HapTurk: Crowdsourcing Affective Ratings for Vibrotactile Icons

Oliver S. Schneider, Hasti Seifi, Salma Kashani, Matthew Chun, and Karon E. MacLean
oschneid@cs.ubc.ca, seifi@cs.ubc.ca, salma@ece.ubc.ca, mchun345@cs.ubc.ca, maclean@cs.ubc.ca
Sensory, Perception, & Interaction (SPIN) Lab, University of British Columbia, Vancouver, Canada

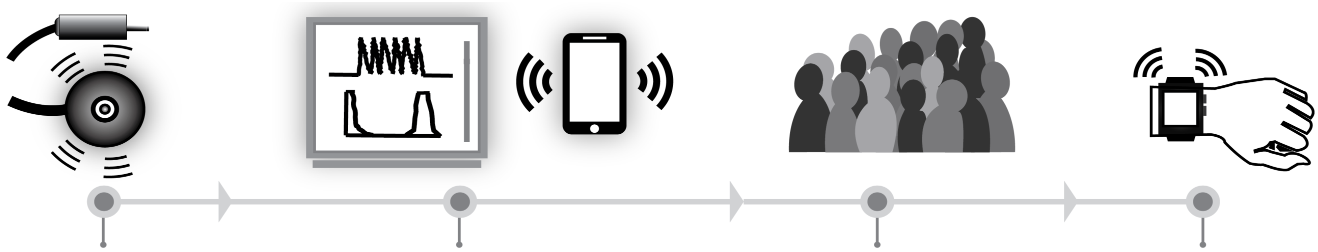


Figure 1: In HapTurk, we access large-scale feedback on informational effectiveness of high-fidelity vibrations after translating them into proxies of various modalities, rendering important characteristics in a crowdsource-friendly way.

ABSTRACT
Vibrotactile (VT) display is becoming a standard component of informative user experience, where notifications and feedback must convey information eyes-free. However, effective design is hindered by incomplete understanding of relevant perceptual qualities. To access evaluation streamlining now common in visual design, we introduce *proxy modalities* as a way to crowdsource VT sensations by reliably communicating high-level features through a crowd-accessible channel. We investigate two proxy modalities to represent a high-fidelity tactor: a new VT visualization, and low-fidelity vibratory translations playable on commodity smartphones. We translated 10 high-fidelity vibrations into both modalities, and in two user studies found that both proxy modalities can communicate affective features, and are consistent when deployed remotely over Mechanical Turk. We analyze fit of features to modalities, and suggest future improvements.

Author Keywords
Haptic; vibrotactile; design; crowdsource; Mechanical Turk; visualization; user study; multimodal.

ACM Classification Keywords
H.5.2. Information Interfaces & Presentation (e.g. HCI): User Interfaces - Haptic I/O

CHI'16, May 07 - 12, 2016, San Jose, CA, USA

DOI: http://dx.doi.org/10.1145/2858036.2858279

INTRODUCTION
In modern handheld and wearable devices, vibrotactile (VT) feedback can provide unintrusive, potentially meaningful cues through wearables in on-the-go contexts [8]. With consumer wearables like Pebble and the Apple Watch featuring high-fidelity actuators, VT feedback is becoming standard in more user tools. Today, VT designers seek to provide sensations with various perceptual and emotional connotations to support the growing use cases for VT feedback (everyday apps, games, etc.). Although low-level design guidelines exist and are helpful for addressing perceptual requirements [5, 6, 22, 32, 47], higher-level concerns and design approaches to increase their usability and information capacity (e.g., a user's desired affective response, or affective or metaphorical interpretation) have only recently received study and are far from solved [4, 25, 26, 35, 36, 42]. Tactile design thus relies heavily on iteration and user feedback [40]. Despite its importance [42, 43], collecting user feedback on perceptual and emotional (i.e., affective) properties of tactile sensations in small-scale lab studies is undermined by noise due to individual differences (IDs).

In other design domains, crowdsourcing enables collecting feedback at scale. Researchers and designers use platforms like Amazon's Mechanical Turk (`www.mturk.com`) to deploy user studies with large samples, receiving extremely rapid feedback in, e.g., creative text production [45], graphic design [49] and sonic imitations [9].

The problem with crowdsourcing tactile feedback is that the "crowd" can't feel the stimuli. Even when consumer devices have tactors, output quality and intensity is unpredictable and uncontrollable. Sending each user a device is impractical.

What we need are crowd-friendly proxies for test stimuli. Here, we define a *proxy vibration* as a sensation that communicates key characteristics of a source stimulus within a bounded error; a *proxy modality* is the perceptual channel and representation employed. In the new evaluation process thus enabled, the designer translates a sensation of interest into a proxy modality, receives rapid feedback from a crowdsourcing platform, then interprets that feedback using known error bounds. In this way, designers can receive high-volume, rapid feedback to use in tandem with costly in-lab studies, for example, to guide initial designs or to generalize findings from smaller studies with a larger sample.

To this end, we must first establish feasibility of this approach, with specific goals: **(G1)** Do proxy modalities work? Can they effectively communicate both physical VT properties (e.g., duration), and high-level affective properties (roughness, pleasantness)? **(G2)** Can proxies be deployed remotely? **(G3)** What modalities work, and **(G4)** what obstacles must be overcome to make this approach practical?

This paper describes a proof-of-concept for proxy modalities for tactile crowdsourcing, and identifies challenges throughout the workflow pipeline. We describe and assess two modalities' development, translation process, validation with a test set translation, and MTurk deployment. Our two modalities are a new technique to graphically visualize high-level traits, and the low-fidelity actuators on users' own commodity smartphones. Our test material is a set of 10 VT stimuli designed for a high-fidelity tactile display suitable for wearables (referred to as "high fidelity vibrations"), and perceptually well understood as presented by that type of display (Figure 7). We conducted two coupled studies, first validating proxy expressiveness in lab, then establishing correspondence of results in remote deployment. Our contributions are:

- A way to crowdsource tactile sensations (vibration proxies), with a technical proof-of-concept.
- A visualization method that communicates high-level affective features more effectively than the current tactile visualization standard (vibration waveforms).
- Evidence that both proxy modalities can represent high-level affective features, with lessons about which features work best with which modalities.
- Evidence that our proxy modalities are consistently rated in-lab and remotely, with initial lessons for compliance.

RELATED WORK

We cover work related to VT icons and evaluation methods for VT effects, the current understanding of affective haptics, and work with Mechanical Turk in other modalities.

Existing Evaluation Methods for VT Effects

The haptic community has appropriated or developed many types of user studies to evaluate VT effects and support VT design. These target a variety of objectives:

1) *Perceptibility:* Determine the perceptual threshold or Just Noticeable Difference (JND) of VT parameters. Researchers vary the values of a VT parameter (e.g., frequency) to determine the minimum perceptible change [31,38].

2) *Illusions:* Studies investigate effects like masking or apparent motion of VT sensations, useful to expand a haptic designer's palette [17,23,44].

