

インタラクション研究でのシングルケース実験についての考察

宮下芳明^{†1}

本稿ではまず、インタラクション系の学会発表予稿集に求められる評価実験はどうあるべきかを考え、2群比較実験法の適切性、平均操作の弊害について論じている。そのうえで「シングルケース実験」を用いることが有益ではないかと考え、その歴史や実験デザイン、他分野での研究事例を紹介している。最後に、インタラクション研究の一事例として、音楽演奏支援システムの評価実験をとりあげ、シングルケース実験ならではの恩恵を享受できているか、方法論に照らせばどのように実験デザインを改善できるかを議論し、向いている分野やシステムについて考察した。

1. はじめに

本稿では、被験者が1〜少数で、群間比較を行わずに経時的な変化を検討するという、シングルケース実験について考察し、本シンポジウムをはじめとするインタラクション系の学会やワークショップで活用できないか議論します。

1.1 インタラクション系の学会発表予稿集に求められる評価実験はどうあるべきか

たとえばこのエンタテインメントコンピューティングシンポジウム予稿集に掲載されている研究は、どれほど評価実験を行っているのでしょうか？論文誌レベルの厳密さをもった実験を行っているものもあるかもしれませんし、被験者にシステムの良さを7段階評価で尋ね平均値を比較しているような実験もあるかもしれませんし、あるいは、全く評価実験を行わずに実装報告を行っている原稿もあるかもしれません。

もし学会が速報的にプロトタイプシステムのデモや議論を行う場であり、発表後、それをより良い手法に改善し、いかに有用かを実証し、査読付き論文誌に投稿していく…という流れであるならば、「主観的な設計思想の説明+実装報告」という発表は勿体ないかもしれません。せっかくシステムを実装したにもかかわらず、実際にうまくいくかどうか試していない（試したデータを見せていない）というのでは、学会参加者は結局、設計思想についてあれこれ議論するしかできません。だからと言って、論文誌レベルの厳密さ…無作為配分等による剰余変数統制やその他の条件を整備し、十分な被験者数を確保して、独立変数と従属変数の間にある因果関係を確実に特定せよというの、上述の学会発表の側面にそぐわないのでないでしょうか。論文誌投稿の前段階で行われる学会発表では「どう評価実験を行ったら良いか」といったことも議論対象であるはずなので、そう考えると、論文誌レベルの評価である必要はないとも考えられます。

インタラクション系の研究においては、「何を作るか」

「どう作るか」を練り、大概是著者・共著者がそのプロトタイプを実装し、プロトタイプやコンポーネントを開発者自身が試して改良するプロセスを何回も回しています。そこにはかなりの労力と時間を割き、「良い手法である」と著者らが実感できているから速報的に発表したいと思っているわけで、厳密な評価に労力と時間をかけられない、という本音もあるのではないかと思います。しかしその一方で、本当に意図された効果をもっているどうかはやはり何らかの客観的な評価がないと信用されませんし、それを示すことが研究者の責務でもあります。

学会発表される手法が「プロトタイプ」として理解され議論対象になるのならば、実験デザインやその結果についても、プロトタイプとして受容できるはずですが、システムをよりよくするためにはどうしたらいいかという議論に加えて、実験デザインや分析方法についての議論も行うことができます。建設的な展開のためにも、荒削りでも良いので何らかの実験とデータが発表される方が、投稿者・参加者ともに有益ではないでしょうか。

1.2 2群比較実験法は適切か

さて、HCI分野においては、心理学の実験手法として正しさが保証されているやり方で「提案システム」「従来システム」の実験群（処置群）・統制群（対照群）の比較が行われることが一般的です。多標本実験計画法における、いわゆる2群比較実験にあたります。ただ、前節で述べた文脈での学会発表予稿集に求められる評価実験という観点だと、2群比較による実験は不向きな点も多いです。

そもそも2群比較実験は、「対象となる標本が無作為に抽出されたもの」「それらの測定値が正規分布をしていること」などの条件が満たされていることを前提にしているわけですが、こうした前提条件をちゃんと満たすことは難しいことが多いです。研究によっては、そもそもターゲットユーザに適合する被験者をそれだけ集めることが困難なケースすらあるでしょう。また、測ろうとするものによっては、実験開始時から被験者のレベルがバラバラ、システム

^{†1} 明治大学 総合数理学部 先端メディアサイエンス学科
Department of Frontier Media Science, School of Interdisciplinary Mathematical Sciences, Meiji University.

