# Active Sampling for Efficient Subjective Evaluation of Tactons at Scale

Marc Demers[1] Pascal E. Fortin[1] Antoine Weill--Duflos[2] Yongjae Yoo[1] Jeremy R. Cooperstock[1]

***Abstract*— Traditional tacton evaluation studies often rely on pre-defined haptic effects that are specifically tailored to explore a handful of design parameters. To prevent combinatorial explosion, researchers are forced to constrain their exploration to very limited subsets of the parameter space. In this work, we propose a hands-off active sampling strategy grounded in probability and information theory that automatically generates tactons to maximize the perceptual information gain at each stimulus presentation. As a proof of concept of the proposed technique, we present the results from a crowdsourced study investigating the perceived similarity between tactons with over 200 participants. Without researcher intervention in the tacton selection process, our method allowed a set of the most salient features for perception of tacton similarity to emerge naturally from the data. This approach is highly scalable and allows for a more efficient exploration of a larger haptic space than typical laboratory study designs aimed at evaluating perceptual attributes of tactons.**

## I. INTRODUCTION

There is increasing interest in the haptics community for the development of authoring tools that allow novice hapticians to *automatically* generate haptic stimuli that fit their design objectives. However, the majority of such tools were developed from perceptual data acquired in laboratory conditions. While we recognize and value these efforts, we argue that the community still lacks a proper understanding of the inter-individual differences in the perception of vibrotactile effects, and all the more so for in-the-wild settings.

While traditional laboratory studies allow researchers to study the human perceptual system in depth, the fact that their findings are obtained in idealized environments makes their applicability to real-world conditions problematic. Indeed, when faced with realistic attention demands, perceptual noise and the variability introduced by different actuators and devices, the users' perception of vibrotactile effects can be significantly altered [1]. In addition, researchers are often forced to limit their investigation to the evaluation of a few values for each studied parameter to avoid combinatorial explosion and lengthy experiment sessions. This may ultimately result in a sparse understanding of the impact of design parameters on the subjective properties of a tacton. It also encourages researchers to pre-select the studied tactons, based on prior literature or their intuition, according to criteria that seem to best match the properties they are seeking. While understandable, such a bias limits the exploration of the parameter space and risks missing out on a local or global optimum that is outside of the traditionally studied range for a given design parameter.

*This work was supported by a Natural Sciences and Engineering Research Council (NSERC Canada) Discovery Grant RGPIN-2017-05013 and the Fonds de Recherche Nature et Technologies (Québec) B2X No.271767

[1]Dept. of Electrical and Computer Engineering, McGill University. 3480 Rue University, Montréal, H3A 0E9, QC, Canada. {marc.demers,pe.fortin}@mail.mcgill.ca, {yyoo, jer}@cim.mcgill.ca

[2]Haply Robotics Co., Montréal, QC, H2T 1V2, Canada. antoine@haply.co

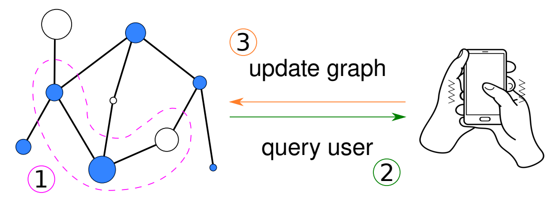



Fig. 1: Schematic representation of the proposed active sampling approach. 1) Select tactons that maximize the expected information gain. 2) Present these tactons through a user's smartphone. 3) Update graph based on the user's perception. All users' ratings update a centralized graph.

In this work, we examine haptic perception from a different angle. Instead of starting from a pre-selected set of tactons, we propose a technique that automatically generates and selects tactons to maximize the information gain for each presented stimulus (Fig. 1). We then iteratively evaluate the perceptual characteristics of the generated tactons *a posteriori* through data mining. Finally, we present and examine an application of this new approach in a crowdsourced study to reveal the most important properties to assessing subjective similarity between two tactons. This study is motivated by the potential benefits of such user-specific knowledge over the status quo of present-day zero-knowledge authoring tools. Indeed, determining how groups of individuals may share certain characteristics of tacton perception that are different from other groups may allow us to overcome the challenge of customization or personalization of tactons to suit a specific user or task [2].

This paper makes the following contributions:

1) 1) it introduces a novel active sampling methodology allowing efficient crowdsourced exploration of a large parameter space without human intervention in stimulus selection; and 2)
2) demonstrates the potential of the proposed approach through the investigation of tacton similarity perception, which leads to identification of three tacton "communities" (clusters), perceived as distinctly dissimilar by a large group of users.

