

Reviews of 8433 - *"Evaluating a Personalizable, Inconspicuous Vibrotactile(PIV) Breathing Pacer for In-the-Moment Affect Regulation"*

Reviewer 4 (AC)

Expertise

Knowledgeable

Recommendation

Possibly Accept: I would argue for accepting this paper; 4.0

Review

See the meta-review below

1AC: The Meta-Review

Taking a mixed-design study approach, this paper aims to present an evaluation of the efficacy of a technological intervention (PIV Breathing Pacer) to regulate anxiety in the presence of a cognitive stressor, the user engagement with the technology, and the potential users who might benefit from it according to the individual differences. Based on the results, the paper provides a couple of implications for future design and evaluation of this sort of regulation technologies.

The evaluation criteria include the clarity of the presentation of the contribution to the CHI community, including the related work, methods applied, the results in terms of efficacy, user engagement and individual differences as well as the discussion and implications for design provided.

Overall all three reviewers seem very positive about this paper and found it as an “extremely well designed” (R1) and “mature” (R3) study as well as an “impressive piece of work” (R2) that is “very interesting” (R2) with a “clever” solution (R2). I agree with the reviewers that the topic and contribution of the paper seems very strong and that can be very attractive for the CHI community as highlighted by R2 that this paper “clearly demonstrate the effectiveness of a device inducing breathing”. Although all reviewers are supporting the possible acceptance of this paper, they all raised many questions and concerns that need to be addressed before publication. I believe the authors would be able to reply to the reviewer’s comments as they do have a strong case here, if a strong rebuttal is provided then this paper will be able to move to the next phase.

Here, I will provide a summary of the main issues reported by the reviewers on this manuscript.

1. Improving the presentation of the study:

While R1 appreciate how the authors could fit the paper into 10 pages, R2 and R3 got issues with it due to the “verbosity” (R2) and “somewhat hastily” presentation using “sup-optimal phrasing” (R3). It might be the case, as highlighted by R3, the conference deadline might have affected the quality of the presentation of the case even though both reviewers R2 and R3 described as well written. In particular, R3 asked for a clear introduction of the contribution of the study right up front in the abstract and introduction.

In addition, R3 found that the contribution of the paper was constrained by the 10 pages and together with R1, and R2

have suggested that a journal publication will enable the authors elaborate more on the “theoretical motivation” (R3) and contributions and some of them suggested/mentioned TOCHI as an alternative. What is the author’s uptake of this suggestion? How do you see a conference paper and a journal paper coming from the same study? And this links to the next section.

2. Improving the Presentation of the Case, Methods and Findings

R1 raised some question about the use of multiple sensors during the study and how the data collected from these sensors were used or contribute to the “general analysis”. In addition, R1 asked for clarifications regarding the types of qualitative data collected and if yes, how were these collected. R1 wonders if there was any user feedback before and/or after the intervention?

Similarly, R2 is concerned about the “actual breathing pattern of participants” and urge the authors to elaborate more on whether or not “users synchronized their breathing with the haptic feedback”? and If yes, is it possible to reply to the reviewers why these results are not in the paper?

While R3 asked the authors for “rewriting the exploratory analyses”, R2 suggested to take a step back and include the relevant results needed to make “solid conclusions” rather than wandering around the “participants’ traits and states”.

R2 also suggested to include a table to make it easier for the reader to follow the results as well as improved the readability of figures. And R3 asked for further connection between the findings, discussion and methods. R2 suggested a reference to help the authors to present and discuss the significance of the results.

All reviewers raised many questions around the methods in their reviews so make sure you read them all e.g., different length of the experiments, whether scales are normalized, and the p-values necessary to explain discrepancies, were training and testing test fixed? Any cross-fold validation? (R2), and an “in-depth explanation of the models and critical features” (R3).

