

Web-script を用いた待ち時間評価システム —信頼性評価と特性分析¹⁾—

筑波大学大学院人間総合科学研究科・KDDI 研究所 新井田 統

KDDI 研究所 上村 郷志

筑波大学人間系 原田 悦子

Web-script based subjective quality assessment system: Reliability assessment and characteristic analysis

Sumaru Niida (*Graduate School of Comprehensive Human Sciences, University of Tsukuba, Tsukuba 305-8572, Japan, Communication Network Planning Laboratory, KDDI R&D Laboratories Inc., 2-1-15 Ohara, Fujimino, Saitama 356-8502, Japan*)

Satoshi Uemura (*IP Communication Quality Laboratory, KDDI R&D Laboratories Inc., 2-1-15 Ohara, Fujimino, Saitama 356-8502, Japan*)

Etsuko T. Harada (*Faculty of Human Sciences, University of Tsukuba, Tsukuba 305-8572, Japan*)

This paper discusses the reliability and characteristic of evaluation method for an effective and efficient analysis of mental process of waiting in ICT use. In behavior analysis of cognitive artifacts, it is required to consider several cognitive factors such as an experience and an attitude toward ICT use. Authors have developed the web-script based assessment system and examined an evaluation method with ecological validity. In the method, participants received programs containing an experimental task written by web-script. They executed the program on a PC or cellular phone terminals as they chose, and transmitted their responses over a network. In the preparatory experiment, we evaluated the time resolution of the system and showed that it has adequate performance for evaluation of the degree of satisfaction considering waiting time. The following experiment indicated its stability and revealed a characteristic of the method by analyzing two different investigations.

Key words: reliability, evaluation method, waiting time, time perception, ICT

背景と目的

近年の携帯電話やインターネットなどの情報通信

技術 (Information and Communication Technologies, 以下 ICT) の普及は、人々に多くの利便性をもたらしてきた。利用者は時間と場所という二つの束縛から解放され、自らの生活様式に適したコミュニケーションの形を選択して取り込んでいっている。こうした高度情報化社会の発達に伴う利用行動の多様化により、サービス品質を一様に扱うことが困難に

1) 本研究の一部は、2010年電子情報通信学会ソサイエティ大会において発表されました (新井田・上村・中村, 2010a)。

なってきた。

こうした流れに呼応して、これまではシステム内で発生する伝送遅延などのみで規定してきた通信ネットワークの品質についても、QoE (Quality of Experience) (ITU, 2007) と呼ばれる状況を考慮した上で利用者が体感する品質に注目が集まっている (阿部・石橋・吉野, 2008, 新井田・原田, 2011)。初期的な検討では、待ち時間に対して主観的に感じる満足度を物理的時間に対する関数として表し、システム設計基準値を作成することが目的とされていた。その結果、国際機関である ITU (International Telecommunication Union) から、PC によるウェブサイトへの接続における利用者の評価を行う手続きが勧告される (ITU, 2005) など、一定の成果が得られている。さらに近年では、物理的時間に加えて利用の文脈を考慮した体感品質の検討が進められ、携帯電話における許容可能な待ち時間の評価 (Niida, Uemura & Nakamura, 2010b) や、コンテンツに支払う料金の影響の検討 (Liu, Ngu & Zeng, 2004) などが行われている。

こうした通信ネットワークの体感品質の評価においては、品質に影響を与えると推測される要因を規定・操作し、実験室環境で他の要因を一定に統制した形で待つ行為に対する満足度を5件法などで測定することが多かった。しかし、通信ネットワークの利用を人-人工物間の相互作用と捉えた場合、待つ行為の評価においては、ユーザの特性、システムの特性を含むユーザー-システム間対話、行われるタスクの特性、そしてそれらが発生する状況特性の四つの要素を同時に考察していく必要がある (原田, 1997)。そのため、特に各種の状況要因をいかに実験的に制御しつつ、生態学的妥当性を有する環境を設定するかを考慮する必要がある。実利用環境からあまりにかけ離れた実験環境での実験データは、仮に安定した結果を得ることができたとしても、実際の自然環境下での経験を正しく反映した結果であるという証明が困難となる可能性が高い。また、結果を工学的問題解決にも利用するためには、利用者全体を母集団としてその効果を検証していくことが必要であり、そのためには、一定程度、実験参加者の数も重要な要因となる。