3) *Perceptual organization:* Reveal the underlying dimensionality of how humans perceive VT effects (which are generally different than the machine parameters used to generate the stimuli). Multidimensional Scaling (MDS) studies are common, inviting participants compare or group vibrations based on perceived similarity [10,20,37,47,48].

4) *Encoding abstract information:* Researchers examine salient and memorable VT parameters (e.g. energy, rhythm) as well as the number of VT icons that people can remember and attribute to an information piece [2,6,10,47].

5) *Assign affect:* Studies investigate the link between affective characteristics of vibrations (e.g., pleasantness, urgency) to their engineering parameters (e.g., frequency, waveform) [30,39,47,50]. To achieve this, VT researchers commonly design or collect a set of vibrations and ask participants to rate them on a set of qualitative metrics.

6) *Identify language:* Participants describe or annotate tactile stimuli in natural language [10,16,21,35,43,47].

7) *Use case support:* Case studies focus on conveying information with VT icons such as collaboration [10], public transit [8] and direction [4,8], or timing of a presentation [46]. In other cases, VT effects are designed for user engagement, for example in games and movies, multimodal storytelling, or art installations [25,51]. Here, the designers use iterative design and user feedback (qualitative and quantitative with user rating) to refine and ensure effective design.

All of the above studies would benefit from the large number of participants and fast data collection on MTurk. In this paper, we chose our methodology so that the results are informative for a broad range of these studies.

Affective Haptics

VT designers have the challenge of creating perceptually salient icon sets that convey meaningful content. A full range of expressiveness means manipulating not only a vibration's physical characteristics but also its perceptual and emotional properties, and collecting feedback on this. Here, we refer to all these properties as affective characteristics.

Some foundations for affective VT design are in place. Studies on tactile language and affect are establishing a set of perceptual metrics [35,43]. Guest *et al.* collated a large list of emotion and sensation words describing tactile stimuli; then, based on multidimensional scaling of similarity ratings, proposed comfort or pleasantness and arousal as key dimensions for tactile emotion words, and rough/smooth, cold/warm, and wet/dry for sensation [35]. Even so, there is not yet agreement on an affective tactile design language [26].

Recently, Seifi *et al.* compiled research on tactile language into five taxonomies for describing vibrations [43]. **1) Phys-**

ical properties that can be measured: e.g., duration, energy, tempo or speed, rhythm structure; **2) sensory properties**: roughness, and sensory words from Guest *et al.*'s touch dictionary [16]; **3) emotional interpretations**: pleasantness, arousal (urgency), dictionary emotion words [16]; **4) metaphors** provide familiar examples resembling the vibration's feel: heartbeat, insects; **5) usage examples** describe events which a vibration fits: an incoming message or alarm.

To evaluate our vibration proxies, we derived six metrics from these taxonomies to capture vibrations' physical, sensory and emotional aspects: 1) duration, 2) energy, 3) speed, 4) roughness, 5) pleasantness, and 6) urgency.

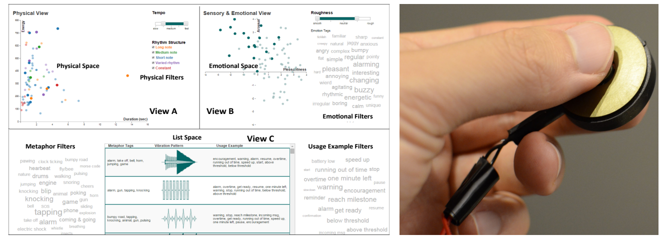
Mechanical Turk (MTurk)

MTurk is a platform for receiving feedback from a large number of users, in a short time at a low cost [18, 29]. These large, fast, cheap samples have proved useful for many cases including running perceptual studies [18], developing taxonomies [11], feedback on text [45], graphic design [49], and sonic imitations [9].

Crowdsourced studies have drawbacks. The remote, asynchronous study environment is not controlled; compared to a quiet lab, participants may be subjected to unknown interruptions, and may spend less time on task with more response variability [29]. MTurk is not suitable for getting rich, qualitative feedback or following up on performance or strategy [33]. Best practices – e.g., simplifying tasks to be confined to a singular activity, or using instructions complemented with example responses – are used to reduce task ambiguity and improve response quality [3]. Some participants try to exploit the service for personal profit, exhibiting low task engagement [12], and must be pre- or post-screened.

Studies have examined MTurk result validity in other domains. Most relevantly, Heer *et al.* [18] validated MTurk data for graphical perception experiments (spatial encoding and luminance contrast) by replicating previous perceptual studies on MTurk. Similarly, we compare results of our local user study with an MTurk study to assess viability of running VT studies on MTurk, and collect and examine phone properties in our MTurk deployment.

Need for HapTurk: Our present goal is to give the haptic design community access to crowdsourced evaluation so we can establish modality-specific methodological tradeoffs. There is ample need for huge-sample haptic evaluation. User experience of transmitted sensations must be robust to receiving device diversity. Techniques to broadcast haptic effects to video [28, 34], e.g., with YouTube [1] or MPEG7 [13, 14] now require known high-fidelity devices because of remote device uncertainty; the same applies to social protocols developed for remote use of high-quality vibrations, e.g. in collaborative turn taking [10]. Elsewhere, studies of VT use in consumer devices need larger samples: e.g., perceivability [27], encoding of caller parameters [7], including caller emotion and physical presence collected from pressure on another handset [19], and usability of expressive, customizable VT icons in social messaging [24]. To our knowledge, this is the first attempt to run a haptic study on a crowdsource site and characterize its feasibility and challenges for haptics.



(a) VibViz interface [43]

(b) C2 tactor

Figure 2: Source of high-fidelity vibrations and perceptual rating scales.

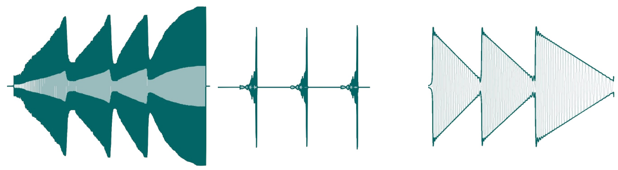


Figure 3: Vis_{DIR} Visualization, based on VibViz

SOURCING REFERENCE VIBRATIONS AND QUALITIES

We required a set of exemplar source vibrations on which to base our proxy modalities. This set needed to 1) vary in physical, perceptual, and emotional characteristics, 2) represent the variation in a larger source library, and 3) be small enough for experimental feasibility.

High-fidelity reference library

We chose 10 vibrations from a large, freely available library of 120 vibrations (VibViz, [43]), browsable through five descriptive taxonomies, and ratings of taxonomic properties. Vibrations were designed for an Engineering Acoustics C2 tactor, a high-fidelity, wearable-suitable voice coil, commonly used in haptic research [43]. We employed VibViz's filtering tools to sample, ensuring variety and coverage by selecting vibrations at high and low ends of energy / duration dimensions, and filtering by ratings of temporal structure/rhythm, roughness, pleasantness, and urgency. To reduce bias, two researchers independently and iteratively selected a set of 10 items each, which were then merged.