3. Discussion and Limitations

While R1 suggested to include more potential applications of the technology beyond the professional therapist in particular in relation to future work, R2 is more critical and request more in-depth discussion in regard to the “effectiveness of PIV—and about reputability of the results – with different stressors, or in different contexts”. R3 raised a very good point and suggested to elaborate a bit more on the limitations of the study to help the reader understand the findings and contribution in context. R1 and R3 in particular would like the paper to consider how the findings of the study can go beyond this particular case to inform the wider community.

4. Others - related work

R1 and R3 has fewer comments about the related work and the way it is presented and provide few suggestions on this.

All reviewers have additional suggestions and questions so please refer to the individual reviews for more details.

This paper also received a short discussion behind the scenes in particular as all three reviewers agrees that this paper would be better presented as a journal paper. Nevertheless, they all find it very relevant and all reviewers saw the potential for this paper to contribute to the CHI community and in particular they also want a reassurance regarding the actual publication of the dataset.

In summary, Reviewers asked for a clear presentation of the contribution upfront in the abstract and introduction, and a more detailed presentation of the methods applied and a more profound description of the analysis and discussion based on the findings but also taking critical perspective making sure that the paper has all the necessary material to make a solid contribution in particular in regards to the “lack of results concerning breathing (and, to a lesser extent, EDA, heart rate, temperature)”[Additional discussion]. For the rebuttal, reviewers will be expecting clear answers to the concerns presented above and if there are further results to be included, how would you achieve this, which in principle

reviewers will not see this as a problem if the data is already available within the study.

We all look forward to reading the rebuttal so make sure you reply to all the points outlined above and/or follow most of the suggestions provided by the reviewers.

Rebuttal response

I would like to thank the authors for submitting the rebuttal and all the additional information provided that has been well received by all the reviewers. However, they got additional comments that I believe the authors will be able to consider/add in the final version of this paper.

- 1) Reviewers appreciate the reply about the physiological data and that they can actually be reported in another paper. However, the physiological data can be used to provide additional context as suggested by R2 that has asked for adding data “concerning respiration, namely to assess up to which point PIV did affect users' breathing”. I believe the authors would be able to provide this as they already got the data.
- 2) The rebuttal states that “anxiety experience were not correlated with the physiological indicators”. I agree with R2 that this needs to be unpacked in the discussion or in the future work section of the paper.
- 3) The details about the methods and the promises for taking additional steps in the analysis were very well received and will enhance the final version of this manuscript. As noted by R2 these can be done in the restricted timeframe of producing the camera-ready version, without diminishing its contribution.
- 4) Although R3 is now satisfied with the rebuttal response, this reviewer would have wished to get part of the supplementary material rather inside the paper to be self-contained.

I strongly suggest the authors to take the last reviewer’s comments on board to improve the quality of the presentation of the contribution of this paper to the CHI community. I do believe that the proposed changes can be doable in the submission timeline and that the authors will publish the dataset.

I wish the authors all the best in working on the revised version of this paper hoping that they will deliver what they have promised.

AT THE PC MEETING

We discussed this paper at the PC meeting extensively. At the end, we argue for accepting this paper believing the authors will make the proposed changes and will account for the last comments in the reviewer's rebuttal response.

Reviewer 3 (2AC)

Expertise

Knowledgeable

Recommendation

Possibly Accept: I would argue for accepting this paper; 4.0

Review

The current manuscript describes a mixed-design experiment that was designed to evaluate a vibrotactile system designed for affect regulation, named PIV. Contributions pertaining to the design of this system was reported in a previous publication. This work solely concerns whether using this system results in reducing anxiety relative to not using it (between-groups; 2 levels), across two stressor manipulations (within-groups; 2 levels). The authors also explore aspects of use difficulty, relative to the different stressors. The authors also employ a machine learning model to account for individual differences might influence the relative benefit from using PIV.