## II. METHODOLOGY

### A. Tacton Psychometrics

As summarized in Table I, the psychometrics of similarity can be expressed by either deterministic or probabilistic models, both in terms of the percept itself, and the mental decision process as to the similarity of two percepts. The decision process is deterministic if the same stimulus always yields the same response, and probabilistic if a response is sampled from a probability distribution.

Perceptual similarity of vibrotactile (VT) patterns has traditionally been investigated with multidimensional scaling (MDS) [3], [4], [5], which considers both the percept and decision process as deterministic. However, most theorists argue that percepts are more probabilistic in nature than deterministic [6]. As such, this work models similarity from a probabilistic point-of-view, i.e., using a Type III joint model.

TABLE I: Classification of psychometric models for similarity. Adapted from Ashby and Ennis [7].

| | | Decision Process | |
|---|---|---|---|
| | | Deterministic | Probabilistic |
| Percept | Deterministic | **Type 0**<br>MDS [8] | **Type II**<br>Logistic |
| Percept | Probabilistic | **Type I**<br>Classical Thurstonian psychophysics [9] | **Type III**<br>Probabilistic extensions of Type II models |

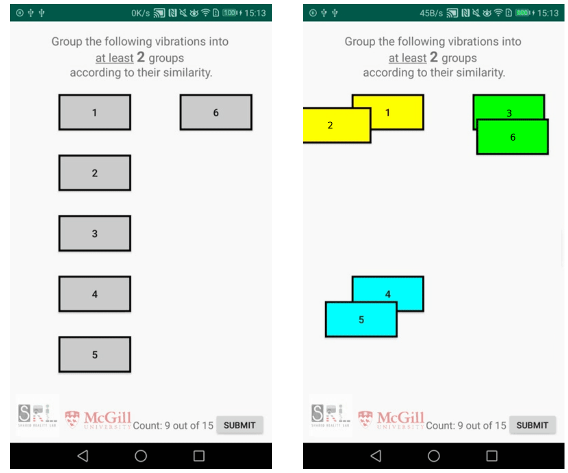
Our study examines the influence of low-level characteristics on our perception of tacton similarity.

### B. Tacton Design

Due to the exploratory nature of the study and to mitigate complexity, we restricted the dimensionality of tactons in several aspects. Each vibrotactile effect was limited to two seconds, as it was found to be their greatest useful length [4]. These two seconds were split in twenty 100 ms segments that were either at maximum or minimum vibration intensity to be compatible with older Android smartphones that do not offer vibration amplitude control. This sampling rate was chosen as it is long enough to resolve the tacton temporal separation of 10 ms [10], yet short enough to carry meaningful content across two seconds of stimulus.

### C. Experimental Procedure

The annotation process takes place directly on the participant's smartphone, via an application downloadable from the Google Play Store. Participants are instructed to hold their smartphone in the palm of their hand while interacting with the application. They were free to use their holding hand or the other hand to touch the screen. Upon query, the tactons are generated server-side and sent to the user for annotation through a RESTful API. Fig. 2 shows the application used by annotators to rate the similarity of a number of tactons. The process can run concurrently for a large number of participants without interruption (batch-crowdsourcing), and data integrity is guaranteed by the use of checksums.




(a) Start of the round. (b) Submitting a grouping.

Fig. 2: Tacton presentation and clustering interface used during the study. Participants can repeatedly experience each tacton as they form clusters they consider to be similar, and then submit their ratings.

We employed a comparison scheme in which six tactons were presented to the participant on each round, who was asked to group them into *at least* two groups.