Because VibViz was designed for a C2 tactor, we used a handheld C2 in the present study (Figure 2b).

Affective properties and rating scales

To evaluate our proxies, we adapted six rating scales from the tactile literature and new studies. Seifi *et al.* [43] proposed five taxonomies for describing vibrations including physical, sensory, emotional, metaphors, and use examples. Three taxonomies comprise quantitative metrics and adjectives; two use descriptive words.

We chose six quantitative metrics from [43] that capture important affective (physical, perceptual, and emotional)

VT qualities: 1) *duration* [low-high], 2) *energy* [low-high], 3) *speed* [slow-fast], 4) *roughness* [smooth-rough], 5) *urgency* [relaxed-alarming], and 6) *pleasantness* [unpleasant-pleasant]. A large scale (0-100) allowed us to treat the ratings as continuous variables. To keep trials quick and MTurk-suitable, we did not request open-ended responses or tagging.

PROXY CHOICE AND DESIGN

The proxies' purpose was to capture high-level traits of source signals. We investigated two proxy channels and approaches, to efficiently establish viability and search for triangulated perspectives on what will work. The most obvious starting points are to 1) visually augment the current standard of a direct trace of $amplitude = f(time)$, and 2) reconstruct vibrations for common-denominator, low-fidelity actuators.

We considered other possibilities (e.g., auditory stimuli, for which MTurk has been used [9], or animations). However, our selected modalities balance a) directness of translation (low fidelity could not be excluded); b) signal control (hard to ensure consistent audio quality/volume/ambient masking); and c) development progression (visualization underlies animation, and is simpler to design, implement, display). We avoided multisensory combinations at this early stage for clarity of results. Once the key modalities are tested, combinations can be investigated in future work.

"REF" denotes high-fidelity source renderings (C2 tactor).

1) Visual proxies: Norms in published works (e.g. [10]) directed [43] to confirm that users rely on graphical $f(time)$ plots to skim and choose from large libraries. We tested the direct plot, VIS_{DIR}, as the status-quo representation.

However, these unmodified time-series emphasize or mask traits differently than felt vibrations, in particular for higher-level or "meta" responses. We considered many other means of visualizing vibration characteristics, pruned candidates and refined design via piloting to produce a new scheme which explicitly *emphasizes* affective features, VIS_{EMPH}.

2) Low-fidelity vibration proxy: Commodity device (e.g. smartphone) actuators usually have low output capability compared to the C2, in terms of frequency response, loudness range, distortion and parameter independence. Encouraged by expressive rendering of VT sensations with commodity actuation (from early constraints [10] to deliberate design-for-lofi [24]), we altered stimuli to convey high-level parameters under these conditions, hereafter referred to as LOFIVIB.

Translation: Below, we detail first-pass proxy development. In this feasibility stage, we translated proxy vibrations manually and iteratively, as we sought generalizable mappings of the parametric vibration definition to the perceptual quality we wished to highlight in the proxy. We frequently relied on a cycle of user feedback, e.g., to establish the perceived roughness of the original stimuli and proxy candidate.

Automatic translation is an exciting goal. Without it, HapTurk is still useful for gathering large samples; but automation will enable a very rapid create-test cycle. It should be attainable, bootstrapped by the up-scaling of crowdsourcing itself. With a basic process in place, we can use MTurk studies to identify these mappings relatively quickly.

Visualization Design (VIS_{DIR} and VIS_{EMPH})

VIS_{DIR} was based on the original waveform visualization used in VibViz (Figure 3). In Matlab, vibration frequency and envelope were encoded to highlight its pattern over time. Since VIS_{DIR} patterns were detailed, technical and often inscrutable for users without an engineering background, we also developed a more interpretive visual representation, VIS_{EMPH}; and included VIS_{DIR} as a status-quo baseline.

We took many approaches to depicting vibration high-level properties, with visual elements such as line thickness, shape, texture and colour (Figure 4). We first focused on line sharpness, colour intensity, length and texture: graphical waveform smoothness and roughness were mapped to perceived roughness; colour intensity highlighted perceived energy. Duration mapped to length of the graphic, while colour and texture encoded the original's invoked emotion.

Four participants were informally interviewed and asked to feel REF vibrations, describe their reactions, and compare them to several visualization candidates. Participants differed in their responses, and had difficulties in understanding VT emotional characteristics from the graphic (i.e. pleasantness, urgency), and in reading the circular patterns. We simplified the designs, eliminating representation of emotional characteristics (color, texture), while retaining more objective mappings for physical and sensory characteristics.

VIS_{EMPH} won an informal evaluation of final proxy candidates (n=7), and was captured in a translation guideline (Figure 5).

Low Fidelity Vibration Design

For our second proxy modality, we translated REF vibrations into LOFIVIB vibrations. We used a smartphone platform for their built-in commodity-level VT displays, their ubiquity amongst users, and low security concerns for vibration imports to personal devices [15]. To distribute vibrations remotely, we used HTML5 Vibration API, implemented on Android phones running compatible web browsers (Google Chrome or Mozilla Firefox).

As with VIS_{EMPH}, we focused on physical properties when developing LOFIVIB (our single low-fi proxy exemplar). We emphasized rhythm structure, an important design parameter [47] and the only direct control parameter of the HTML5 API, which issues vibrations using a series of on/off durations. Simultaneously, we manipulated perceived energy level by adjusting the actuator pulse train on/off ratio, up to the point where the rhythm presentation was compromised. Shorter durations represented a weak-feeling hi-fi signal, while longer durations conveyed intensity in the original. This was most challenging for dynamic intensities or frequencies, such as increasing or decreasing ramps, and long, low-intensity sensations. Here we used a duty-cycle inspired technique, similar to [24], illustrated in Figure 6.

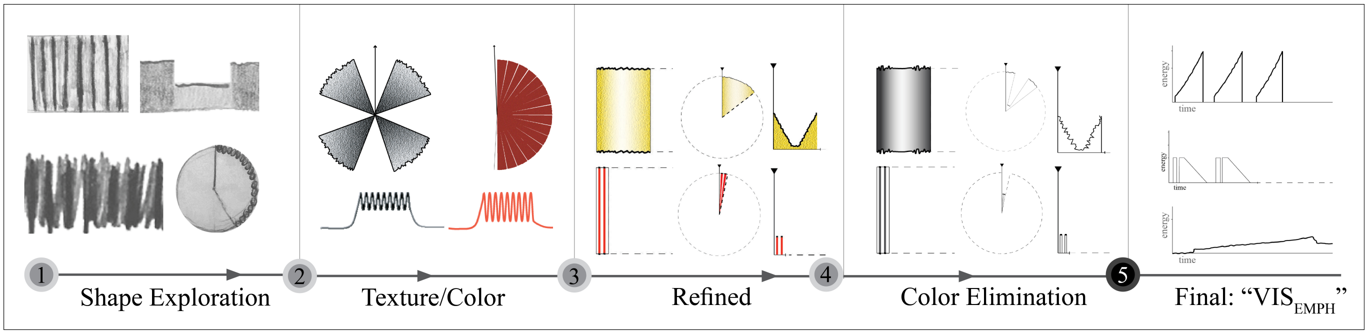


Figure 4: Visualization design process. Iterative development and piloting results in the VIS$_{EMPH}$ visualization pattern.

