

実践講座

ヒューマンインタフェース学会論文誌への投稿のすすめ 第2回 人間特性評価論文をどう書くべきか

関西大学 小谷 賢太郎

1. はじめに

本記事は前号に掲載された「ヒューマンインタフェース学会論文誌への投稿のすすめ第1回：論文誌と査読・編集プロセス」に続く記事である。HI学会論文誌編集委員会にはHIという広領域性から、様々な論文が投稿されてくる。したがって、一般にHI領域の論文とはどうあるべきかと論じるのは困難であり、いくつかのジャンル分けを行ったうえで、少しでも指針のようなものをまとめられればどうだろうという話が持ち上がり、この連載となった。モノづくり論文やケーススタディのような論文、そして、今回ターゲットとした人間特性評価論文など書き方にもそれぞれ特徴があると思う。次号ではモノづくり論文を中心とした書き方についての記事が掲載される予定であるが、本記事では人間特性評価論文、すなわち、システムやインタフェースに関わる人間の特性を得て、分析することにより、それらのシステムやインタフェースなどを評価しようとするもの、あるいはその評価手法・技術そのものを提案しようとする論文を指す。私自身はこれまで自身の研究テーマとして生体信号を計測してその特性をヒューマンインタフェース分野に利用してきた。本年度から本学会の論文誌編集委員長を務めている関係上、このような依頼を引き受けることになった。とはいうものの、ここに記す内容は一般的な内容が中心であり、それ以上のことは書ける実績があるわけではない（正直に言うと、ここにあるタイトルのような内容の記事なら私自身が真っ先に読まねばならぬと思っているのだから）。本記事ではHI学会論文誌編集のプロセスについての紹介と今後の論文誌の方向性の紹介も交えながら、現在のHI学会論文誌における人間特性評価論文について思うところを記したいと思う。

2. HI学会論文の専門性と人間特性とのつながり

HI学会論文誌編集委員会は現在25名の編集委員で構成されているが、毎年新たに編集委員会に加わっていただく委員にはキーワードリストを提示し、その中で各自の専門分野は何であるかを申告していただいている。これはもちろん様々な領域から投稿されてくる論文に対して適切な担当編集委員を割り当てるためのものである。編集委員は査読者（レビューワ）ほど担当論文について十分専門的なコメントが出せるほど専門性を有している必要はない。しかし、この新しい編集委員の専門分野の確認はその関連領域における専門家である2名のレビューワによって評価された査読報告書を「メタレビューワ」として適切に査読されたものであるかを十分判断できる専門性が必要なためである。キーワードリストを表1に示す。今回、どの程度の領

表1 論文誌編集委員の専門分野リスト（2011年度版）。

委員会はこの中から複数領域を担当する25名の編集委員から構成されている。

認知・知覚	生理
感性	感覚
表情	心理
ノンバーバル	マルチモーダル
アルゴリズム	マルチメディア
ネットワークメディア	エージェント指向
協調作業	バーチャルリアリティ
Augmented Reality, Mixed Reality	グループウェア
ユビキタス	実世界指向
ウェアラブル	モバイル
視覚障害支援	言語・記号・メタファ
福祉・障害支援	高齢者支援
コミュニケーション支援	共生・ユニバーサルデザイン
ユーザ行動・モデル	教育・学習支援
遊び・エンタテインメント	身体的インタラクション
インタフェースデザイン	ユーザビリティ
情報デザイン	環境デザイン
触覚・力覚	生体計測
入力方式・入力デバイス	設計支援
音・音響・音声	出力方式・出力デバイス
ディスプレイ技術	テーブルトップ
HRI/HAI	医療機器
技術・技能の伝承	BMI
アンビエントインタフェース	安心・安全
遠隔コミュニケーション	デザイン、思想
手書き入力、タッチパネル、書字	交通、ドライバ支援、モビリティ
画像処理	ゲーム、シミュレーション
異文化、多言語	ヒューマンエラー
子育て支援	災害、避難
デジタルサイネージ	ライフログ

域（キーワード）が人間特性に関連付けられているのかを示そうと挙げてはみたのだが、この表から見ると、そう単純に切り分けられないことが分かる。生体特性と直接的に関連する研究領域として、生理、心理、感覚、生体計測などが含まれている。キーワードの中には画像処理、アルゴリズムといった人間特性とは直接関連しないようなキーワードも散見する。では、本当にこれらが人間特性とは無縁な領域なのかというと、これらのキーワードを挙げている論文でさえ、人間特性を実験的に評価している論文を見つけることが出来た。実のところ、領域によって人間特性評価の必要性が変わるわけではなく、どうやら、投稿されてくる論文が研究全体のどのフェーズに位置づけられ、そのフェーズで人間特性を評価指標とするべきところに来ているか否かに依存していると判断できる。