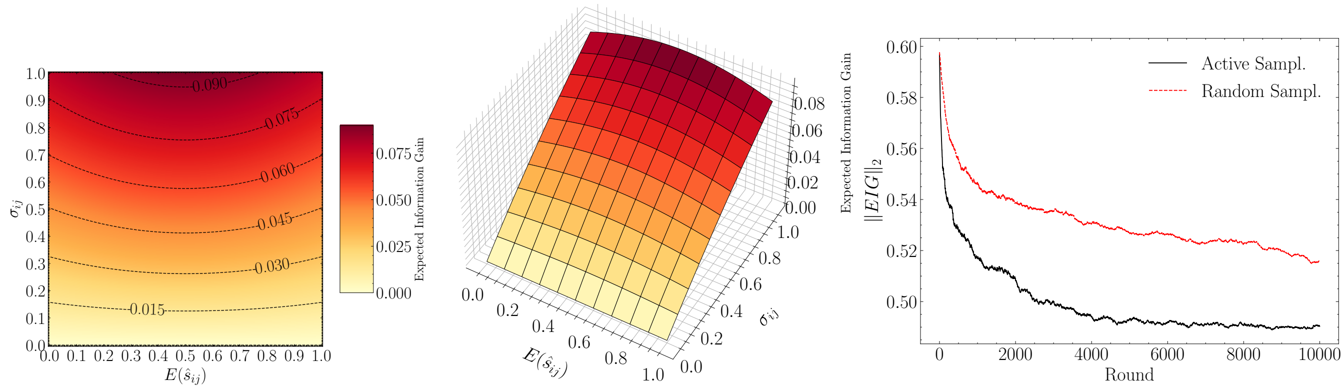
*1) Attention Tests:* With the exception of two identical "attention test" tactons, each round's stimuli were different. The attention test was passed if the identical tactons were assigned to the same cluster. We used this test to quantify compliance with the protocol, evaluate the quality of data provided, and determine participant compensation.

*2) Participants and Compensation:* The experimental procedure builds on HapTurk [11], recruiting a large number of participants over Amazon Mechanical Turk (AMT). We did not impose restrictions on geographical location or worker qualifications (e.g., Master workers). Participants were compensated USD 0.40 for submitting a human intelligence task (HIT), and were awarded a bonus payment of USD 1.60 for obtaining an accuracy score of 80% or greater on the attention tests. Participants performed the experiment once.

In total, 210 workers participated in the study. The data from 129 participants were retained after excluding participants who missed more than 20% of the attention tests. These 129 participants each performed 5 rounds of ratings of 6 tactons for a total of 3870 ratings of non-unique binary tactons. Of these, 252 were unique tactons, as selected by the active sampling procedure.

### D. Similarity Probabilistic Model

We extend the Bradley-Terry model for paired comparisons [12] to the case of an arbitrary number of comparisons to determine similarity on a global scale. This model is probabilistic in nature, as the outcome of each rating is assumed to follow the distribution of a binomial random variable $y$, where the success outcome is considered to be a positive similarity rating.




(a) 2D view of the EIG surface.
(b) 3D view of the same.
(c) Active vs. random sampling performance.

Fig. 3: (a,b) Expected information gain (EIG) as a function of the mean rating ($E(\hat{s}_{ij})$) between two tactons and its associated variance ($\sigma_{ij}$). As expected, pairs that have as many similarity ratings as dissimilarity ratings exhibit the highest EIG. (c) Actively sampling tactons from the tacton space leads to a 6.5× reduction of exploration time compared to random sampling.