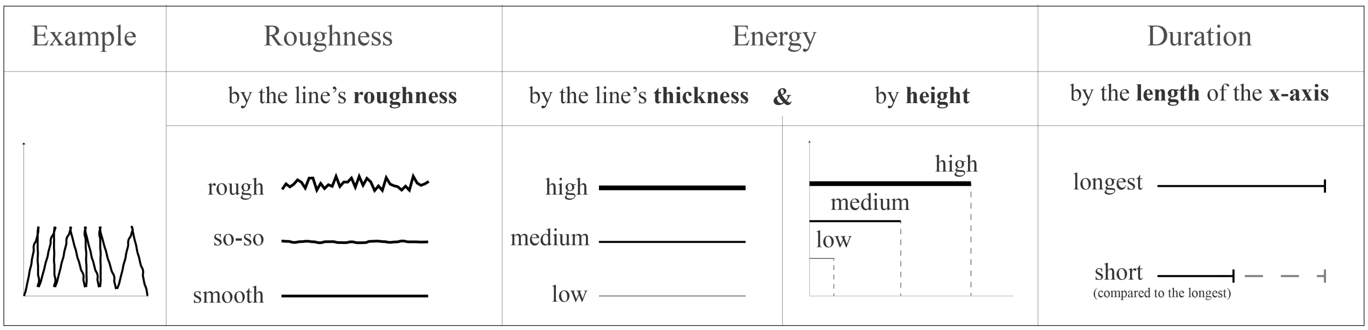


Figure 5: Final VIS$_{EMPH}$ visualization guide, used by researchers to create VIS$_{EMPH}$ proxy vibrations and provided to participants during VIS$_{EMPH}$ study conditions.

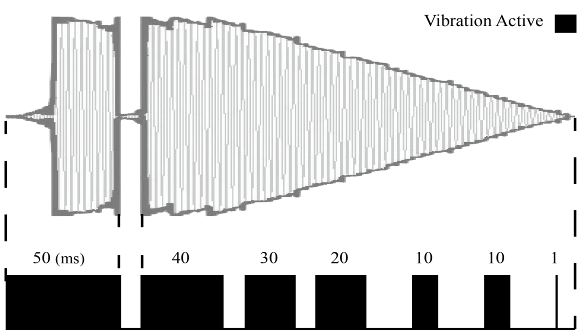


Figure 6: Example of LOFIVIB proxy design. Pulse duration was hand-tuned to represent length and intensity, using duty cycle to express dynamics such as ramps and oscillations.

To mitigate the effect of different actuators found in smartphones, we limited our investigation to Android OS. While this restricted our participant pool, there was nevertheless no difficulty in quickly collecting data for either study. We designed for two phones representing the largest classes of smartphone actuators: Samsung Galaxy Nexus, which contains a coin-style actuator, and a Sony Xperia Z3 Compact, which uses a pager motor resulting in more subdued, smooth sensations. Though perceptually different, control of both actuator styles are limited to on/off durations. As with VIS$_{EMPH}$, we developed LOFIVIB vibrations iteratively, first with team feedback, then informal interviews (n=6).

STUDY 1: IN-LAB PROXY VIBRATION VALIDATION (G1)

We obtained user ratings for the hi-fi source vibrations REFand three proxies (VIS$_{DIR}$, VIS$_{EMPH}$, and LOFIVIB). An in-lab format avoided confounds and unknowns due to remote MTurk deployment, addressed in Study 2. Study 1 had two versions: in one, participants rated visual proxies VIS$_{DIR}$ and VIS$_{EMPH}$ next to REF; and in the other, LOFIVIB next to REF. REF$_{VIS}$ and REF$_{LOFIVIB}$ denote these two references, each compared with its respective proxy(ies) and thus with its own data. In each substudy, participants rated each REF vibration on 6 scales [0-100] in a computer survey, and again for the proxies. Participants in the visual substudy did this for both VIS$_{DIR}$ and VIS$_{EMPH}$, then indicated preference for one. Participants in the lo-fi study completed the LOFIVIB survey on a phone, which also played vibrations using Javascript and HTML5; other survey elements employed a laptop. 40 participants aged 18-50 were recruited via university undergraduate mailing lists. 20 (8F) participated in the visual substudy, and a different 20 (10F) in the low-fi vibration substudy.

Reference and proxies were presented in different random orders. Pilots confirmed that participants did not notice proxy/target linkages, and thus were unlikely to consciously match their ratings between pair elements. REF/proxy presentation order was counterbalanced, as was VIS$_{DIR}$/VIS$_{EMPH}$.

Comparison Metric: Equivalence Threshold

To assess whether a proxy modalities were rated similarly to their targets, we employed *equivalence testing*, which tests the hypothesis that sample means are within a threshold δ,