における反応の仕方も人それぞれであったりすることがあります。そういった被験者を「群」として扱い、比較して統計的な有意差を出すことは原理上困難になります。「被験者を増やせば、有意差が出るようになる」という意見もよく聞かれますが、大数と統計に頼るあまり、ますます人間ひとりひとりを観察しなくなるのではないかと危惧してまいります。

1.3 平均操作の弊害について

筆者は、「少数の例外を無視する結果になる」という平均操作の側面にやや乱暴さを感じますし、平均するという行為に恣意性を感じることもあります。平均を行うと、個々のデータの傾向がつぶれてしまうことが多いです。たとえば7割ぐらいが増加傾向、残り3割が逆に減少傾向だったとしても、グラフで平均すると増加傾向の単調増加のグラフになったりします。

図1 上部グラフは、4羽のハトにおける、個別別のキーつき反応の遂行経過図です。実験箱の中の照明状況に従って反応数が変化していますが、これを平均すると図1下部のようになり、個別のグラフから議論できた事柄が多く埋もれてしまいます。

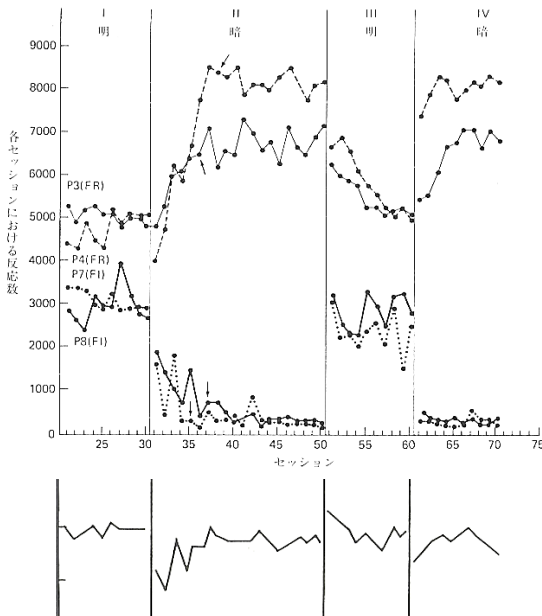


図1 4羽のハトにおける個別別のキーつき反応の遂行経過図（上部）と平均（下部）（[1]より引用・配置改変）

図2のグラフは、A・B二人の発達曲線の例で、どちらも、ある年齢から徐々に発達し始め、ある年齢で最も勾配が大きくなって、その後は次第に緩やかになるという形状をしています。ところが、これらの平均をとると、元のふたつの曲線と異なる特徴をもった曲線になってしまっています。ある年齢から徐々に上がり始め、いったん勾配がゆるやかな時期がしばらく続いてからもう一度勾配が大きくなるかのような意味が生まれてしまっています。

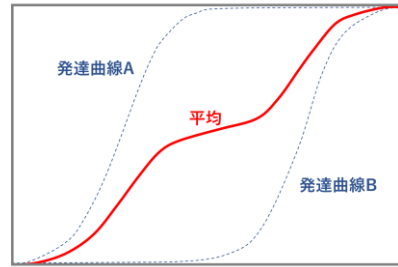


図2 個人A・Bの発達曲線（青点線）と平均発達曲線（赤実線）（[2]より引用・着色）

被験者が多くても平均することによって異なる意味が生まれることがあります。図3のグラフ左側は、10人の被験者における学習曲線です。全被験者において、どこかのトライアルで一気に学習成果が出る、いわゆる悉無律が顕著に表れています。ところが、これらを平均すると図3右側のようになり、個人ごとに位相がずれているがために、ゆるやかなS字状の学習曲線として表されてしまいます。

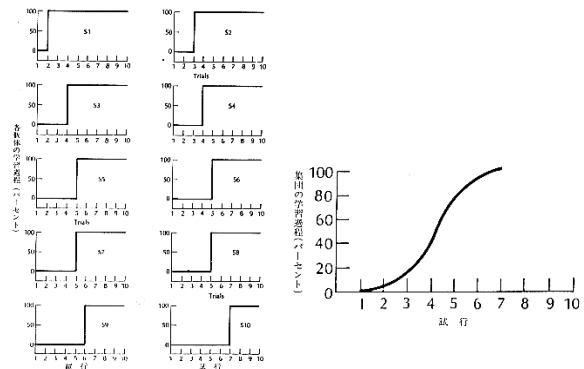


図3 学習経過が悉無律となっている10人の被験者の学習曲線（左）、およびそれらを平均した学習曲線（右）（[1]より引用・配置改変）

つまり平均するという行為は、ものごとを見えにくくするどころか、異なるものを見せてしまうことがかなりあるのです。現象の本質にちゃんと向き合うためには、ひとりひとりを見る必要があります。仮説通りの反応を示す被験者も、個別に観察すればタイプが違う人がいるかもしれません。仮説通りの効果が現れない人たちについても、個別にデータを詳細に見れば原因が究明できるかもしれません。それに基づいてシステムを改良すればその人たちにも効果を及ぼすことができるかもしれません。そして、仮説と逆の反応を示す被験者はさらに興味深い知見につながる可能性があります。少数派のデータをノイズとして扱い、平均でつぶすなんてとんでもない、と筆者は思っています。