これらを考慮して、我々は一般利用環境下で待ち時間を評価するシステムを開発した (Uemura, Niida & Nakamura, 2011)。本評価システムでは、実験参加者は、自分自身の ICT 端末に実験課題が含まれた評価用プログラムを導入して実験を行い、結果をネットワーク経由で返送する。このため、“実験参加者が自分の都合のよい時間に都合のよい場所

で、自分自身が通常使っている端末を利用して課題を実施することができるのみならず、より普段の状況に近い環境で実験を行うことが可能である。また、スクリプト言語を使用して作成された評価用プログラムは、実験参加者の通信環境にかかわらず待ち時間を実験的に統制した利用状況を模擬的に作り出すことができる。さらに、ネットワークを介してデータの収集が行えるため、大量の回答を得ることが比較的容易である。

しかし、シミュレーション型の評価実験をこうした実利用環境で行う試みは稀であり、統制不能な要因が結果に影響を与える可能性も考えられる。このため、提案システムを用いた実験が、操作対象である独立変数・待ち時間が従属変数である満足度に与える影響を安定的に評価可能であることを実証することは重要である。

このため本稿は、異なる時期に行った同一の二回の実験を比較することで、提案した評価システムにより得られる結果の再現性により、手法の信頼性を評価することを目的とした。同時に、参加者の特性による結果の変化として特性の効果を分析することで、こうしたシステムによる今後の評価において考慮すべき点を明らかにする。

予備実験

ICT 利用時の待つ行為の評価においては、待ち時間に関する条件の統制が最も重要な要因であり、そのため評価システムの時間制御の精度が問題となる。こうした評価を行う認知心理学実験において、PC を用いて刺激を提示する手法は一般的に用いられているが、本提案システムのように携帯電話上で Web-script を用いた評価システムを用いた方法については先例が少なく、検討されていない。そこで予備実験として、提案システムの時間分解能についての評価を行った。

手続き 時間分解能を評価するため、Flash Lite で作成をした実験刺激を、4種類の異なる携帯電話端末 (au iida G9, au W63SA, DoCoMo N905i, Softbank 912T) を用いて再生をして、その画面遷移をハイスピードビデオカメラにて撮影し、オフラインで再生状況をフレーム単位で分析した。実験では、画面が転換する間隔設定値を7条件 (0.1, 0.2, 0.5, 0.8, 1.0, 2.0, 3.0秒) に定め、各設定値に対して、それぞれの端末で10回の再生を行った。

結果と考察 結果を Table 1 に示す。表中には平均と標準偏差に加えて、設定値との差分を示すため設定値と測定値内の最大値及び最小値との差、およ

び $\sqrt{\mu_2}$ (設定値周りの二次モーメントの累乗根、設定値を中心としたバラツきの指標となる)を示した。設定値はWeb-scriptを作成する際に画面遷移にかかる時間として設定した値である。また、測定値は、撮影された映像をフレーム毎に確認して、画面転換開始から終了までのフレーム数にフレームレートを積算して算出した。

測定値について端末種別(4)と設定値(7)を要因とする2要因分散分析を行ったところ、端末種別($F(3,252)=122.98, p<.001$)と設定値($F(6,252)=38696.58, p<.001$)の主効果のいずれもが有意であり、併せて端末種別と設定値の交互作用($F(18,252)=3.22, p<.001$)も有意であった。これより、測定値が設定した時間長や使用する端末により差があることが示された。

しかしTable 1に示したとおり、設定された時間長に対して標準偏差が単調増加とはなっておらず、設定時間長を長くした場合に誤差が累積して大きくなるような問題は発生していない。また、設定値に対して最大で300msecのズレ(設定値2秒)が発生しているものの、標準偏差が0.072秒以内、 $\sqrt{\mu_2}$ でも0.15秒以内に収まっており、データのバラツキが小さい安定した結果が得られている。1秒以上の設定値に限れば、誤差は10%以内に収まっており、端末による変動を考慮しても、秒オーダーでの実験には問題なく利用可能であると言える。

実 験

本実験では、異なる時期に異なる実験参加者群に対して同一の実験を2回実施し、その結果を比較することで再現法による本手法の信頼性を検討することとした。ここで、実験全体を示す“実験”や、一つ一つの課題の実行を示す“試行”などとの混乱を避けるために、新たに「実査」という言葉を設け、本稿ではこれ以降二回行われた実験を“第1実査”

