용성 측정치를 중심으로 진행하였다. 그러나 사용자 인터페이스는 물론 전체적인 사용자 경험(User Experience)이 더욱 중요해진 이때, 절대 무시할 수 없는 것이 측정이 불가능한 상대적인 가치와 감성적인 부분이다. 따라서 향후 연구과제로 이러한 정성적이고 상대적인 가치에 대해서 대안적인 연구가 필요하다고 생각한다. 또한 제품이나 서비스를 소비하고 사용하는 행태의 전반적인 맥락(Context)과 사용자태도가 사용성에 미치는 영향에 대한 연구가 뒤따라야 할 것이다.

참고문헌

- 박남춘 (2006). 사용자 인터페이스 디자인에 있어서 식스시그마 방법론의 적용에 관한 연구. *기초조형학 연구*, 7(4), 241-251.
- 이건표 (1996). 제품 디자인에 있어서의 사용성 평가에 관한 연구. *디자인학연구*, 18, 137-150.
- Ann Chadwick-Dias, Michelle McNulty, & Tom Tullis. (2003). Web usability and age: how design changes can improve performance. *Proceedings of the 2003 conference on Universal usability*, Vancouver, British Columbia, Canada.
- Chin, J. P., Diehl, V. A., and Norman, K. L. (1988). Development of an instrument measuring user

satisfaction of the human-computer interface. *in Proceedings of SIGCHI '88* (pp. 213-218), New York: ACM/SIGCHI.

Laura M. Leventhal & Julie A. Barnes. (2007). *Usability Engineering : Process, Product, and Examples*. Prentice Hall.

Lewis, J. R. (1992). Psychometric evaluation of the Post-Study System Usability Questionnaire: The PSSUQ. *in Proceedings of the Human Factors Society 36th Annual Meeting* (pp. 1259-1263). Atlanta, GA: Human Factors Society.

Lynne Cooke. (2005). Eye Tracking: How It Works and How It Relates to Usability. *Technical Communication, Washington, 52*, 456-464.

Mary Beth Rosson & John M. Carroll. (2002). *Usability Engineering : Scenario-Based Development of Human-Computer Interaction*. Morgan Kaufmann.

Molich, R., Ede, M., Kaasgaard, K., & Karyukin, B (2004). Comparative Usability Evaluation. *Behaviour & Information Technology, 23*(1), 65-74.

Nielsen, J. (2003). *Measuring usability: preference vs. performance*. New York: Communications of the ACM.

Sauro, J. & Kindlund E. (2005). A Method to Standardize Usability Metrics into a Single Score. *in Proceedings of the Conference in Human Factors in Computing Systems (CHI 2005) Portland, OR*, 401-409.

Tom Tullis, Bill Albert. (2008). *Measuring User Experience*. Morgan Kaufmann.

<http://www.measuringusability.com/>

원고접수 : 11.02.07

수정접수 : 11.02.25

게재확정 : 11.02.28