異なる人の集まった群を比較するのではなく、「同じ人」を見ることには意義があります。しかしやっかいなのは、人間は勝手に学習し、学習してしまうと元の状態には戻れ

ない不可逆性をもつことです。もしタイムマシンがあれば、従来システムを被験者に使わせて記録をとったあと、実験開始時刻までタイムトラベルし、今度は提案システムを与えて、従来システムと提案システムを使ったふたつの「世界線」を比較したいわけです。ですが、そんなことは実現のしようがありませんし、だからこそ異なる被験者グループによる 2 群比較の手法があるわけなのですが、「同じ被験者で」それをなんとか実現するのがシングルケース実験です。たとえば、被験者が従来システムで習熟に苦しむ状況があり、提案システムによる支援でそれを打開できるとしたとき、その経時変化を観察すれば、観測不能なふたつの「世界線」からその差異を観察し、支援効果を推論できるものです。

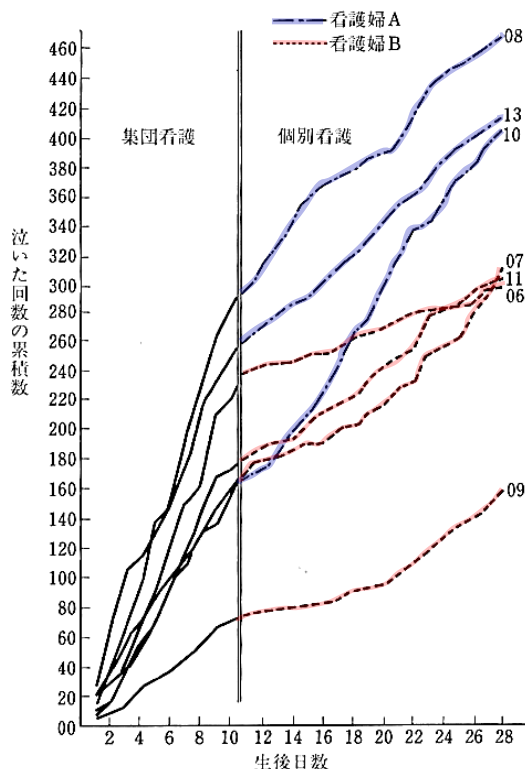


図4 新生児の泣く回数に及ぼす保育担当者の効果 ([1]より引用・着色)

図4は、新生児の生後2週間目における保育担当者の相違が、泣く回数に影響を与えるかを検証したグラフです。横軸は生後日数、縦軸は泣いた回数の累積数、折れ線グラフについての番号は新生児につけられた識別番号です。この図から、看護婦Aに比べると、看護婦Bによって世話された新生児たちの泣く回数が少ないことを見るができます。新生児はもともと個体ごとに泣く頻度は数倍も異なっていますが、こうして個別にみることで考察ができます。実際にタイムトラベルすることはできませんが、「もし生後11日目に新生児08・13・10を看護婦Bのほうに担当させたら」という世界線がどうなっていたかを推論することはできるというわけです。

2. シングルケース実験

シングルケース実験は、他の呼称としては単一事例実験、N=1 実験、少数n計画などがあり、一つの事例あるいは一人の被験者を対象にして、途中から提案システムを使わせるといった介入を行い、その前後の変化をみて介入が有効だったかどうかを調べる方法、およびその応用を指します。統計的手法に頼らず、グラフの読み取りによって解釈します。[2]によれば、「処遇の結果がグラフから一目瞭然であれば、特に統計的分析を加えなくてもよい」というのがシングルケース実験における基本的な了解事項となっているようです。