そこで、本論文誌に採録される論文は一体どの程度実際に人間特性を扱っているのかを過去2年間（2009年11巻1号～2010年12巻4号）にさかのぼって調査することにした。調査するに当たり、まず評価指標となる人間特性を含む論文をそのタイプにより3つに分類した。一つは生体データ計測論文と名付けた。その名の通り、人間の生体信号を実際に計測し、評価に応用しようというものである。

HI 論文誌で取り扱われている生体信号の中には自律神経系(心拍、呼吸、血圧、脈波、皮膚温など)、中枢神経系(脳波を代表とする脳計測によって得られる生体信号が中心)、筋電図、眼電図、顔表情などの生体機能・特徴を表す信号などが存在する。第2の論文タイプを行動データ計測論文とした。このタイプには反応時間、作業成績、エラー率、誤認識率など行動に関連する指標をもとに得られた人間特性データからインタフェースを評価しようとするタイプとなる。第3の論文タイプを主観データ計測論文と名付けた。これはアンケート、主観評価と呼ばれるデータを扱って、インタフェースの適性やユーザビリティなどを評価しようとしている論文に多数みられる。もちろん、これら3つのタイプは場合によっては一つの論文の中で重複して評価されて、そのおのおのが関連付けられていることも多い。それらの論文については重複してカウントしているため、実際のタイプ別論文を3タイプともすべて合わせると全論文数よりも多くなっていることに注意してほしい。また、論文の中には単に開発したシステムの動作を確認するためにヒトが実際に試してみても問題ないかを確認しているような記述のある論文もあるが、このような論文は実際にヒトの特性を調べているとはいえないため、人間特性評価論文としてカウントしなかった。

その結果は図1の通りである。全92編の採録論文のうち、実に92.4%にあたる85編が何らかの人間特性を求めることにより評価していることが分かった。分類した3タイプの論文の中では行動データ計測論文が最も多かった。行動データはアンケートなどの主観データとは異なり、一定の客観性が存在するため、定量的な評価結果に信頼性を与えてくれる。また、作業成績やエラー率といった指標は直接インタフェースの効果につながるため説得力を持つ結論を導くことができる。主観データ計測論文がその次に多く、最も少ないのは生体データ計測論文であった。生体信号は行動データや主観データと比べてデータ取得に一定のノウハウが必要であることや、生体信号の変動が具体的にどういった意味を持つのかという点の多様性が解釈の困難さにつながっているためであろうと思われる。

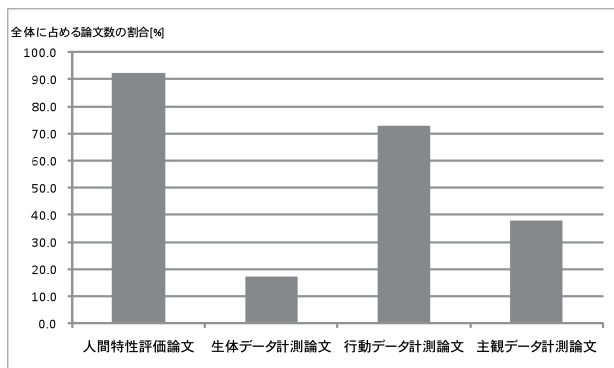


図1 過去2年間に採録された論文誌全体に占める生体計測を含む論文の割合(生体データ計測論文、行動データ計測論文、主観データ計測論文はそれぞれ相互に重複を含む)

残りの7.6%である、生体計測を含まない論文には一体どのようなものがあつたのかというと、その多数は新しいシステムやインタフェースの報告であった。これらの論文はヒューマンインタフェースに関連する原理やシステムの開発を中心として報告したものであって、それらの論文の切り口は、研究の初期フェーズであり、実装したシステム自体のパフォーマンス(計算速度や、アプリケーションの紹介)を示すことがそれらの論文を示す有効性となっているようである。したがって、それらのシステムを利用するヒトの側面を紹介するフェーズは今後の方向性として示されている場合が多い。