“第2実査”とする。

実験は、ウェブ調査会社を通して実験参加者の募集を行い、実際の参加者を性別と年代により選定するリソース型の属性絞り込み方式(大隈, 2002)による調査を行った。吉村(2001)によれば、こうしたインターネット調査の参加者は20代後半から40代前半で女性より男性が多く、生活の中でインターネットが占める割合が大きい人が多いとされている。こうした実験参加者の特性、すなわち性別や年齢、インターネットサービスの利用頻度が、提案する評価手法へ与える影響について評価した。

方法

実験参加者 実験参加者はネット広告で集められ、実験への自発的参加に対し規定の謝礼が支払われた。実験は、約2ヵ月の間隔で、異なる時期に異なるウェブ調査会社により参加者の募集が行われて二回実施された。実験参加者数は、1回目の実査は405名(男性175名、女性230名)で、2回目の実査は275名(男性135名、女性140名)であった。年齢のレンジは15-59歳であり、10代から50代の各年代の参加者数がほぼ等しくなるよう各回の実査内で統制された。

手続き 2回の実査は、同一の課題、同一の手続きで行われた。実験参加者は、まず評価プログラムを自分の携帯端末に導入した。評価プログラムは、携帯電話を用いたウェブサイトへの接続における、接続開始ボタン押下後からトップページ(最初に表示されるページ)が表示されるまでの状況を模擬している。実験参加者に与えられた課題は、トップページへ接続されるまでの待ち時間に対する満足度評価であり、トップページ表示直後に“接続開始から完了までの待ち時間について満足度を回答して下さい”として5件法にて回答を求めた(非常に満足:5, 満足:4, どちらでもない:3, 不満:2, 非常に不満:1)。待ち時間長は10条件(5, 6, 7, 8, 9, 10, 11, 12, 15, 20秒)が設定され、ランダムな順序で表示された。

実験参加者には、実験を実施する場所と時間は特に指定せず、参加者の任意とした。10回の待ち時間に対する満足度評価が全て終了した後に、携帯電話によるウェブサイト接続の利用頻度と、現状満足度(「携帯電話のインターネットにアクセスするとき、トップページへの接続時間に対してどの程度満足されていますか。’)および性別と年代を、待ち時間に対する満足度評価の回答と併せて回答を求め、ネットワークを介して返送させた。

Table 1 時間分解能の評価結果(単位は秒)

設定	平均	標準偏差	$\sqrt{\mu^2}$	最大値 - 設定値	設定値 - 最小値
0.1	0.19	0.040	0.095	0.14	0.01
0.2	0.27	0.043	0.081	0.13	0.02
0.5	0.54	0.049	0.066	0.16	0.05
0.8	0.82	0.048	0.053	0.13	0.05
1.0	0.99	0.062	0.062	0.11	0.10
2.0	2.13	0.072	0.15	0.30	0.05
3.0	3.09	0.061	0.11	0.23	0.03

結 果

まず実査条件の影響を評価した。実査ごとの待ち時間に対する満足度の平均と標準偏差を Fig. 1 に示す。回答された満足度について、待ち時間条件(10; 参加者内要因)と実査条件(2; 参加者間要因)を要因として、2要因混合型分散分析を行った。その結果、待ち時間条件の主効果は有意($F(9, 6102) = 471.26, p < .001, \eta^2 = .22$) (ただし η^2 は相関比)であったが、実査条件の主効果($F(1, 678) = 1.81, p = .18$)は有意でなかった。また、待ち時間条件と実査条件の交互作用は有意傾向を示した($F(9, 6102) = 1.80, p = .064, \eta^2 = .00085$)。単純主効果検定の結果は、11, 12, 20秒の待ち時間において、実査条件間に有意な差があり、いずれも第1実査の満足度が高かった。更に Levene の等分散性の検定を行ったところ、全ての待ち時間において等分散性の仮説は棄却されなかった(5%水準)。

次に、満足度評価と各種属性との関係を明らかにするため、実査条件と待ち時間条件に、属性パラメータのうち一つを加えた3要因分散分析を、パラメータを変えて行った。

まず、性別を属性パラメータとして分析を行った。その結果、二次の交互作用は有意でなく($F < 1, ns$)一次の交互作用については、待ち時間条件と性別の交互作用が有意($F(9, 6084) = 2.98, p = .002, \eta^2 = .0014$)で、待ち時間条件と実査条件($F(9, 6084) = 1.87, p = .051, \eta^2 = .00089$)および性別と実査条件($F(1, 676) = 2.72, p = .10, \eta^2 = .0018$)は有意傾向であった。また主効果については、待ち時間条件のみ有意($F(9, 6084) = 465.89, p < .001, \eta^2 = .22$)であり、実査条件($F(1, 676) = 1.35, p = .25$)および性別($F < 1, ns$)は有意でなかった。