2.1 歴史

「シングル・ケース研究法」[1]には、15 ページ程度にわたってその歴史が記されています。もともと心理学における実験は 19 世紀までシングルケース実験として発展しており、たとえばエビングハウスによる忘却曲線など、この実験手法によって導かれた知見が心理学の礎を築いていました。ところが 20 世紀にフィッシャーの推測統計学が登場し、処理条件の異なる群間比較実験が中心となりはじめ、1916 年時点では心理学の分野に群間比較実験はほとんどなかったのが、1951 年にはほぼ半数、1962 年には 9 割以上となりました。しかしスキナーらは多標本実験計画法のやりにくさを訴え、他分野の研究者と連携してシングルケース実験の意義を「実験行動分析誌」として発刊します。それ以来、オペラント心理学や行動生理学などの学問分野が、シングルケース実験を武器として勢力を伸ばし、1960 年代後半には心理学会全体から認知されるようになったそうです。国内では、書籍「シングル・ケース研究法」[1]にこうした方法論が 1990 年にまとめられています。

とはいえ、筆者が心理学実験に関する書籍をいくつか調べてみた限りにおいては、「シングル・ケース研究法」に書かれてるほどの市民権を得ていないように思えます。書籍「心理学研究法」[3]では、「単一か少数である被験者・被験体がたまたま特殊という可能性もあるため、そこで得られた知見をただちに一般的なものとはとらえられないという問題がある。この実験の方法論は成熟しているが、心理学全体の中では、やや限定された分野で用いられているように思われるので、ここでくわしくは述べない。」と記されています。一方で「心理学研究法入門」[2]では、「心理学の歴史においては、ただひとりだけの被験者（ときには研究者自身）を対象にした実験的研究のほうが長い歴史をもっており、重要な成果を上げてきた。」「臨床や教育の分野の研究、とくに行動療法とよばれる治療アプローチの適用と効果の評価において用いられている」と記され、詳細な説明がなされています。

2.2 シングルケース実験のデザイン

シングルケース実験のデザインについては、[1]で網羅的に紹介されており、反復型実験計画、非反復型実験計画、基準移動型実験計画、交替操作型実験計画、その他の実験計画といったかたちで解説されています。[2]においても要点が押さえられているので、その説明をここに記すと、シングルケース実験は、従属変数の継続的な測定を行い、特定の処遇を導入することの効果が測定値に反映されるかどうかを検討するところにポイントがあります。実験デザインの出発点は AB デザインで、処遇の効果をはっきりと検出できるように、処遇導入前の測定値の推移（ベースライン）を安定させるための実験統制を積極的に行います。この処遇を与えない時期をベースライン期と呼びます。ベースラインが安定してきたらそこで処遇を導入し、処遇期に移行するというわけです。もちろん、うまくいったとしても他の可能性がいくつか危惧されます。たとえば、処遇期への移行の前後に処遇の導入以外の出来事があり、それが変化の原因になっているのではないか？（履歴の脅威）とか、時間の経過による自然変化ではないか？（成熟の脅威）といった可能性への対処は十分ではありません。そこで、こうした脅威に対する防衛力を高めて、研究の内的妥当性を高めるデザインが以下の ABA デザイン、ABAB デザイン、多重ベースラインデザインです。

2.2.1 ABA デザイン・ABAB デザイン

ABA デザインとは、処遇期の後に、もう一度ベースラインに戻してみる手法です。もとのベースラインの水準に戻ったとしたら、これはかなりの説得力をもって「処遇によって引き起こされた」と主張できる、というわけです。また、ABAB デザインというのは、2 度目のベースライン期あとのさらにもう一度処遇期を置き、処遇の結果を再確認するとともに、好ましい行動の定着をはかる手法です。二度目の処遇期は、好ましい行動がある程度定着したと思われる時点まで続けられます。また、その後処遇を除いても行動水準が維持されるかどうかを、一定期間後の追跡調査で確認します（図 5）。

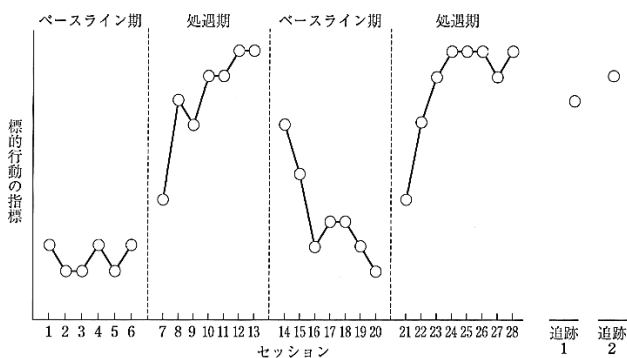


図 5 ABAB デザインを用いた結果の例[2]