Review

In my opinion, I feel that this submission is better suited as a journal article instead of a conference paper. It is a mature study and there is a lot to like it. In particular, the thoughtful procedures adopted in carrying out the study. A conference paper sells this short because it does not allow the authors to expand on the theoretical motivations for affective regulation and to address why a systematic evaluation such as this is necessary (and currently lacking in the growing field of such applications). Instead, this report currently serves as a service evaluation of an existing implementation, which the reviewers are unable to access given the double blind procedure. Having access to the tailored evaluation without a clear link to the design of implementation itself is a wasted opportunity.

On a less critical note, I would like to point out that this manuscript was written somewhat hastily (perhaps due to the conference deadline) and a reiterative process with peer reviewers would help to iron out some avoidable oversights or sub-optimal phrasing.

To conclude, I like this work. However, I do not think that it should be a conference paper. Instead, the authors should consider revising this manuscript and submitting this to TOCHI instead, as an extension to the implementation of the system. The thoughtful design of systematically evaluating this implementation has implied nuances that are currently lost within the format of a conference paper. In its current form, I doubt that it will have long-lasting impact in spite of the efforts clearly taken by the researchers to perform a robust and replicable evaluation. If this work is accepted, the authors should focus primarily on rewriting the exploratory analyses to fit better with the methods, design implications, and conclusions. They appear to have been written in haste, without a careful integration into the global narrative. Thus, one can only take the authors' interpretations at face-value without having the opportunity to evaluate the analytical process critically.

Given my mixed feelings, I rate this submission 3.5. However, I am likely to revise my opinion given the authors' response.

Title

This title is difficult to process. Instead of being descriptive of the work performed, which is not particularly exciting in my subjective opinion, consider reporting the findings or the research question instead. Being explicit about the research contribution should increase perceived impact.

Introduction

I would prefer if the authors were more explicit in reporting their findings right at the beginning. They currently frame the contribution of this work in abstract terms and also with regards to the work performed. Instead, I would like them to clearly state their findings and to position these findings in light of other work. This is because affect regulation technology is a semi-mature field and it is, thus, suitable for the readers to expect a consolidation of main findings across various attempts (also see, comments on 'Discussion']).

There is an overemphasis on the pre-registration. While I am personally in favor of Open Science, your practice is currently phrased in a way that appears to be irrelevant to the scientific work per se; please note that the original statement released by CHI for pre-registration is no longer actively pursued. Hence, a footnote would suffice unless the authors feel that there is a stronger point to be made, namely how pre-registration is necessary in order to lend strength to their current findings.

The motivation, namely "We have used their recommendations in our research to conduct a thorough evaluation of PIV.", is vague. It will be hard for the reader to determine what constitutes a "thorough evaluation". I advise the authors to paraphrase this in terms of research questions.

Comparison with related work

The authors provide a reasonable selection of prior implementations to compare PIV against. They methodically express the scope and rationale for their selective comparisons. In addition, they provide a systematic delineation between PIV and two targeted implementations (i.e., Doppel and EmotionCheck). Table 1 is an invaluable summary. I further appreciated that the authors reported the effect size of previous work and based their sample size accordingly.

My main recommendation is for the authors to consider illustrating the vibrotactile pattern of PIV as a time-series to contrast this with EmotionCheck and/or Doppel. This would be a more intuitive description than p3 col1 par1.

Method

This section is especially well-presented. Figure 2 provides a concise and clear representation of the experiment design. I am confident of replicating these procedures given the current report.

Results and Discussion

The results are systematically reported and correspond well with the earlier reports of the independent and dependent variables. The authors adopt a consistent research approach and report effect sizes that are consistent with their a priori expectations.

I found the exploratory analyses frustrating to review. For "Stressor and PIV-User Engagement", it was difficult for me to readily understand the differences between the levels of Engagement-type. I reviewed the Methods section again but could not readily establish a link between this variable and the methods. I feel that this section presents some interesting findings but was not able to easily understand what the authors' intended to communicate within the space afforded here. Similar challenges to the reader were experienced when I read the subsection on the predictive model of the efficacy of this intervention given individual differences. Once again, I feel that the authors did interesting work here. However, I'm not able to fully appreciate the work or to evaluate it critically without an in-depth explanation of the models and critical features. The way that it is currently reported, I can only accept the interpretation of the researchers at face-value. It should also be noted that I am not familiar with the classification approach adopted by the researchers and, hence, could require more information or interaction with the reviewers in order to effectively evaluate this work.