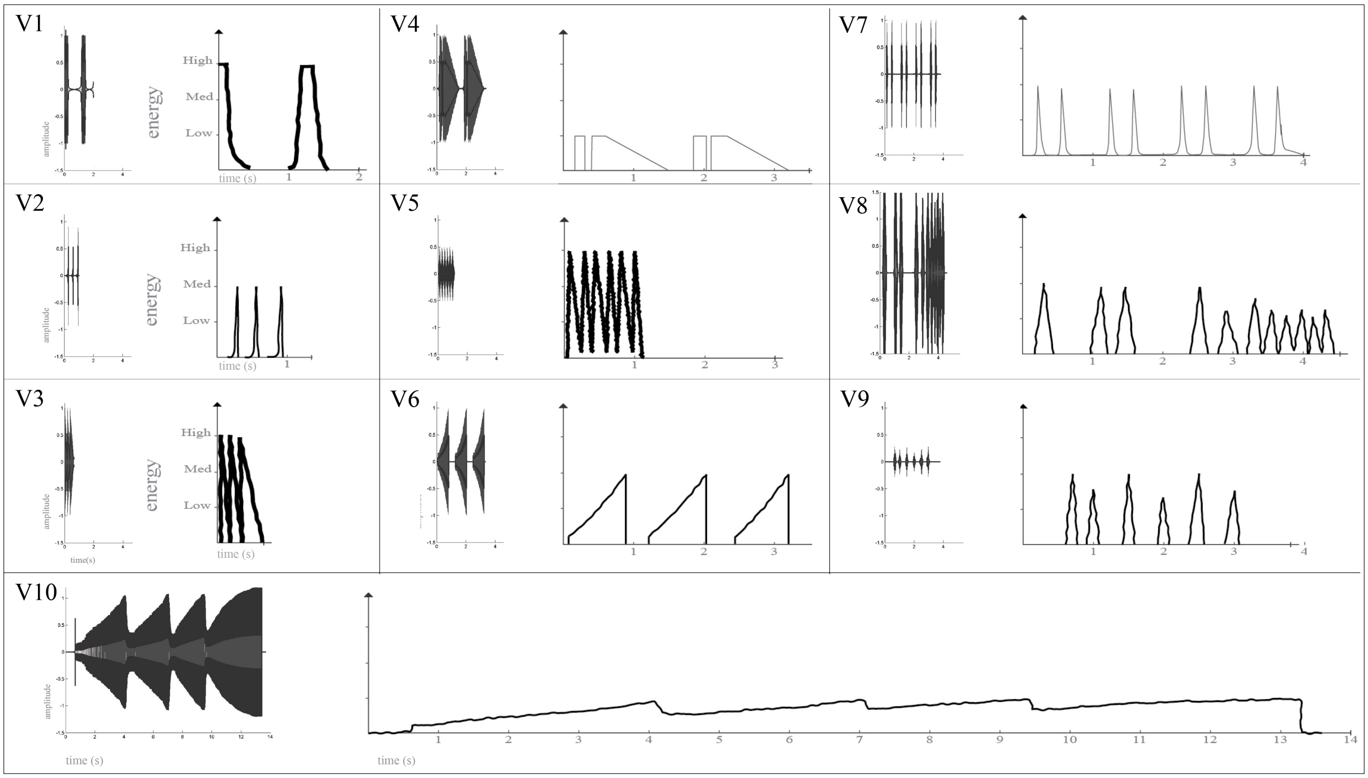


Figure 7: Vibrations visualized as both VIS_{DIR} (left of each pair) and VIS_{EMPH}.

against the null of being outside it [41]. This tests if two samples are equivalent with a known error bound; it corresponds to creating confidence intervals of means, and examining whether they lie entirely within the range $(-\delta, \delta)$.

We first computed least-squares means for the 6 rating scales for each proxy modality and vibration. 95% confidence intervals (CI) for REF rating means ranged from 14.23 points (Duration ratings) to 20.33 (Speed). Because estimates of the REF "gold standard" mean could not be more precise than these bounds, we set equivalence thresholds for each rating equal to CI width. For example, given the CI for Duration of 14.23, we considered proxy Duration ratings equivalent if the CI for a difference fell completely in the range $(-14.23, 14.23)$. With pooled standard error, this corresponded to the case where two CIs overlap by more than 50%. We also report when a *difference* was detected, through typical hypothesis testing (i.e., where CIs do not overlap).

Thus, each rating set pair could be *equivalent*, *uncertain*, or *different*. Figure 9 offers insight into how these levels are reflected in the data given the high rating variance. This approach gives a useful error bound, quantifying the precision tradeoff in using vibration proxies to crowdsource feedback.

Proxy Validation (Study 1) Results and Discussion

Overview of Results

Study 1 results appear graphically in Figure 8. To interpret this plot, look for (1) equivalence indicated by bar color, and CI size by bar height (dark green/small are good); (2) rating richness: how much spread, vibration to vibration, within a cell indicates how well that parameter captures the differences users perceived; (3) modality consistency: the degree to which the bars' up/down pattern translates vertically across rows. When similar (and not flat), the proxy translations are being interpreted by users in the same way, providing another level of validation. We structure our discussion around how the three modalities represent the different rating scales. We refer to the number of *equivalents* and *differents* in a given cell as $[x{:}z]$, with y = number of *uncertains*, and $x+y+z = 10$.

Duration and Pleasantness were translatable

Duration was comparably translatable for LOFIVIB [5:1] and VIS_{EMPH} [6:1]; VIS_{DIR} was less consistent [7:3] (two differences very large). Between the three modalities, 9/10 vibrations achieved equivalence with at least one modality. For Duration, this is unsurprising. It is a physical property that is controllable through the Android vibration API, and both visualization methods explicitly present Duration as their x-axis. This information was apparently not lost in translation.

More surprisingly, Pleasantness fared only slightly worse for LOFIVIB [4:2] and VIS_{EMPH} [4:1]; 8 / 10 vibrations had at least one modality that provided equivalence. Pleasantness is a higher-level affective feature than Duration. Although not an absolute victory, this result gives evidence that, with improvement, crowdsourcing may be a viable method of feedback for at least one affective parameter.

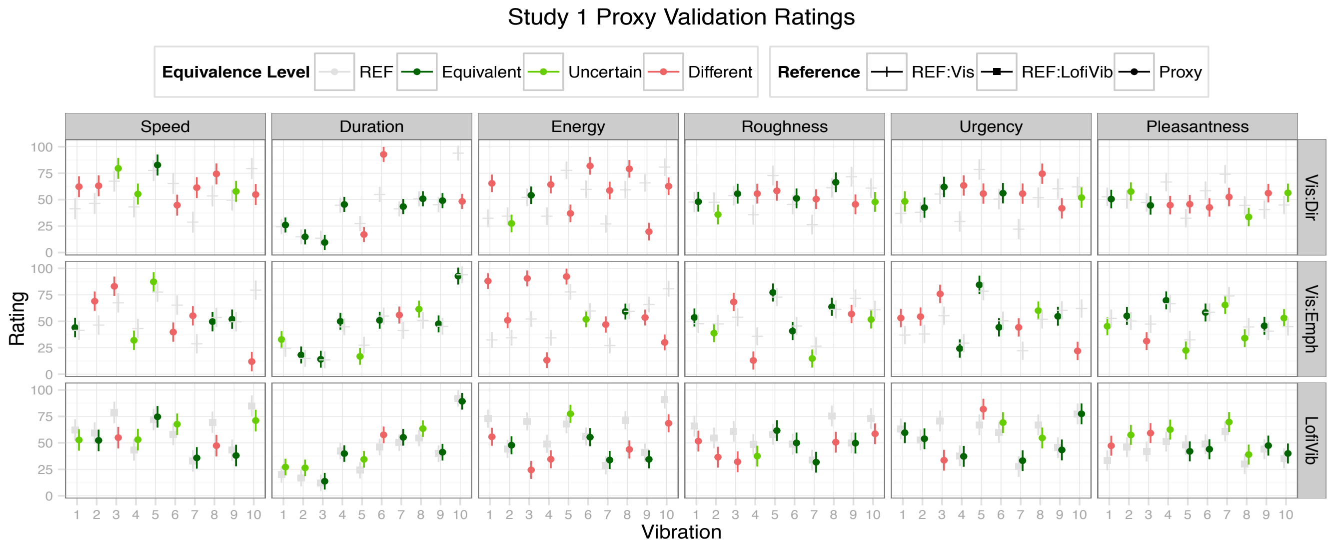


Figure 8: 95% confidence intervals and equivalence test results for Study 1 - Proxy Validation. Grey represents Ref ratings. Dark green maps equivalence within our defined threshold, and red a statistical difference indicating an introduced bias; light green results are inconclusive. Within each cell, variation of Ref ratings means vibrations were rated differently compared to each other, suggesting they have different perceptual features and represent a varied set of source stimuli.