2.2.2 多重ベースラインデザイン

とはいえ、人間の行動には、「いったん行動が変容したら処遇や他の要件を除いても行動が元に戻らない」こともよくあると思います。たとえば自転車に乗れるようにする何らかの訓練法ですと、ベースライン期と処遇期の比較は簡単ですが、いったん処遇期を経て自転車乗りになり成功してしまつたら、そこで訓練を中止したところで、また自転車に乗れなくなるわけではありません。こうしたときには、複数のベースラインを対象にし、他の被験者において、処遇導入時期を変えて検証する「多重ベースラインデザイン」などが行われます。異なる被験者に異なる時期に処遇を導入した結果、どの被験者についても処遇導入後に顕著な効果が見られたとしたら、そこから説得力のある推論ができるというわけです。

2.3 実際の研究事例

図 6 は、学級における二人の問題児に対して、命令・無視・賞賛という指示の仕方、およびその組み合わせを実施した論文[4]からのもので、横軸は 2 日ごとのセッションブロック、縦軸は問題児が不適切な行動を起こす割合を示しています。教示の仕方は、ベースライン - 命令 - 命令&無視 - 命令&無視&賞賛 - ベースライン - 命令&無視&賞賛となっています。

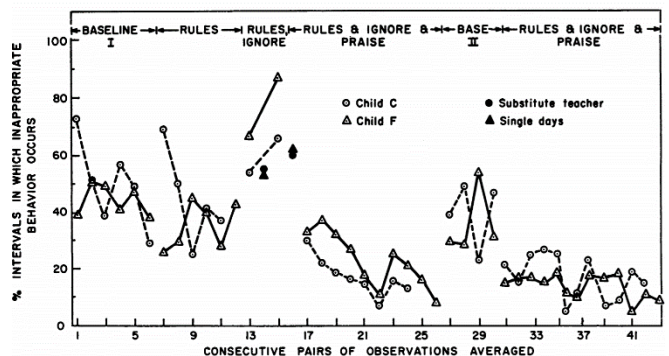


図 6 学級における二人の問題児の、指示の仕方と不適切な行動が起こる割合の関係[4]

いわば「A-BCD-A-BCD 型」とよべる複雑な事例ではありますが、[1]によれば、これはややデザインが不十分であると指摘されています。まず、ベースラインの省略は危険と記されています。もし第二ベースライン期に復元が起こらなかった場合は得られた結果が 3 つのどれに起因するか明らかにできないので、必ずその都度ベースラインを挿入すべきであると書かれています。また、2 種類の変数を同時に提示しても良いが同時に変化させないように注意しなければならない、とも記されています。

3. インタクション研究への適用例と考察

このように、シングルケース実験の事例をみていると、インタクション研究において、支援システムにおけるフィードバック機構をどうしたらよいかとか、ゲームにおける効果音やスコア表示がモチベーションにどのような影響を与えるかといった分析に使えそうに思えます。もともと、インタクションとかエクスペリエンスというものは個人的なものであり、メディアとしての特性もマスではなくパーソナルなものであることが多いので、分野的にも向いていると思われます。そこで、本章では、かつて研究会報告で行った実験を例に、そこからわかることや、シングルケース実験の教科書に基づくどのような改善点があるかを考察します。本稿で事例としてあげるのは、温度を用いた即興演奏支援システム Thermosupport についての評価実験です[5]。内容およびデータは既発表ですが、実験手法の適用という観点から再執筆しています。

3.1 支援システムの概要

この支援システムは即興演奏を対象としています。即興演奏においては、演奏者は音を出し、全体の響きを聞いたうえで次の音を決するというプロセスをとります。そのなかで、初心者が特に困難としているのは「全体の響きを聞いて理解する」という過程 (Representational Talkback) です。支援システム Thermosupport (図 6) は、この Representational Talkback を強化すべくデザインされたもので、ペルチェ素子搭載鍵盤を加熱する (45°C前後) ことによって、不協和度の高い音を演奏者に提示します。耳でとらえられる響きに加え、指先から伝わってくる熱によってもそれが強化されるというわけです。演奏者は鍵盤に触れたときにのみ温度を知覚するので、結果としてその音に関する情報だけを選択的に、そして打鍵直後に得ることになります。光るキーボードのように事前提示でないところひとつのポイントになります。

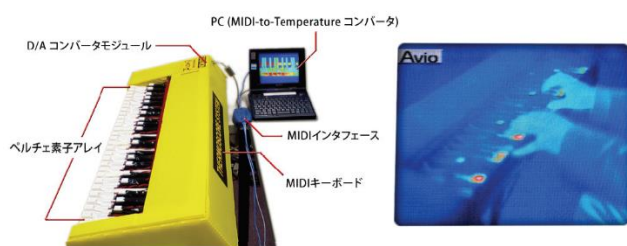


図 6 支援システム Thermosupport とサーモグラフィ画像

3.2 評価実験のデザイン

このシステムの評価実験は、シングルケース実験に非常に類したことを行っています。何も支援を行わない鍵盤楽器と、入力音を自動的に補正するシステム(典型的な即興演奏支援システム)、そして提案システム Thermosupport を切