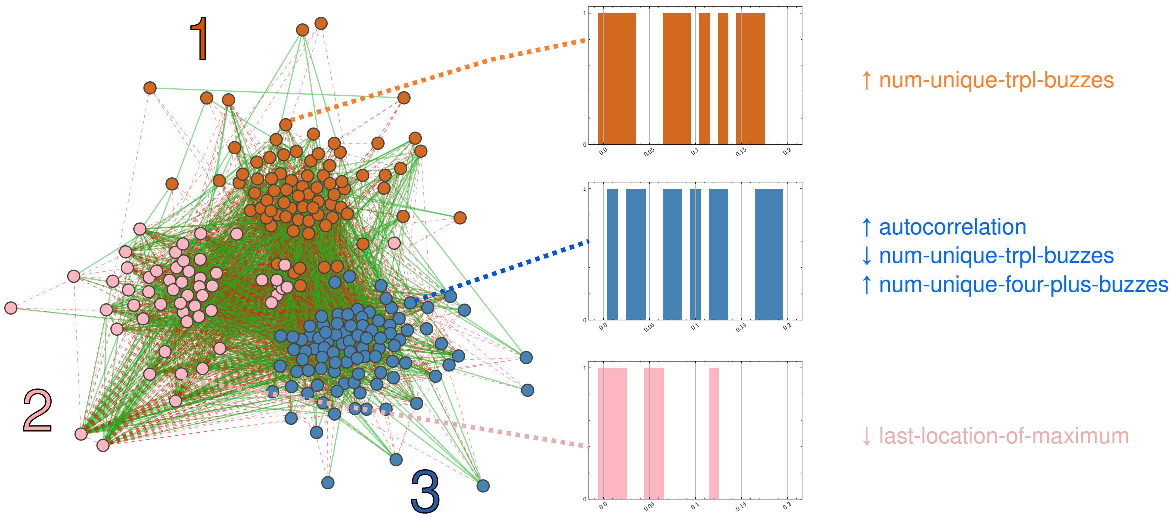



Fig. 4: Community detection using the Leiden algorithm in the similarity network. Edges representing similarity are solid green lines, and those representing dissimilarity edges are dashed red. We show an example tacton and its associated characteristics for each community. On average, there are 9 edges per tacton in the network, giving a dense and rich graph.

To avoid data imbalance, we weighed the outcomes by a factor $w$ such that the sum of the outcomes is equal to the number of groupings performed by the annotators.

We use the similarity pair comparison matrix (PCM) as a positive square matrix A in which each {row, column} ({i, j}) locus on the upper triangle ($i < j$) represents the similarity weight $\alpha_{ij}$ for objects $T_i$ and $T_j$, and every locus on the lower triangle represents their dissimilarity weight $\alpha_{ji}$ ($i > j$).

We model the error on each outcome $y^r_{ij}$ between tactons $T_i$ and $T_j$ by a Gaussian random variable $\varepsilon_{ij}$ [13]. We use the subscripts to denote tacton identifiers and superscript to denote the rating round. The pair similarity value $\alpha_{ij}$ is thus defined as the sum of weighted outcomes up until round $R$:

$$\alpha_{ij} = \sum_{r=1}^{R} \eta^r w^r_{ij} y^r_{ij} + \varepsilon_{ij}, \qquad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2_{ij}) \tag{1}$$

where $\eta^r$ is the *annotator performance rating*, which acts as a surrogate for the reliability of the rating [14]. In practice, we use $\eta^r = 1$ for rounds where the attention test was a success, and $\eta^r = 0.5$ otherwise.

We find the probability that tacton $T_i$ and $T_j$ are rated as similar through logistic binomial regression, such that:

$$P(A_i \sim A_j) = \hat{s}_{ij} = \frac{1}{1 + e^{-(\alpha_{ij} - \alpha_{ji})}} = \frac{e^{\alpha_{ij}}}{e^{\alpha_{ij}} + e^{\alpha_{ji}}}, i < j \tag{2}$$

Assuming the outcomes $y_{ij}$ are independent in probability across time and across annotators, we can define the variance $\sigma^2_{ij}$ of the binomial similarity probability $s_{ij}$ as:

$$\hat{\sigma}_{ij}^2 = \hat{s}_{ij}(1-\hat{s}_{ij}) = \frac{\alpha_{ij}\alpha_{ji}}{(\alpha_{ij}+\alpha_{ji})^2} \tag{3}$$

We then obtain the probability distribution of the similarity pairs $\hat{s}_{ij}$ with their associated variance $\hat{\sigma}_{ij}$ in the $\zeta_{ij}$ matrix:

$$\zeta_{ij} \sim \mathcal{N}(\hat{s}_{ij}, \hat{\sigma}_{ij}) \tag{4}$$

### E. Maximal Information Gain

This probabilistic model allows us to efficiently sample the tacton space for pairs whose (dis-)similarity would yield the highest expected information gain (EIG). We use an approach similar to Li *et al.* [15] to sample pairs of items but adapt it to the case where we do not wish to infer a global ranking of test candidates.

We define the expected Kullback-Leibler divergence (KLD) between the prior probability distribution $P(\zeta_{ij})$ and the posterior distribution given the current outcome $P(\zeta_{ij} \mid y_{ij})$ as a surrogate to the EIG to be the distance function $D_{ij}$:

$$D_{ij} = D_{KL}(P(\zeta_{ij} \mid y_{ij}) \parallel P(\zeta_{ij})), \quad i<j \tag{5}$$