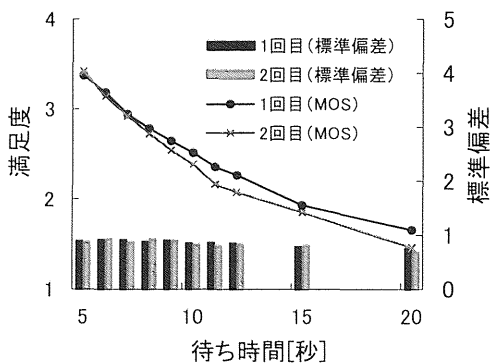


Fig. 1 各実査条件における待ち時間に対する満足度の平均と標準偏差

待ち時間条件と性別の交互作用が有意であったため、単純主効果検定を行った結果、待ち時間条件が8秒の点のみで、男性と女性間に有意な差が見られた。男性は待ち時間条件が8秒と9秒の間のみ有意差が消失しており、女性は、待ち時間条件が11秒と12秒の間で、有意差が消失していた。性別での群分けによる、待ち時間に対する満足度の平均を Fig. 2 に示す。

有意傾向であった待ち時間条件と実査条件の交互作用について単純主効果検定を行ったところ、待ち時間条件が11, 20秒の点において、実査条件間に有意な差が見られ、いずれも第1実査の満足度が高かった。また、性別と実査条件の交互作用に対する単純主効果検定の結果は、女性のみ実査条件間に有意な差が存在し、第1実査での満足度が第2実査での値よりも高かった。

次に、年代を属性パラメータとして分析を行った。各年代の実験参加者数は、10代(15-19歳)が125名、20代(20-29歳、以降同様)が160名、30代が136名、40代が139名、50代が120名であった。分析の結果、一次の交互作用のうち、待ち時間条件と実査条件のみ有意傾向が見られ($F(9, 6030) = 1.75, p = .072, \eta^2 = .00083$)、実査条件と年代($F(4, 670) = 1.59, p = .18$)および年代と待ち時間条件($F < 1, ns$)は有意でなかった。また、二次の交互作用についても有意ではないという結果となった($F < 1, ns$)。それぞれの条件の主効果については、待ち時間条件のみ有意($F(9, 6030) = 463.70, p < .001, \eta^2 = .22$)であり、実査条件($F(1, 670) = 1.86, p = .17$)および年代($F < 1, ns$)は有意でなかった。

次に、携帯電話によるウェブ利用頻度を属性パラメータとした分析を行った。実査条件と利用頻度のクロス集計の結果を、Table 2 に示す。第2実査で

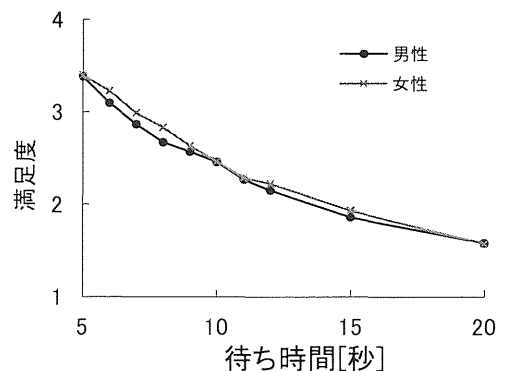


Fig. 2 各性別における待ち時間に対する満足度の平均

は、月に1-3回程度以下の回答が無いため、1日1回程度以下を全てまとめて5つの利用頻度群を構成し、カイ二乗検定を行ったところ、有意な結果が得られた ($\chi^2(4) = 187.95, p < .001$)。すなわち第1実査に比べ、第2実査の参加者はより頻繁にネットワーク利用を体験しているものと考えられた。そこで、利用頻度群(5)を独立変数に加え、待ち時間条件と実査条件による3要因分散分析を行ったところ、待ち時間条件と実査条件 ($F(9, 6030) = 1.11, p = .35$) および実査条件と利用頻度群 ($F < 1, ns$)、利用頻度群と待ち時間条件 ($F < 1, ns$) のいずれの一次の交互作用も有意でなかった。また、二次の交互作用も有意ではなかった ($F < 1, ns$)。更に、それぞれの条件の主効果については、待ち時間条件のみ有意 ($F(9, 6030) = 138.22, p < .001, \eta^2 = .077$) であり、実査条件 ($F < 1, ns$) および利用頻度群 ($F(4, 670) = 1.88, p = .11$) は有意でなかった。