り替えていく実験で、その経時変化を被験者ごとに記録していくというものです。

実験に際しては、即興演奏の未経験者に、伴奏に従って即興を試みてもらいます。最初は支援のないシステム A (ベースライン)、次にアベイラブル音以外の音(アボイド音を除く。以下アウト音とよぶ)を近くのアベイラブル音に自動補正するシステム B、そして最後に、アウト音の鍵盤を加熱するシステム C Thermosupport と切り替えていきます。なお、打鍵前に鍵盤温度を探る行為をインストラクションで禁じています。システムはソフトウェアとして内部的に切り替えられます(図 7)。伴奏には、循環コードを 8 小節で繰り返す分散和音を用いています。各システムでの演奏時間は 4 分 16 秒 (60bpm、64 小節) となります。

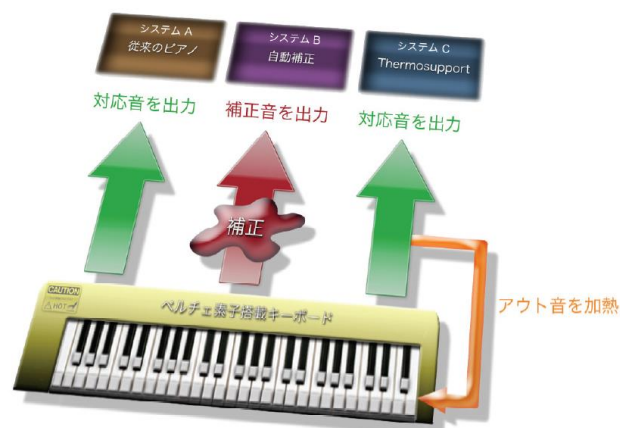


図 7 実験で比較する 3 つのシステム

3.3 実験結果の分析

結果は図 8 になります。図 8 上部の縦軸は、アウト音の打鍵頻度であり、いわばエラー数を示します。つまり、折れ線が降下するほど、伴奏に協和した演奏が行えていることとなります。図 8 下部は、打鍵数を単純にカウントしたのものになります。

実際に実験を行うと、何人かの被験者は循環コードを一度聞いた後に即興をはじめの傾向がありました。そのため、データの取得範囲はその次の 2 ループ分 (9~24 小節目) と最終 2 ループ分 (49~64 小節目) としています。

システム A のところを見ますと被験者 2 と 7 がアウト音を高い頻度 (約 80%) で打鍵しています。また、システム B では被験者 2 と 7 に加えて、3 と 5 もアウト音を高い頻度で打鍵しています。なお、記録を調べると、これらの被験者は白鍵を打鍵しがちな傾向をもつことがわかりました。実は、この実験の伴奏となる循環コードにおけるアベイラブル音は、多くの黒鍵を含んでおり、もしランダムに白鍵のみを演奏すると、その音がアウト音となる確率は 75% になります。同様に、黒鍵・白鍵に関わりなくランダムに演奏が行われた場合は、上記確率は 50% となります(図 8 点線部)。

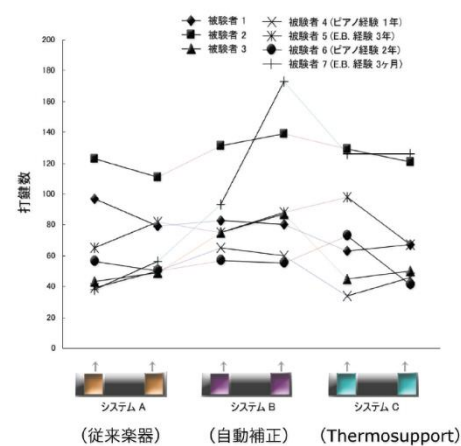
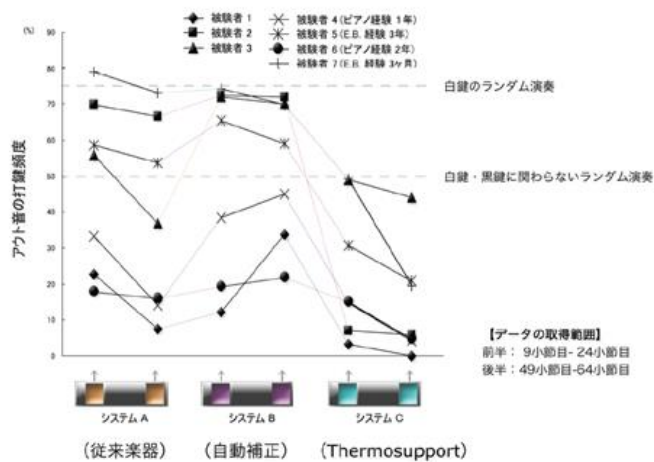