From Eqs. 4 and 5, and simplying, we obtain:

$$\begin{aligned} D_{ij} = \int f_1(y)e^{-y^2}\partial y - \int f_2(y)\, log f_2(y) e^{-y^2}\partial y + \\ \int f_3(y)e^{-y^2}\partial y - \int f_4(y)\, log f_4(y) e^{-y^2}\partial y, \quad i<j \end{aligned} \tag{6}$$

in which the four components are defined as:

$$f_1(y) = \frac{1}{\sqrt{\pi}} \frac{1}{1+e^{-(\sqrt{2}\hat{\sigma}_{ij}y+\Phi(\hat{s}_{ij}))}} \log \frac{1}{1+e^{-(\sqrt{2}\hat{\sigma}_{ij}y+\Phi(\hat{s}_{ij}))}} \tag{7}$$

$$f_2(y) = \frac{1}{\sqrt{\pi}} \frac{1}{1+e^{-(\sqrt{2}\hat{\sigma}_{ij}y+\Phi(\hat{s}_{ij}))}} \tag{8}$$

$$f_3(y) = \frac{1}{\sqrt{\pi}} \frac{1}{1+e^{\sqrt{2}\hat{\sigma}_{ji}y+\Phi(\hat{s}_{ji})}} \log \frac{1}{1+e^{\sqrt{2}\hat{\sigma}_{ji}y+\Phi(\hat{s}_{ji})}} \tag{9}$$

$$f_4(y) = \frac{1}{\sqrt{\pi}} \frac{1}{1+e^{\sqrt{2}\hat{\sigma}_{ji}y+\Phi(\hat{s}_{ji})}} \tag{10}$$

where $\Phi(\cdot)$ is a function that maps the logistic output range $[0,1]$ to a hyperbolic tangent range $[-1,1]$ such that the mean of the normal distribution is not biased towards similarities.

We plot the resulting EIG surface in Figs. 3a and b. The EIG is maximal for pairs with similar scores $s_{ij}$ and $s_{ji}$ and high variance $\sigma_{ij}^2$, and minimal for pairs with different scores and small $\sigma_{ij}^2$. This is consistent with the intuition that we should sample from pairs with maximal *uncertainty*, meaning pairs that do not exhibit clear similarity or dissimilarity.

To actively sample the tacton space, we therefore select the batches of tacton pairs that yield the greatest EIG using the minimum spanning tree (MST) of $-D_{ij}$. We add a new tacton in the global mix when the batches from the MST are exhausted. Advantages of using the MST include a lower computational budget and the possibility of querying *batches* of pairs of tactons. For a more thorough explanation, we refer the reader to Li *et al.* [15]. Each participant receives six tactons from the batch to evaluate in a pairwise fashion, and the process repeats for subsequent rounds (Fig 1).

While a random sampling strategy would have quadratic time complexity ($\mathcal{O}(n^2)$), the active sampling process reduces this to $\mathcal{O}(n \log n)$, which was confirmed empirically by the comparison of performance of these strategies, shown in Fig. 3c. Both curves were calculated on 30 tacton pairs[1] over 10,000 rounds. As a means to show the EIG across all tacton pairs, we depict the Euclidean norm of the EIG: in this case, a lower norm indicates lower uncertainty with respect to the groupings, and shows that the global ratings have lower entropy. We observe significantly faster improvement in the information content of the groupings via the active sampling strategy vs. random sampling, reducing by a factor of approximately 6.5 the number of rounds needed to obtain a comparable EIG norm.

## III. RESULTS AND DISCUSSION

### A. Analysis

The pair comparison matrix obtained from the experiments can be viewed as the adjacency matrix of an acyclic undirected graph, where vertices are tactons, and edge weights represent the similarity probabilities $\hat{s}_{ij}$. Analogous to detecting "cliques" in social networks, we apply the Leiden algorithm [16] to detect the "communities" (clusters) of tactons that are perceived similarly, as seen in Fig. 4, in which each community is represented by a distinct node color and number. The Leiden algorithm not only detects similarity clusters, but also provides a hierarchy of similarity for each cluster, which allows us to learn, for example, that the orange (1) and pink (2) communities are closer to each other than to the blue (3) community.

To describe the binary tactons, we used 15 distinctive features: energy, complexity, non-linearity, autocorrelation, spectral rolloff, binary entropy, first and last location of maximum, first and last location of minimum, number of unique double and triple buzzes, number of four or more unique buzzes, number of ramp-ups, and number of ramp-downs. The first six features were taken from the literature on tacton analysis [4], while the rest were designed specifically for the case of describing binary tactons. The features were measured in the *ad hoc* analysis on the generated tactons to attempt to explain all of the interactions in the network.

In addition, we characterize these emergent communities by analyzing the feature distribution of their constituent tactons. The features distributions were standardized so as to make a fair and clear comparison between the different scales of each feature. We performed a Kruskal-Wallis H-test to determine which features could distinguish all three communities with 95% confidence, and illustrate the top five such features (by p-values) in Fig. 5. The top five p-values obtained from the test were (in order of appearance in Fig. 5)

[1] Note that in actual experimental conditions, the number of tactons is not constant, but increasing as the experiment progresses.