Design implications

Without having access to the prior work on PIV, I can only say that the design implications sub-section is well-written and represent reasonable generalizations given the results. The authors also highlight limitations of their interpretations, for example with regards to the relative effectiveness of implicit and explicit involvement technologies. I feel that this aspect ought to be expanded upon and, if so, will have a wider impact than currently.

Conclusions

Effective, yet generic, summary of the potentially good work done.

Rebuttal response

I am satisfied with the authors' response and believe that this work is a valuable contribution to the community. However, I reserve judgment as to whether or not the camera-ready version will indeed be sufficiently revised to address our concerns. This work can be improved with more focus and the manuscript should be self-sufficient, without relegating critical details to supplementary materials or an external repository.

Reviewer 1 (reviewer)

Expertise

Passing Knowledge

Recommendation

. . . Between possibly accept and strong accept; 4.5

Review

In this work, the authors report on an experiment used to determine the (1) efficacy of the PIV breathing pacer in reducing anxiety with using stressors (2) the engagement with PIV and (3) the effect of individual differences on reducing anxiety. The paper is dense with information, and reports on an extremely well designed study which I congratulate the authors on. I appreciate the way the authors presented the information in the paper in 10 pages, although this study is clearly a building block in a bigger project which would be perfect for a journal publication (especially that the supplementary material is already the size of another paper).

I would have liked to see a bit more on the related work apart from a reference to another publication, however, I find the way they presented the comparison nicely done. A question I have is why did the authors use multiple other sensors during the study: EDA/Breathing Gauge/Pulse and temperature? Was the data from these sensors used in any way to confirm the authors' hypothesis or for general analysis of the stressor task? Also did the authors collect any qualitative data that is not in questionnaire form? It would have been nice to know the actual feedback from the users on the effect of the stressor task and their "perceived" feelings of anxiety before and after the stressor and breathing, especially that they were shown their scores during the task. Was there a difference between those who scored well and those who didn't with respect to their responses to the questions?

In the real world, for the PIV to work it would need to be incorporated in a bigger loop of sensing anxiety (implicitly or explicitly by the user) and regulating it. The author's mentioned that the PIV could be used by professional therapists to be prescribed to people suffering from anxiety or other mental health issues, but definitely more situations may exist. Reflecting on them in the grand scope of HCI research would be useful for future work.

Finally, I find that this paper, however dense and more suited as a journal, offers an extremely well designed and presented study, which could be used by others in the emotion regulation research area. I look forward to seeing this paper presented in CHI this year.

Rebuttal response

I have read the authors' rebuttal and thank them for addressing some of the concerns charted out in the reviews. I hope the authors do deliver on the suggested changes for the camera ready as promised.

Reviewer 2 (reviewer)

Expertise

Expert

Recommendation

Possibly Accept: I would argue for accepting this paper; 4.0

Review

In this paper authors study the effectiveness of PIV, a device providing breathing guidance through haptic feedback in order to reduce anxiety. They test the system using a stressor -- a task inducing cognitive stress -- and comparing two groups (about 50 participants each), one with and one without the device (control). Authors employ various questionnaires aimed at measuring anxiety and effect, and they proceed on investigating further how PIV interact with participants' traits and states.

Overall this is an impressive piece of work. The topic is very interesting, the proposed solution clever, the protocol thorough. To my knowledge this is the first paper that so clearly demonstrate the effectiveness of a device inducing

breathing. The paper is well written -- albeit some verbosity, for example in the introduction or in the discussions -- the sample size important (97), most stats appear sound -- actually I rarely come across such thorough analyses, that include the exact formula ran in R. I also salute the openness of authors'. There are still very few occurrences of papers in HCI that register their protocol beforehand, and beyond that author should release their dataset, opening new opportunities for researchers interested in such topic.