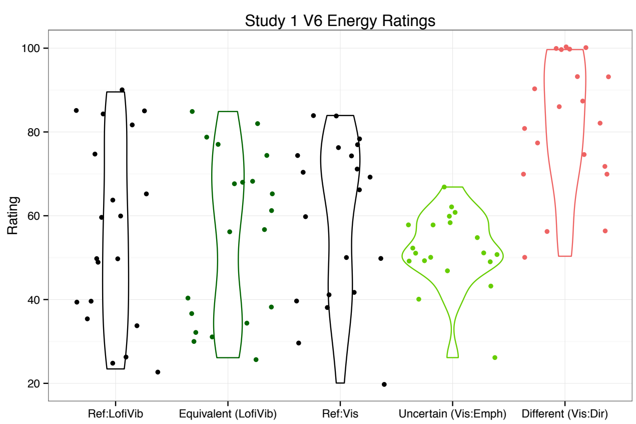


Figure 9: Rating distributions from Study 1, using V6 Energy as an example. These violin plots illustrate 1) the large variance in participant ratings, and 2) how equivalence thresholds reflect the data. When equivalent, proxy ratings are visibly similar to Ref. When uncertain, ratings follow a distribution with unclear differences. When different, there is a clear shift.

Speed and Urgency translated better with LofiVib

LofiVib was effective at representing Urgency [6:2]; Vis$_{EMPH}$ attained only [4:5], and Vis$_{DIR}$ [3:5]. Speed was less translatable. LofiVib did best at [4:2]; Vis$_{DIR}$ reached only [1:6], and Vis$_{EMPH}$ [3:5]. However, the modalities again complemented each other. Of the three, 9/10 vibrations were equivalent at least once for Urgency (V8 was not). Speed had less coverage: 6/10 had equivalencies (V3,4,6,10 did not).

Roughness had mixed results; best with Vis$_{EMPH}$

Roughness ratings varied heavily by vibration. 7 vibrations had at least one equivalence (V2,4,10 did not). All modalities had 4 equivalencies each: Vis$_{EMPH}$ [4:3], Vis$_{DIR}$ [4:4], and LofiVib [4:5].

Energy was most challenging

Like Roughness, 7 vibrations had at least one equivalence between modalities (V1,4,10 did not). LofiVib [4:5] did best with Energy; Vis$_{EMPH}$ and Vis$_{DIR}$ struggled at [1:8].

Emphasized visualization outperformed direct plot

Though it depended on the vibration, Vis$_{EMPH}$ outperformed Vis$_{DIR}$ for most metrics, having the same or better equivalencies/differences for Speed, Energy, Roughness, Urgency, and Pleasantness. Duration was the only mixed result, as Vis$_{DIR}$ had both more equivalencies and more differences [7:3] versus [6:1] In addition, 16/20 participants (80%) preferred Vis$_{EMPH}$ to Vis$_{DIR}$. Although not always clear-cut, these comparisons overall indicate that our Vis$_{EMPH}$ visualization method communicated these affective qualities more effectively than the status quo. This supports our approach to emphasized visualization, and motivates the future pursuit of other visualizations.

V4,V10 difficult, V9 easy to translate

While most vibrations had at least one equivalency for 5 rating scales, V4 and V10 only had 3. V4 and V10 had no equivalences at all for Speed, Roughness, and Energy, making them some of the most difficult vibrations to translate. V4's visualization had very straight lines, perhaps downplaying its texture. V10 was by far the longest vibration, at 13.5s (next longest was V8 with 4.4s). Its length may have similarly masked textural features.

V8 was not found to be equivalent for Urgency and Pleasantness. V8 is an extremely irregular vibration, with a varied rhythm and amplitude, and the second longest. This may have made it difficult to glean more intentional qualities like Urgency and Pleasantness. However, it was only found to be different for VIS_{DIR}/Urgency, so we cannot conclude that significant biases exist.

By contrast, V9 was the only vibration that had an equivalency for every rating scale, and in fact could be represented across all ratings with LOFIVIB. V9 was a set of distinct pulses, with no dynamic ramps; it thus may have been well suited to translation to LOFIVIB.

Summary

In general, these results indicate promise, but also need improvement and combination of proxy modalities. Unsurprisingly, participant ratings varied, reducing confidence and increasing the width of confidence intervals (indeed, this is partial motivation to access larger samples). Even so, both differences and equivalencies were found in every rating/proxy modality pairing. Most vibrations were equivalent with at least one modality, suggesting that we might pick an appropriate proxy modality depending on the vibration; we discuss the idea of triangulation in more detail later. Duration and Pleasantness were fairly well represented, Urgency and Speed were captured best by LOFIVIB, and Roughness was mixed. Energy was particularly difficult to represent with these modalities. We also find that results varied depending on vibration, meaning that more analysis into what makes vibrations easier or more difficult to represent could be helpful.

Though we were able to represent several features using proxy modalities within a bounded error rate, this alone does not mean they are crowdsource-friendly. All results from Study 1 were gathered in-lab, a more controlled environment than over MTurk. We thus ran a second study to validate our proxy modality ratings when deployed remotely.

STUDY 2: DEPLOYMENT VALIDATION WITH MTURK (G2)

To determine whether rating of a proxy is similar when gathered locally or remotely, we deployed the same computer-run proxy modality surveys on MTurk. We wanted to discover the challenges all through the pipeline for running a VT study on MTurk, including larger variations in phone actuators and experimental conditions (G4). We purposefully did not iterate on our proxy vibrations or survey, despite identifying many ways to improve them, to avoid creating a confound in comparing results of the two studies.

The visualization proxies were run as a single MTurk Human Intelligence Task (HIT), counterbalanced for order; the LOFIVIB survey was deployed as its own HIT. Each HIT was estimated at 30m, for which participants received $2.25 USD. In comparison, Study 1 participants were estimated to take 1 hour and received $10 CAD. We anticipated a discrepancy in average task time due to a lack of direct supervision for the MTurk participants, and expected this to lead to less accurate participant responses, prompting the lower payrate. On average, it took 7m for participants to complete the HIT while local study participants took 30m.

We initially accepted participants of any HIT approval rate to maximize recruitment in a short timeframe. Participants were post-screened to prevent participation in both studies. 49 participants were recruited. No post-screening was used for the visual sub-study. For the LOFIVIB proxy survey, we post-screened to verify device used [33]. We asked participants (a) confirm their study completion with an Android device via a survey question (b) detected actual device via FluidSurvey's OS-check feature, and (c) rejected inconsistent samples (eg. 9 used non-Android platforms for LOFIVIB). Of the included data, 20 participants participated each in the visual proxy condition (6F) and the LOFIVIB condition (9F).

For both studies, Study 1's data was used as a "gold standard" that served as a baseline comparison with the more reliable local participant ratings [3]. We compared the remote proxy results (from MTurk) to the REF results gathered in Study 1, using the same analysis methods.

Results

Study 2 results appear in Figure 10, which compares remotely collected ratings with locally collected ratings for the respective reference (the same reference as for Figure 8). It can be read the same way, but adds information. Based an analysis of a different comparison, a red star indicates a statistically significant difference between remote proxy ratings and corresponding local *proxy* ratings. This analysis revealed that ratings for the same proxy gathered remotely and locally disagreed 21 times (stars) out of 180 rating/modality/vibration combination; i.e., relatively infrequently.