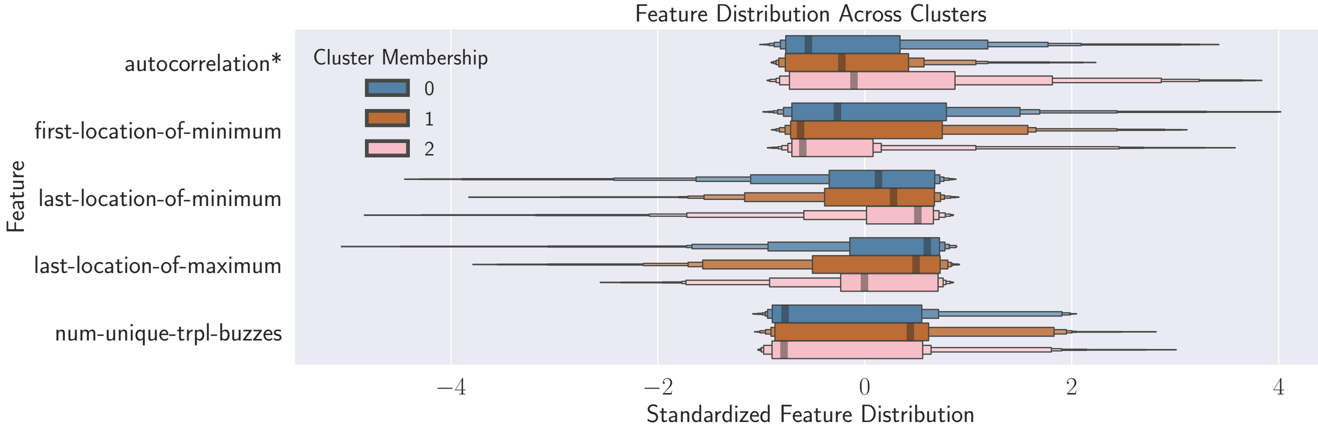



Fig. 5: Feature distribution across all clusters. The distributions were normalized so as to compare them on a similar scale. Only the autocorrelation feature was statistically significant, as determined by a Kruskal-Wallis H-test with 95% confidence. The boxenplots are plotted with the median of the distribution and the slim tails on both sides include outliers.

0.0309, 0.1425, 0.2441, 0.2952, and 0.3368. No post hoc tests were performed. We notice that the orange (1) and blue (3) communities exhibit strong preference for distinct features. Although the pink community exhibits a lower last-location-of-maximum relative to the other communities, this was, on its own, insufficient to discriminate it reliably.

## B. Discussion

These results present the first successful attempt at modeling perceptual similarity in a "bottom-up" fashion, providing a new perspective for subjective evaluation of tactons. While previous experiments for evaluation of haptic similarity typically involved fewer than 100 participants [3], [5], [17], [18], we believe that the results of this "in the wild" study, obtained from a high number of participants, comprising a wide diversity of backgrounds, are more representative of the haptic perception characteristics of the general population. Beyond providing a better understanding of perceptual similarity, the mapping can further be used in applications where a haptician might want users to discern coarse changes between VT patterns (e.g., alerts, notifications), or to propagate sparse subjective ratings in tools to produce content for such applications, such as tacton generation.

With regard to tacton similarity, one of the main findings of the study is that all the tactons could be grouped into three main perceptual similarity "communities" (clusters). From examples of tactons associated with each family, and their most salient features, we anecdotally labelled the pink, orange and blue communities, "short", "coarse" and "jittery" respectively. As seen in Fig. 4, the pink community tends to have short tactons, with a low value of last-location-of-maximum ("short"); the orange cluster has a high number of buzzes of duration longer than a single period ("coarse"); the blue community has high autocorrelation and a high number of buzzes, indicating fast tempo ("jittery"). The importance of these features in the assessment of tacton similarity is not new [4], [5], [18]. This is, however, the first time that these features *emerge* from a bottom-up exploration of tacton similarity perception, as opposed to being part of the initial research hypotheses.