Most is well, however despite these appraisals my score is low because I still have few issues with the paper. My major concern is that there is no data concerning the actual breathing patterns of participants. As stated late in the conclusions, the analysis of breathing measurements would unveil key elements for the validation of the system, as it would help to determine up to which point users synchronized their breathing with the haptic feedback. Authors did not write anything about that, and yet as per the experimental protocol they did record physiological signals, and yet as per the TOCHI paper they do have the entire signal processing pipeline ready. Hence I cannot fathom why such results are not part of this paper. This should not be relegated to future work. I would rather remove the entire section about the model instead; better to have solid conclusions about PIV before investigating further interaction with participants' traits and states.

My other issues/remarks:

- Compared to the robustness of the stats I was slightly disappointed to see that a threshold of 0.05 was chosen for significance, a value might be too high to ensure strong conclusions [a]. That, combined with the fact that authors might not have corrected p-values for multiple comparisons could explain some discrepancies authors pointed out, as for example about interaction with Distraction in the model.
- Page 3, authors mention no significant difference between the two groups, without stating which tests were employed.
- Why does the length of the experiment vary to much between participants (60 to 90m)?
- In the meditation stage, was the slow-pace breathing fixed or tuned depending on participants breathing rate baseline? Did author investigating confounding factor regarding the actual breathing rate of participants?
- I am unsure up to which point it is possible to directly compare the different scores related to user engagement and state that (perceived) synchronization could be more difficult than noticing vibrations. Up to which point these various scales are normalized with one another?
- While studying the model, why did authors chose to transform variable to binary categories? I might miss a point here, but regression models could be used to find relationship between continuous values and observed variable, with results that might be easier to interpret and put into practice by other researchers.
- PIV is not that inconspicuous if it necessitate noise cancellation to operate and various elements to be strapped on the body. More iterations are needed for the design as much as for the engineering.
- The readability of various figures should be improved, e.g. Figure 3 can present bigger plots; it is next to impossible to read any values in Figure 6
- The paper might be shorten and easier to follow if the (many) results were presented in a synthetic table
- Authors submitted Q&A and various additional details as supplementary materials. This practice is unusual to me, and while it could be a good and handy one, it feels a bit like cheating, the actual page count going beyond well 10 (+ new references!). For instance important details about model training and testing are omitted in the main paper, in particular about training and testing sets.

- Speaking about the model, were the training and testing test fixed? Without cross-fold validation I fear there might be over-fitting, and it would be hard to test the reliability of the predictions, diminishing the importance of the section.

- Up to which point authors decided on some criteria before studying some of the features selected by their model?

- Could author comment about the effectiveness of PIV -- and about reputability of the results -- with different stressors, or in different contexts?

Despite these various questions I still tend to recommend this paper for acceptance, eagerly waiting for authors rebuttal to see up to which point my score could increase.

[a] Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., ... Johnson, V. E. (2018). Redefine statistical significance. *Nature Human Behaviour*, 2(1), 6–10. <https://doi.org/10.1038/s41562-017-0189-z>

Rebuttal response

I thank authors for their thorough rebuttal.

They addressed my main concern, which is the lack of physiological data. Even though I still think that they could fit in current paper and would just make it stronger, I also understand that those results can benefit from a publication on their own. If authors provide at least some data concerning respiration, namely to assess up to which point PIV did affect users' breathing, that is fine with me. Now I am only teased by the "anxiety experience were not correlated with the physiological indicators" -- authors might share this kind of information as future work in the camera ready.

Authors also gave more details about their analyses, which inclusion in the paper will be welcomed; and they described additional analyses they still have to perform (e.g. more splits, which should be doable within the time frame before camera ready). Even though all results might not "hold" in the final version of the paper, the work and the contribution will not be diminished.

I increased my score after considering these various elements.

[Return to submission and reviews](#)