Overall, we found similar results and patterns in Study 2 as for Study 1. The two figures show similar up/down rating patterns; the occasional exceptions correspond to red-starred items. Specific results varied, possibly due to statistical noise and rating variance. We draw similar conclusions: that proxy modalities can still be viable when deployed on MTurk, but require further development to be reliable in some cases.

DISCUSSION

Here we discuss high level implications from our findings and relate them to our study goals (G1-G4 in Introduction).

Proxy modalities are viable for crowdsourcing (G1,G2:feasibility)

Our studies showed that proxy modalities can represent affective qualities of vibrations within reasonably chosen error bounds, depending on the vibration. These results largely translate to deployment on MTurk. Together, these two steps indicate that proxy modalities are be a viable approach to crowdsourcing VT sensations, and can reach a usable state with a bounded design iteration (as outlined in the following sections). This evidence also suggests that we may be able to deploy directly to MTurk for future validation. Our two-step validation was important as a first look at whether ratings shift dramatically; and we saw no indications of bias or overall shift between locally running proxy modalities and remotely deploying them.

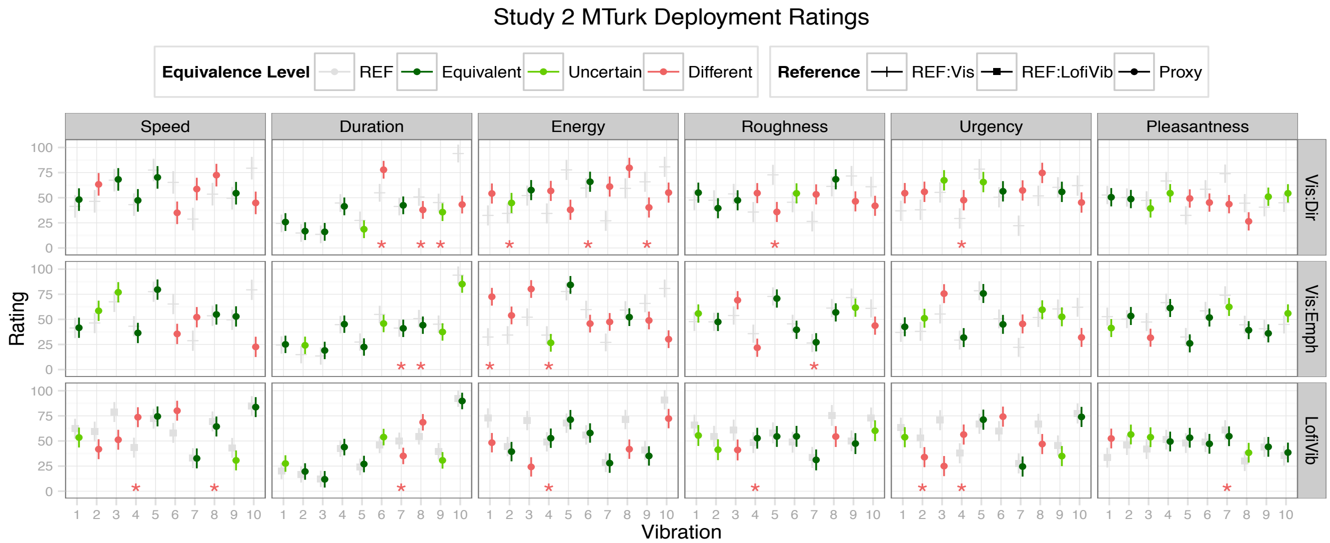


Figure 10: 95% Confidence Intervals and Equivalence Test Results for Study 2 - MTurk Deployment Validation. Equivalence is indicated with dark green, difference is indicated with red, and uncertainty with light green. Red star indicates statistically significant difference between remote and local proxy ratings.

Triangulation (G3:promising directions/proxies)
Most vibrations received equivalent ratings for most scales in at least one proxy modality. Using proxy modalities in tandem might help improve response accuracy. For example, V6 could be rendered with LOFIVIB for a pleasantness rating, then as VIS_{EMPH} for Urgency. Alternatively, we might develop an improved proxy vibration by combining modalities - a visualization with an accompanying low-fidelity vibration.

Animate visualizations (G3:promising directions)
Speed and Urgency were not as effectively transmitted with our visualizations as with our vibration. Nor was Duration well portrayed with VIS_{DIR}, which had a shorter time axis than the exaggerated VIS_{EMPH}. It may be more difficult for visual representations to portray time effectively: perhaps it is hard for users to distinguish Speed/Urgency, or the time axis is not at an effective granularity. Animations (e.g., adding a moving line to help indicate speed and urgency), might help to decouple these features. As with triangulation, this might also be accomplished through multimodal proxies which augment a visualization with a time-varying sense using sounds or vibration. Note, however, that Duration was more accurately portrayed by VIS_{EMPH}, suggesting that direct representation of physical features *can* be translated.

Sound could represent Energy (G3:promising directions)
Our high-fidelity reference is a voice-coil actuator, also used in audio applications. Indeed, in initial pilots we played vibration sound files through speakers. Sound is the closest to vibration in the literature, and a vibration signal's sound output is correlated with the vibration energy and sensation.

However, in our pilots, sometimes the vibration sound did not match the sensation; was not audible (low frequency vibrations); or the C2 could only play part of the sound (i.e, the sound was louder than the sensation).

Thus, while the raw sound files are not directly translatable, a sound proxy definitely has potential. It could, for example, supplement where the VIS_{DIR} waveform failed to perform well on any metric (aside from Duration) but a more expressive visual proxy (VIS_{EMPH}) performed better.

Device dependency and need for Energy model for Vibrations (G4:challenges)
Energy did not translate well. This could be a linguistic confusion, but also a failure to translate this feature. For the visualization proxies, it may be a matter of finding the right representation, which we continue to work on.

However, with LOFIVIB, this represents a more fundamental tradeoff due to characteristics of phone actuators, which have less control over energy output than we do with a dedicated and more powerful C2 tactor. The highest vibration energy available in phones is lower than for the C2; this additional power obviously extends expressive range. Furthermore, vibration energy and time are coupled in phone actuators: the less time the actuator is on, the lower the vibration energy. As a result, it is difficult to have a very short pulses with very high energy (V1,V3,V8). The C2's voice coil technology does not have this duty-cycle derived coupling. Finally, the granularity of the energy dimension is coarser for phone actuators. This results in a tradeoff for designing (for example) a ramp sensation: if you aim for accurate timing, the resulting vibration would have a lower energy (V10). If you match the energy, the vibration will be longer.

Knowing these tradeoffs, designers and researchers can adjust their designs to obtain more accurate results on their intended metric. Perhaps multiple LoFiVib translations can be developed which maintain different qualities (one optimized on timing and rhythm, the other on energy). In both these cases, accurate models for rendering these features will be essential.