図8 実験結果

(上：アウト音の打鍵頻度推移 下：打鍵数推移)。

システム A についてみてみると、被験者間の個人差はこの時点でとても大きく、たとえば「ピアノ経験者が他者よりアウト音を打鍵していない」ことがみてとれます。そして、アウト音の頻度は次第に減少し、この傾向は例外なく全ての被験者にみることができます。即興演奏の未経験者であっても、自分の演奏と伴奏の協和度のある程度は聴感的に把握し、どのような音その調性感を崩すのかを経験を通してゆっくり学習できていることを示していると考えられます。

しかし、これがシステム B に切り替わったとたん、ほとんどの被験者についてアウト音は増加し、システム A よりも高くなっています。システム B を使ううちにさらに増加した被験者も 3 名います。

システム B は、アウト音を自動的に補正するので、どの鍵盤からもそれらしい音が出ることが出来ます。打鍵している音がアベイラブルか否かを考える必要はありません。したがって、たとピアノ経験者であっても、より多くアウト音に相当する鍵盤を打鍵するようになってしまっています。打鍵数の上昇についても図 8 下部のグラフを参照いただければと思います。

そして最後に、システム C (Thermosupport) に移行した箇所をみますと、アウト音は劇的に減っており、多くがシステム A の後半部よりも下回っています。そして、システム C の使用に伴ってさらに減少していて、この傾向は全ての被験者にわたっています。

前述の通り、システム C の場合、鍵盤の温度を知ることができるのは「打鍵後」になります。打鍵するまでその音がアベイラブルか否かはわかりません。にもかかわらず減少がみられたというのはどういうことかということ、演奏しようとしている鍵盤がアベイラブルか否かを、被験者が予測できるようになっている、ということになります。これはすなわち、被験者が「学習した」ことを示唆します。

3.4 実験の手法に関する考察

さて、ここからは、この実験および分析手法が、[1]などのシングルケース実験の教科書に照らしてどの程度正しく、どういった改良の余地があるかを考えていきます。

まず、この実験は、個々人の行動の変化を、時系列変化の折れ線グラフとして追っているという点で、シングルケース実験に類していると言えます。また、被験者ごとのレベルが最初から大きく異なるので、平均値による評価を行うほうがむしろリスクがあるように思えます。被験者ごとに経過を追うことによる発見も多く、たとえば被験者 2 をみると、それぞれのシステム移行タイミングを境に大きな変化が起こっていて、システム C の支援効果を裏付けています。一方で、被験者 3 を見てみると、システム C よりもシステム A の最終段のほうで、アウト音打鍵頻度が低くなっており、必ずしも提案システムが良いとは限らないことが議論できます。

この実験の大きな特徴は、時間経過とともに学習が常に進行し、アウト音の打鍵頻度がどんどん下がっていく状況であることです。システム C による効果があるように見えますが、もしかしたら単に時間経過による影響かもしれません。2.2 節で「成熟の脅威」と呼んでいたものがそれになります。「もしシステム A を使い続けていたとしたらどこまで下がったか」を考え、それと現状を比較しないといけません。今回の場合、システム A-B 間の差、B-C 間の差はそれなりにはっきりしているように思えるものの、システム C がシステム A と比べて本当に優れているのかを見極めるのはやや難しいといえます。

統制群というほどでなくても、数人だけシステム A のまま実施していれば、よりはっきりと支援システムの効果を実証できたかもしれません。あるいは、2.3 節で述べた多重ベースラインデザインのやり方を用い、より早く、あるいは遅くシステムを移行させる被験者を用意して比較すればよかった、といえるかもしれません。また、完全に元に戻るとはいえませんが、システムの移行時や最後にもう一度ベースラインに戻す実験を行うべきだったかもしれません。

このように、わずか7人によるシングルケース実験（しかもやや不備あり）ではありますが、それによって提案システムの評価法、および効果に対するめどをつけることができます。前述の方法で追実験・追検証を行えば、他の可能性（脅威）に対する防衛力を高め、より内的妥当性が高い推論が行えるようになると考えられます。

4. 考察

以上のように、既発表の音楽系の支援システムについての実験を例として考えた場合、シングルケース実験ならではの恩恵を享受できており、シングルケース実験の方法論にしっかり照らせば、よりよい実験ができたはずだということもわかります。