While the number of communities may seem low, it is consistent with the findings of Seifi *et al.* [19] who reported that participants typically preferred higher-level changes to vibrations when asked to design them. This may very well be the case here: the number of perceptual similarity groups that can be distinguished within the entire tacton population at a high-level could be quite low, regardless of the amount of data gathered. It is hypothesized that the low number of communities could also result from the variability introduced by inter-individual perceptual differences, different smartphones and contexts in which the experiment was conducted. We therefore argue that the emergent tacton communities presented in this paper are a better representation of the in-the-wild reality than those obtained under laboratory conditions. A logical follow-up study investigating the possible existence of perceptual "personas" would allow us to obtain a better understanding of such inter-individual differences.

As a first proof of concept and due to the variable availability of vibration amplitude control on smartphones, this study only investigated fixed length binary-amplitude tactons. As such, the features describing the tactons focus on characteristics of the rhythm or the tempo of the VT patterns. Of the 15 features considered, only signal autocorrelation was found to be statistically significantly different between the three emergent tacton communities (see Fig. 5). Supported by the data, our advice to designers seeking highly differentiable binary tactons is to maximize the differences in the autocorrelation of each signal, relative to the others.

## C. Limitations

Despite offering promising methodological possibilities, the proposed active sampling approach suffers from limitations. While it can be applied to any experiment investigating a *subjective comparison* of a number of items, it is not suited to the investigation of sequential vibrotactile stimuli [20]. The scalability of this active sampling approach makes it particularly well suited for evaluating and modeling the subjective experience at a community level. However, given the minimal overlap of evaluated stimuli across participants, it does not support the intra-individual richness offered by traditional within-subjects studies. In addition, despite being advantageous for its perception and decision-making modelling, the probabilistic model used fails to consider the source of the uncertainty generated by the ratings of the annotators. Decomposing the sources of errors that make up the variability of the ratings, e.g., human judgement, human error, or cultural background, would provide insight into our perception of vibrotactile stimuli.

The proposed technique does not aim to be a replacement of traditional lab studies. Due to the uncontrolled nature of the data collection process and its reliance on heterogeneous hardware, it is inadvisable to employ it in the study of low-level fundamental perceptual research questions. For these same reasons however, we argue that this methodology is particularly promising for the large scale investigation of tactons' perceptual properties and their validity in realistic deployment settings.

## IV. Conclusion

In this paper, we introduced a novel active sampling methodology that enables efficient crowdsourced studies of subjective properties of tactons. Unlike traditional laboratory studies, this data-driven method can be deployed without the need to pre-select tactons, which arguably reduces the bias introduced by experimenters. In addition, the proposed approach is theoretically less computationally complex than traditional within-subject randomized tacton presentation, allowing the exploration of larger parameter design spaces in less time. In the presented proof of concept study exploring participants' tacton similarity perception, three communities of similar tactons emerged from the data, consistent with existing literature on the topic based on top-down research approaches. This method is anticipated to facilitate the creation of zero-knowledge haptic authoring tools that can better generalize to the population. The code for the experiments is available from http://bit.ly/3eTSkgm.

## Acknowledgment

We thank Hasti Seifi and Oliver Schneider for their support in the project's early stages, and Jeffrey R. Blum and all SRL lab members for their comments and insight. This study approved by McGill Research Ethics Board, (REB #432-0416). We thank Oleksandr Panasovskyi from the Noun Project for the smartphone icon in Fig. 1.

## References

[1] J. Blum, P. Fortin, F. Taha, P. Alirezaee, M. Demers, A. Weill-Duflos, and J. Cooperstock, "Getting Your Hands Dirty Outside the Lab: A Practical Primer for Conducting Wearable Vibrotactile Haptics Research," *IEEE Transactions on Haptics*, p. 1, July 2019.

[2] H. Seifi, *Personalizing Haptics: From Individuals' Sense-Making Schemas to End-User Haptic Tools*. Springer, June 2019.

[3] J. Pasquero, J. Luk, S. Little, and K. MacLean, "Perceptual Analysis of Haptic Icons: an Investigation into the Validity of Cluster Sorted MDS," in *14th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*. Alexandria, VA, USA: IEEE, 2006, pp. 437–444. [Online]. Available: http://ieeexplore.ieee.org/document/1627122/

[4] D. Ternes and K. E. MacLean, "Designing Large Sets of Haptic Icons with Rhythm," in *Haptics: Perception, Devices and Scenarios*, ser. Lecture Notes in Computer Science, M. Ferre, Ed. Springer Berlin Heidelberg, 2008, pp. 199–208.