こういった論文の多くは、その種別としては技術報告に分類されるべき論文であるともいえる。論文の種類の違い(原著論文か、あるいは技術報告か)は、その定義以外に種別のもたらす印象、すなわち、査読プロセスの違い、読者にとっての受け止め方の違い、あるいは究極にはその論文自体の評価、には影響があるかもしれない。現在のところ、論文種類に対する査読の違いは規定としては存在しないし、ガイドラインもない。したがって、両者の種別は著者の申告により決定されるカテゴリーである。この点については今後論文誌編集委員会で検討すべき課題であろう。

3. 人間特性評価論文をわかりやすい論文にするために

以下に、これまで投稿されてくる論文に目を通したり、自分自身が読者として論文に不満を持った経験から、特に人間特性評価論文についてそのわかりやすさを高めるために留意すべき点について述べたいと思う。

●実験の手順やデータ解析手法を正確に詳細に表現すること

例えばあるモデルを示そうとする場合、そのモデルは数式などによって正確に表現できる。ところが、人間特性を用いて評価するためにおこなう実験にはその表現があいまいになる傾向がある。したがって、その実験がどのように行われたのかを読者に明瞭に分かるように示すことが、人間特性評価論文に信頼性を与える観点から特に重要になる。よくみられるケースとして、論文では実験装置などは比較的詳細に表現されている場合が多いが、実験の手順はあいまいで不完全な記述になっている論文が多い。投稿された論文の著者の中にはかなりの割合で、「読者(あるいは査読者)は実験の結果に関心があるわけだから、ていねいに結果を解釈すれば(あわよくば結果を大きく解釈してくれば)よいので、これを意識して論文を書こう」と意識していることが読み取れる論文が少なからず存在することを実感している。読者の中には実験結果よりも実験の手順やデータ解析の実施方法に関心を抱く読者がたくさんいることを知ってもらいたい。実験手順に関する情報はあまりにも多様すぎて、もれなく記述するチェックリストのような手法は存在しない。データ解析の手法も、きっと皆同様の手法でやっているに違いないと思込んでいる著者が多くと予想される。したがって、著者自身が十分記述しているだろうとの自己判断に基づく記述内容となっており、それらは往々にして不十分である。査読者は専門分野が近いこ

とから、実験内容が手に取るように見えることも多いため、比較的实验手法、データ解析手法の不十分さを見逃してしまう場合が見受けられるが、編集委員にはそのような不十分さが論文の価値を下げてしまっていることに気づくことがよくある。HI学会のような広領域の読者をもつ論文誌にとって、新しい実験のアプローチが明瞭に理解できなければ、その論文は読者にとって役に立たない場合が多い。できればコアな専門家だけでなく、異なった領域の専門家にもその手法論が可視化されるよう気を配るべきである。最も適切な改善策としては投稿前に「好意的な査読」を周囲の研究者にってもらうことだと思う。

●写真や図の明瞭さについての配慮

編集委員の立場で論文誌に目を通して特に感じることは効果的な図、写真類の活用の乏しさである。実験の写真に気を配らず投稿してくる論文がなんと多いことだろう。実験の風景(のようなもの)を載せて「図1.実験風景」としたりしているが、これでは読者にとって得られるものは少ない。写真が対象としている「伝えたい内容」に十分吟味したものを選んで載せてほしい。実験の雰囲気のようなものを伝えるメディアとして、論文誌はあまり適切だとは言えないので、実験内容を視覚的に伝える対象を写したものを明瞭に示すよう心掛けてほしい。さらに実は大変読者の理解を促すのに重要なものに、写真のクオリティがある。映りの悪い写真は結局何も伝えることができない。国際誌では図や写真が相応のクオリティを満たさなければ、受け付けてもらえない電子投稿システムをもつジャーナルも多く、編集委員も写真のクオリティの悪さに対しては積極的にリジェクトしようとする傾向があると私は感じている。写真の選択については、もっとシビアになるべきである。特に実験の様子などは後から再現できない。例えば査読後に返却されてきたコメントの中に「もっとクオリティの良い図を載せるように」と、指摘があっても、もうすでに実験環境はすでに片づけられて何も無い、といったこともあるだろうから、余計実験時には気を配らなければならない。また、カメラレディ原稿などでページ数を抑えようと、安易に図のサイズを調整して、判別性を著しく失っているのに配慮しない著者もいる。当然ながら、そのように縮小された図は必要な情報をすでに失っている。今後はHI学会では論文誌が電子化されるため、論文に掲載されるカラー写真にも標準的に対応が可能になり、これまでよりも図や写真を用いた視覚表現の重要性は高まる。現在のところはページ数をもって掲載料金を定めているため、たくさん写真や図を載せてほしいとは安易には言えないが、電子ファイルの容量には制限を設けていないので、クオリティの高い図をどんどん使用して可読性を高めた論文にすることをもっと意識していただきたい。