最後に、携帯電話によるウェブサイト接続の現状満足度を属性要因として分析を行った。実査条件と現状満足度のクロス集計の結果に対して、カイ二乗検定を行ったところ、有意な結果は得られなかった ($\chi^2 < 1, ns$)。また、現状満足度を加えた待ち時間条件と実査条件による3要因分散分析を行ったところ、待ち時間条件と実査条件 ($F < 1, ns$) および実査条件と現状満足度 ($F(4, 670) = 1.32, p = .26$)、現状満足度と待ち時間条件 ($F(36, 6030) = 1.04, p = .40$) のいずれの一次の交互作用も有意でなかった。また、二次の交互作用も有意ではなかった ($F(36, 6030) = 1.30, p = .11$)。それぞれの条件の主効果については、待ち時間条件 ($F(9, 6030) = 147.53, p < .001, \eta^2 = .082$) および現状満足度 ($F(4, 670) = 15.01, p < .001, \eta^2 = .044$) が有意であり、実査条件 ($F < 1, ns$) は有意でなかった。現状満足度での群分けによる、待ち時間に対する満足度の平均を Fig. 3 に示す。

Table 2 携帯電話によるウェブ利用頻度

利用頻度	第1実査	第2実査	合計
1日10回以上	121	214	335
1日6-9回程度	46	38	84
1日4-5回程度	58	11	69
1日2-3回程度	73	7	80
1日1回程度	23	4	27
週に1-3回程度	38	1	39
月に1-3回程度	17	0	17
ほとんど利用していない	29	0	29
合計	405	275	680

考 察

本実験では、開発したシステムを用いた実験方法の安定性を検討するため、異なるパネルに対して行われた二回の実査について結果の比較を行った。分析の結果、実査条件の主効果および交互作用は、同時に測定した属性パラメータの影響を考慮しても、5%水準で有意な差は得られておらず、提案システムを用いた評価手法による実験は、待ち時間に対する満足度の評価を行う上では、再現性のある結果が得られたと言えよう。

ただし、2要因分散分析の交互作用において有意傾向が見られており、特に11秒以上の長い待ち時間において実査間に違いが見られている。ここで、性別を加えた3要因分散分析の単純主効果検定において、第2実査での女性参加者の評価が、第1実査に対して低くなっていることから、両者が関連しあっている可能性も示唆される。しかし、2要因分散分析および性別を加えた3要因分散分析のいずれにおいても、効果量 $\eta^2 < .002$ と小さく、その影響は限定的であると考えられる。

それぞれの実査においては、性別・年代はある程度統制されており、通常満足度の分布にも差は無い。ただ、ウェブの利用頻度については違いが見られ、「1日10回以上」利用する参加者が、第1実査では約30%であるのに対して、第2実査の参加者は75%以上が集中している。今回の評価では、利用頻度の効果は見えていないが、「1日10回以上」利用する中での分布の違いが影響している可能性がある。実際、Niida, Harada, Uemura and Nakamura (2011) では、携帯電話を用いた電子メールの送信頻度の差が、待ち時間に対する満足度評価に影響を与えるという結果が報告されている。この点について

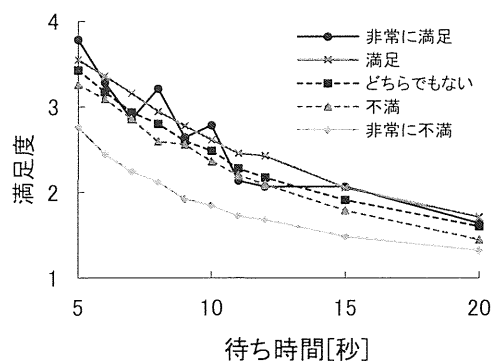


Fig. 3 各現状満足度における待ち時間に対する満足度の平均

ては、今後詳細な検討を要する。

現状満足度は、待ち時間に対する満足度との強い関係性を示しており、現状満足度の高い群ほど、待ち時間に対する満足度が高くなっている。このため、提案システムを用いた評価実験においては、こうした特徴について、募集段階で統制を行う、もしくは本項目を質問紙に含めるなどして、分析においてその影響を考慮する必要があるだろう。