[5] I. Hwang and S. Choi, "Perceptual space and adjective rating of sinusoidal vibrations perceived via mobile device," *IEEE Haptics Symposium, HAPTICS 2010*, pp. 1–8, Mar. 2010.

[6] D. M. Ennis and N. L. Johnson, "Thurstone-Shepard Similarity Models as Special Cases of Moment Generating Functions," *Journal of Mathematical Psychology*, vol. 37, no. 1, pp. 104–110, Mar. 1993. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0022249683710059

[7] F. Ashby and D. Ennis, "Similarity measures," *Scholarpedia*, vol. 2, no. 12, p. 4116, 2007. [Online]. Available: http://www.scholarpedia.org/article/Similarity_measures

[8] F. W. Young and R. M. Hamer, *Multidimensional scaling: history, theory, and applications*. Hillsdale, N.J.: L. Erlbaum Associates, 1987, oCLC: 13792004.

[9] L. L. Thurstone, "A law of comparative judgment," *Psychological Review*, vol. 34, no. 4, pp. 273–286, 1927.

[10] M. Pastor, B. Day, E. Macaluso, K. Friston, and R. Frackowiak, "The Functional Neuroanatomy of Temporal Discrimination," *The Journal of neuroscience : the official journal of the Society for Neuroscience*, vol. 24, pp. 2585–2591, Apr. 2004.

[11] O. S. Schneider, H. Seifi, S. Kashani, M. Chun, and K. E. MacLean, "HapTurk: Crowdsourcing Affective Ratings of Vibrotactile Icons," in *Proc. Human Factors in Computing Systems*. New York, NY, USA: ACM, 2016, pp. 3248–3260, event-place: San Jose, California, USA. [Online]. Available: http://doi.acm.org/10.1145/2858036.2858279

[12] R. A. Bradley and M. E. Terry, *The rank analysis of incomplete block designs 1: the method of paired comparisons*. Blacksburg, VA: Virginia Agricultural Experiment Station, 1952, oCLC: 48031947.

[13] J. P. Dotson, J. R. Howell, J. D. Brazell, T. Otter, P. J. Lenk, S. MacEachern, and G. M. Allenby, "A Probit Model with Structured Covariance for Similarity Effects and Source of Volume Calculations," *Journal of Marketing Research*, vol. 55, no. 1, pp. 35–47, Feb. 2018. [Online]. Available: https://doi.org/10.1509/jmr.13.0240

[14] X. Chen, P. N. Bennett, K. Collins-Thompson, and E. Horvitz, "Pairwise ranking aggregation in a crowdsourced setting," in *Proc. Web search and data mining - WSDM '13*. Rome, Italy: ACM, 2013, p. 193. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2433396.2433420

[15] J. Li, R. K. Mantiuk, J. Wang, S. Ling, and P. L. Callet, "Hybrid-MST: A Hybrid Active Sampling Strategy for Pairwise Preference Aggregation," *arXiv:1810.08851 [cs, stat]*, Oct. 2018, arXiv: 1810.08851. [Online]. Available: http://arxiv.org/abs/1810.08851

[16] V. A. Traag, L. Waltman, and N. J. van Eck, "From Louvain to Leiden: guaranteeing well-connected communities," *Scientific Reports*, vol. 9, no. 1, p. 5233, Mar. 2019. [Online]. Available: https://www.nature.com/articles/s41598-019-41695-z

[17] N. Gaißert, H. H. Bülthoff, and C. Wallraven, "Similarity and categorization: From vision to touch," *Acta Psychologica*, vol. 138, no. 1, pp. 219–230, Sept. 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0001691811001302

[18] I. Hwang, J. Seo, and S. Choi, "Perceptual Space of Superimposed Dual-Frequency Vibrations in the Hands," *PLOS ONE*, vol. 12, no. 1, p. e0169570, Jan. 2017. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0169570

[19] H. Seifi, C. Anthonypillai, and K. E. MacLean, "End-user customization of affective tactile messages: A qualitative examination of tool parameters," in *Haptics Symposium*. IEEE, Feb. 2014, pp. 251–256.

[20] J. R. Cauchard, J. L. Cheng, T. Pietrzak, and J. A. Landay, "ActiVibe: Design and Evaluation of Vibrations for Progress Monitoring," in *Proc. Human Factors in Computing Systems*. San Jose, California, USA: ACM, May 2016, pp. 3261–3271. [Online]. Available: https://doi.org/10.1145/2858036.2858046