●被験者に関する情報と適性に対する検討

被験者特性はデータの性質を決める重要な要素である。よくみられる記述に「被験者は男子大学生10名であった。」この一文だけで「2.1節 被験者」の項が終わっている場合さえある。これではデータの信頼性に疑問を持たざるを得ない。

被験者としての評価の適性、正当性については明確に記述すべきである。またもちろんであるが、男子大学生であれば一般性が保たれるとも限らない。ユーザビリティ評価などでは被験者となるユーザの経験や興味の対象などが異なれば、評価実験のデータに異常値を与える可能性はある。被験者の適性については評価結果の制約範囲として論文中に考察してあれば、データの意義を把握しやすくなり、論文自体の信頼性は高まる。また、考察に感度分析の結果を与えると、実際にシステムの実用化に際してイメージしやすくなる。対象となるユーザと実験に参加していただいた被験者の母集団は等質であろうか。この点については考察を加えてほしい。これらは、実験結果として得たものを過度に一般化することを防ぐことになる。また、実験の方法が実験者と被験者の関係がシングルブラインド性かダブルブラインド性の2つが存在するが、この点もその後のデータの解釈を重要にする情報なので、どちらであったのかは述べられるべきである。被験者の男子大学生はあなたがどういう実験結果を望んでいるのかを知ってはいなかっただろうか。

最後に、被験者を伴う人間特性評価実験で重要な点であるが、最近では倫理的配慮や被験者の個人情報保護の問題がクローズアップされ始め、HI学会でも最近これらを反映して論文誌投稿規定に修正が施されている。読者の所属でも倫理委員会などが立ち上がっていて、実際の実験データはどのように処理され取り扱われるか、実験内容と実験から受ける危険の可能性などは被験者に十分説明がなされているかなどが議論され始めているケースが多いと思う。こういった点に配慮がなされていることについても論文中に記述があると、実験のデザインが適切に行われていることを明確に伝えることができる。

●ヒトはあいまいなもの

私の研究室では主に人間特性を計測しているが、時には自作の実験装置の特性試験のデータをとることもある。そのような力学特性のような見事にきれいなカーブを描くデータを学生に見せられると、普段はノイズが多く含まれる人間特性データを見慣れているだけに、感嘆させられることがある(かといって、ノイズに一見埋もれた人間のデータを見るのが嫌いなわけではないのだが)。良きにつけ悪きにつけ人間から得られる特性には概してノイズに隠れているデータが得られることが普通であり、そのようなデータに信頼性を担保するためには統計情報は重要である(どのように得られる人間特性データのS/N比を良くするかも人間特性評価の上ではきわめて重要なのであるが、これらの方法論の詳細は他に譲る)。記述統計だけではなく、検定を用いて、得られた情報の信頼性を付加することで、論文の説得力は大いに増すことになる。統計学的有意差を表すためには検定を行うが、たいいていの論文では適切に第一種の過誤については適切に求められている。これは人間特性で差のあることを明示したいという著者の意識の表れかもしれない。しかしながら、第二種の過誤についてはあまり適切に表現されている例は少ない。HI領域の研究では、何か新しいシステムやプロトタイプを利用した時のユーザ特性

が現行のシステムと比べて差がないことを示したい場合があるが、そのような際には検出力を意識せずに、両システム間には差がないことを主張する場合がある。このような場合は被験者数（サンプル数）が十分に確保されていない場合が多い。図2は図1で示した過去2年間の論文誌に採録された人間特性評価論文のうち、実際に何名の被験者からデータを得たのかを人間特性評価論文全体と3つの論文のタイプごとにまとめたものである。最小被験者数と最頻被験者数（平均ではなく、何名の人数を被験者としたケースが多いかを調べるために、モードを用いた）を表している。この図を見る限り、人間特性評価論文では10名の被験者からデータを得た論文が最も多い。最小被験者数は2名であった。被験者数を多く取ること、すなわち、被験者数を母集団数に近づくこと、で得られたデータの平均値の精度が高くなるだけでなく、第二種の過誤も小さくすることができる。図でわかる通り、被験者数は論文のタイプにより差があることが分かる。3つのタイプのうちで、被験者数の最も多かった主観データ計測論文では、10名の被験者からデータをとっているケースが最も多かった。しかし、これは主観データの計測方法の簡便さから考えると、評価実験としては被験者数が少なすぎるのではないかというのが私の感想である。最も少なかったのは生体データ計測論文であり、多くの論文では被験者は4名にすぎなかった。被験者数は、提示された仮説に対し、信頼のおける解答を与えられるよう十分多く設定するべきである。誰もが統計的には被験者数を上げること、検出力、すなわち第二種の誤りを起こさない確率、を高めることが重要なことを知っている。しかしながら、実際の被験者数は時間や実験条件の制約、評価対象にもたらず評価結果の意義に左右するプラクティカルな要因により決められることが多い。ちなみに、今回の調査範囲で最も少ない被験者数だった論文はわずか2名の被験者での評価であった。1名のデータのみで評価を行おうとしている論文はさすがにこの2年間にはなかった。当然の結果であろう。主観データ計測論文、行動データ計測論文、生体データ計測論文の順に被験者数が少なくなっているのも、実験の困難さや、データ取得に要する時間・努力