同様なかたちで、インタラクション系の研究において、シングルケース実験が向いているといえる領域は多くありそうです。ターゲットユーザに該当する被験者を数多く集めにくいとき、被験者に大きな個人差があるときは多標本実験計画法で実施することが難しいので、シングルケース実験の恩恵があるでしょう。

シングルケース実験による評価が向いていると思われるシステムは、何らかの気づきを伴うもの、支援効果が即時的に望めるものではないかと筆者は考えます。それは短時間ではっきりと効果を観測できる見込みが強いからです。たとえば、問題解決に対してヒントを与えるような支援[6]などとは相性が良いかもしれません。

一方で、提案システムのままさらに訓練に時間をかけないといけない手法の検証は、難しい可能性があります。効果が出るのが遅いほど、それがシステムを切り替えたことによる効果なのかどうかが見極めにくくなるからです。また、従来システムからの覚え直しなど、提案システムを使うことによる弊害があるものも注意を要するかもしれません（それこそじっくり経過を観察するのが向いているとも言えるかもしれませんが）。

5. おわりに

人間はすべからず、日常生活をよりよいものにすべく日々実験を行う存在なのかもしれません。たとえば、起床時間を変えてみたら体調がよくなるかどうかとか、新しいソフトウェアを使ってみて仕事はかどるかとか、栄養ドリンクを飲んでみて効果があるかどうか、けっこう「自分で」試して、その効果を「自分で」検証して、その結果に基づいて自らの生活を変えているといえます。

何かを試してみて、それ以後の自分を、以前の自分と比較し判断を行うという行動は、まさにシングルケース実験に近いものと言えます。そして人間は、自らを被験者にしたシングルケースの実験結果を信頼します。何万人もの統

計データによって有意に効果が出たものでも、自分でやってみて、自分に効果がなければ、きっと継続しないでしょう。もちろん、こうした考え方でいると「げんかつぎ」や「おまじない」といった非科学的な現象にすら因果関係を感じてしまう危険性もあります。一方で、インタラクション研究の分野では、そういったプラセボも含めて応用可能性を切り拓いており、虚偽情報フィードバック[7]などは有用なインタラクション手法として研究・検証・応用がなされています。

実験手法についての書籍を読んでいると、実験手法にも多様な工夫が積み重なっており、それでも十分に完成されているわけではなく、さらに議論や改良が行われていることがわかります。学会では常に新しく多様なインタラクション技術が発明されていますが、同じように、インタラクション分野だからこそ、新しい評価実験手法も発明してほしいと願っています。それによって本コミュニティが加速するのはもちろんですが、たとえば良い評価実験が行えなかったがゆえに日の目をみることのなかった過去のアイデアが、再び注目されることすらあるのではないかと期待しています。

参考文献

- 1) 岩本隆茂, 川俣甲子夫. シングル・ケース研究法—新しい実験計画法とその応用, 勁草書房, 1990.
- 2) 南風原朝和, 下山晴彦, 市川伸一. 心理学研究法入門—調査・実験から実践まで, 東京大学出版会, 2001.
- 3) 高野陽太郎, 岡隆. 心理学研究法 補訂版, 有斐閣, 2017.
- 4) Madsen CH, Becker WC, Thomas DR. Rules, praise, and ignoring: elements of elementary classroom control. *Journal of Applied Behaviour Analysis*, Vol.1, pp.139-150, 1968.
- 5) 宮下芳明, 西本一志. 温度で制約を緩やかに提示するシステム Thermoscore を用いた即興演奏支援. 情報処理学会ヒューマンインタフェース研究会報告, Vol.90, pp.13-18, 2004.
- 6) 鈴木宏昭, 福田玄明. 洞察問題解決の無意識的性質: 連続フラッシュ抑制による閾下プライミングを用いた検討. *認知科学*, Vol.20, pp.353-367, 2013.
- 7) 中村憲史, 片山拓也, 寺田努, 塚本昌彦. 虚偽情報フィードバックを用いた生体情報の制御手法, 情報処理学会論文誌 Vol.54, No.4, pp.1882-7764, 2013.

※本稿は当初「である調」で書いたものの、文体が強くなるためか、どうしても主張のニュアンスが伝わらず、内容を誤解されてしまう可能性もあと危惧されました。そのため、「ですます調」の文体で再執筆し投稿したものです。筆者も「ですます調」での論文執筆は初めてであり、この文体で投稿したことによる影響やリスクは予想し切れていません。今日の学会では「ですます調」の論文は極めて稀であり、この文体が一般的でないことを末尾に付しておきます。