本稿では、ICTによるネットワーク利用時の待つ行為を評価するための評価手法について、我々の提案手法により収集されたデータについて詳細な分析を行うことで一定の信頼性を示し、またそこで示される諸特性との関連性から待ち時間評価の特性を報告した。本手法では、利用者が普段ICTを利用している環境で実験が可能であり、かつ評価対象となる待ち時間を統制可能である。実験において同一の実験を異なる実験参加者を得て二回行い比較した結果、実験室実験と比較して生態学的妥当性の高く、多くの回答を比較的容易に集めることが可能な、有効な手法であることが明らかとなったといえよう。

ICT利用における待つ行為は、日常生活において頻度の高い事象であり、携帯電話サービスの品質およびそれへの満足度に強い影響を与える。また、サービス品質改善のために待ち時間を数秒改善する改修を行うには莫大な費用が発生するため、ICT利用時の待つ行為における認知プロセスの特徴を明らかにし、より効率的な技術の開発に繋げることは技術的・工学的にも大きな意義を持つ。こうした評価手法を用いてICT技術の使いやすさや満足度を評価する研究は、心理学的研究の応用先として非常に重要な意味が有ると考えられ、今後の発展が期待される。

引用文献

- 阿部威郎・石橋 豊・吉野秀明(2008). 次世代のサービス品質技術動向 電子情報通信学会誌, **91**, 82-86. (Abe, T., Ishibashi, Y., & Yoshino, H. (2008). State of the Art of Service Quality. *The Journal of the Institute of Electronics, Information and Communication Engineers*, **91**, 82-86.)
- 原田悦子(1997). 人の視点からみた人工物研究 認知科学モノグラフ6 共立出版. (Harada, E. T.)
- ITU (2005). Estimating end-to-end performance in IP networks for data applications. *ITU-T Recommendation G.1030*.
- ITU (2007). Definition of QoE. *ITU-T Appendix I to P.10/G.100*.
- Liu, Y., Ngu, A. H., & Zeng, L. Z. (2004). QoS computation and policing in dynamic Web service selection. *Proceedings of International World Wide Web Conference*. 66-73.
- 新井田統・原田悦子(2011). 高度情報化社会におけるコミュニケーションの分析と理解. 電子情報通信学会誌, **94**, 226-231. (Niida, S., & Harada, E. T. (2011). The Analysis and Understanding of Communication in Advanced Information Society. *The Journal of the Institute of Electronics, Information and Communication Engineers*, **94**, 226-231.)
- Niida, S., Harada, E. T., Uemura, S., & Nakamura, H. (2011). Effect of Using Frequency on valuation of Waiting Time When Using ICT. *Psychologia*, **54**, 96-106.
- 新井田統・上村郷志・中村 元(2010a). Web-scriptを用いた主観品質評価システムの時間分解能と再現性 電子情報通信学会2010ソサイエティ大会, BS-5-6. (Niida, S., Uemura, S., & Nakamura, H. (2010a). Time resolution and reproducibility of web-script based subjective quality assessment system. *Proceedings of the IEICE 2010 Society Conference*, BS-5-6.)
- Niida, S., Uemura, S., & Nakamura, H. (2010b). Mobile services -User tolerance for waiting time-. *IEEE Vehicular Technology Magazine*, **5**, 61-67.
- 大隈 昇(2002). インターネット調査の適用可能性と限界—データ科学の視点からの考察—行動計量学, **29**, 20-44. (Ohsumi, N. (2002). Internet Surveys: A Review of Several Experimental Results: Applying Data Science Approach to the Exploration of Internet Surveys. *The Japanese Journal of Behaviormetrics*, **29**, 20-44.)
- Uemura, S., Niida, S., & Nakamura, H. (2011). A Web script-based field evaluation method to assess subjective quality of mobile services. *IEICE Transactions on Communications*. E94-B, 639-648.
- 吉村 宰(2001). インターネット調査にみられる回答者像. その特性, 統計数理研究所公開講演会要旨 統計数理, **49**, 223-229. (Yoshimura, O.)

(受稿3月30日:受理5月7日)