を考えれば予想通りといえる。被験者数、あるいはデータ量は評価実験の信頼性を高めるため重要なのであるが、当該論文の論点が評価実験の信頼性にないものもある（とりあえず、人間特性実験で評価してみて、問題なく効果を示すできそうだという主張のためにのみ使われるケース）ので、一概には言えない。つまり、自分の論文が少ない被験者でしか評価できなかったのであれば、それは一般のユーザを対象とするものであるという主張は取り下げて、別のポイントで論文の有効性や信頼性を主張すべきであろう。

4. おわりに

この「投稿のすすめ」の第2回、昨年度の編集委員長である加藤先生から現編集委員長という流れで私に執筆担当が回ってきた際、「いいですよ」と安請け合いましたものの、脱稿を迎えるまでは正直苦しい作業であった。少し書いては「こんなこと書いて意味あるかなあ」と、自問自答の末、書きなおすという繰り返しであった。いま読み返してみると、もっと網羅的に書けたのではないかと、参考文献をつけてしっかりと裏付けを明瞭に示した文章の方がよかったのだろうかなどと、あれこれ反省している。こういったある種リテラシーともとらえられる情報はHI学会の会員という母集団にはターゲットが広すぎて狙いが定まらないなあと感じた次第である。会誌委員会の皆様にはなかなか原稿を送ることができずご迷惑を書けてしまい、紙面を借りてお詫びしたい。ただ、私としては、少しでも読者の皆様が人間特性評価論文を執筆する際に読者にとって有益な情報として提供できるかを考えたり、また、この記事を読んで「それなら、次は書いてあることを意識してHI学会論文誌に投稿してみよう」（あるいは、「こんな程度の記事を書いている編集委員長なら、私の論文でも簡単に採録されそうだから投稿してみよう」でもよい）と考えたりする読者が一人でも増え、論文誌が活性化されることを祈っている。今後HI学会論文誌は次年度からの論文誌電子化を皮切りにマルチメディア化など、新たな方向性について現在議論を行っているところである。今後のHI学会論文誌に期待してほしい。

著者紹介



小谷 賢太郎 (こたに けんたろう) :

1996年ペンシルバニア州立大学ドクターコース修了。同年4月より関西大学工学部助手、2005年ハーバード大学客員研究員。2010年関西大学システム理工学部教授。関西大学医工業連携研究センター長。HCI、生体情報工学に関する研究に従事する。ヒューマンインタフェース学会理事、日本人間工学会理事、日本生理人類学会理事、International Ergonomics Association Council member、計測自動制御学会等の会員。Ph. D.

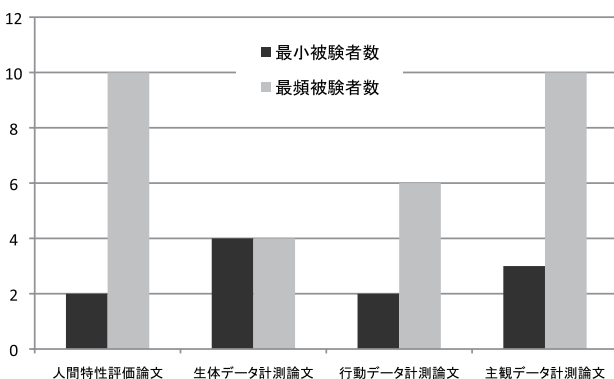


図2 過去2年間で採録された論文誌において対象となった被験者の数（人間特性評価論文全体、および論文タイプ別）