VT affective ratings are generally noisy (G4:challenges)

Taken as a group, participants were not highly consistent among one another when rating these affective studies, whether local or remote. This is in line with previous work [43], and highlights a need to further develop rating scales for affective touch. Larger sample sizes, perhaps gathered through crowdsourcing, may help reduce or characterize this error. Alternatively, it gives support to the need to develop mechanisms for individual customization. If there are "types" of users who do share preferences and interpretations, crowdsourcing can help with this as well.

Response & data quality for MTurk LoFiVib vibrations (G4:challenges)

When deploying vibrations over MTurk, 8/29 participants (approximately 31%) completed the survey using non-Android based OSes (Mac OS X, Windows 7,8.1, NT) despite these requirements being listed in the HIT and the survey. One participant reported not being able to feel the vibrations despite using an Android phone. This suggests that enforcing a remote survey to be taken on the phone is challenging, and that additional screens are needed to identify participants not on a particular platform. Future work might investigate additional diagnostic tools to ensure that vibrations are being generated, through programmatic screening of platforms, well-worded questions and instructions, and (possibly) ways of detecting vibrations actually being played, perhaps through the microphone or accelerometer).

Automatic translation (G4:challenges)

Our proxy vibrations were developed by hand, to focus on the feasibility of crowdsourcing. However, this additional effort poses a barrier for designers that might negate the benefits of using a platform of MTurk. As this approach becomes better defined, we anticipate automatic translation heuristics for proxy vibrations using validated algorithms. Although these might be challenging to develop for emotional features, physical properties like amplitude, frequency, or measures of energy and roughness would be a suitable first step. Indeed, crowdsourcing itself could be used to create these algorithms, as several candidates could be developed, their proxy vibrations deployed on MTurk, and the most promising algorithms later validated in lab.

Limitations

A potential confound was introduced by VIS_{EMPH} having a longer time axis than VIS_{DIR}: some of VIS_{EMPH}'s improvements could be due to seeing temporal features in higher resolution. This is exacerbated by V10 being notably longer than the next longest vibration, V8 (13.5s vs. 4.4s), further reducing temporal resolution vibrations other than V10.

We presented ratings to participants by-vibration rather than by-rating. Because participants generated all ratings for a single vibration at the same time, it is possible there are correlations between the different metrics. We chose this arrangement because piloting suggested it was less cognitively demanding than presenting metrics separately for each vibration. Future work can help decide whether correlations exist between metrics, and whether these are an artifact of stimulus presentation or an underlying aspect of the touch aesthetic.

Despite MTurk's ability to recruit more participants, we used the same sample size of 40 across both studies. While our proxies seemed viable for remote deployment, there were many unknown factors in MTurk user behaviour at the time of deployment. We could not justify more effort without experiencing these factors firsthand. Thus, we decided to use a minimal sample size for the MTurk study that was statistically comparable to the local studies. In order to justify a larger remote sample size in the future, we believe it is best to iterate the rating scales and to test different sets of candidate modalities.

As discussed, we investigated two proxy modalities in this first examination but look forward to examining others (sound, text, or video) alone or in combination.

CONCLUSION

In this paper, we crowdsourced high-level parameter feedback on VT sensations using a new method of *proxy vibrations*. We translated our initial set of high-fidelity vibrations, suitable for wearables or other haptic interactions, into two proxy modalities: a new VT visualization method, and low-fidelity vibrations on phones.

We established the most high-risk aspects of VT proxies, namely feasibility in conveying affective properties, and consistent local and remote deployment with two user studies. Finally, we highlighted promising directions and challenges of VT proxies, to guide future tactile crowdsourcing developments, targeted to empower VT designers with the benefits crowdsourcing brings.

ACKNOWLEDGMENTS

We are grateful for participant feedback, reviewer suggestions, and extra tactors shared by Hong Tan. Research was funded by NSERC and conducted under UBC BREB H13-01646.

REFERENCES

1. Md. Abdur Rahman, Abdulmajeed Alkhaldi, Jongeun Cha, and Abdulmotaleb El Saddik. 2010. Adding haptic feature to YouTube. In *Proceedings of the international conference on Multimedia - MM '10*. ACM Press, New York, New York, USA, 1643–1646. DOI: http://dx.doi.org/10.1145/1873951.1874310

2. M. Allen, J. Gluck, K. E. MacLean, and E. Tang. 2005. An Initial Usability Assessment for Symbolic Haptic Rendering of Music Parameters. In *7th Int'l Conf on Multimodal Interfaces (ICMI '05)*. Trento, Italy, 244–251.

3. Amazon.com Inc. 2015. Amazon Mechanical Turk Requester Best Practices Guide. (2015).
4. Farah Arab, Sabrina Paneels, Margarita Anastassova, Stephanie Coeugnet, Fanny Le Morellec, Aurelie Dommes, and Aline Chevalier. 2015. Haptic patterns and older adults: To repeat or not to repeat?. In *2015 IEEE World Haptics Conference (WHC)*. IEEE, 248–253. DOI: http://dx.doi.org/10.1109/WHC.2015.7177721
5. Stephen Brewster and Lorna M. Brown. 2004. Tactons: structured tactile messages for non-visual information display. In *AUIC '04*. 15–23. http://dl.acm.org/citation.cfm?id=976310.976313
6. Lorna M. Brown, Stephen A. Brewster, and Helen C. Purchase. 2006. Multidimensional tactons for non-visual information presentation in mobile devices. In *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services (MobileHCI '06)*. ACM Press, New York, USA, 231–238. DOI: http://dx.doi.org/10.1145/1152215.1152265
7. Lorna M. Brown and Topi Kaaresoja. 2006. Feel who's talking: Using Tactons for Mobile Phone Alerts. In *CHI '06 extended abstracts on Human factors in computing systems - CHI EA '06*. ACM Press, New York, New York, USA, 604. DOI: http://dx.doi.org/10.1145/1125451.1125577
8. L. Brunet, C. Megard, S. Paneels, G. Changeon, J. Lozada, M. P. Daniel, and F. Darses. 2013. "Invitation to the voyage": The design of tactile metaphors to fulfill occasional travelers' needs in transportation networks. In *2013 World Haptics Conference (WHC)*. IEEE, 259–264. DOI: http://dx.doi.org/10.1109/WHC.2013.6548418
9. Mark Cartwright and Bryan Pardo. 2015. VocalSketch: Vocally Imitating Audio Concepts. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. ACM Press, New York, New York, USA, 43–46. DOI: http://dx.doi.org/10.1145/2702123.2702387
10. A Chan, K MacLean, and J McGrenere. 2008. Designing haptic icons to support collaborative turn-taking. *International Journal of Human-Computer Studies* 66, 6 (2008), 333–355. http://www.sciencedirect.com/science/article/pii/S1071581907001565
11. Lydia B. Chilton, Greg Little, Darren Edge, Daniel S. Weld, and James A. Landay. 2013. Cascade: Crowdsourcing Taxonomy Creation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 1999–2008. DOI: http://dx.doi.org/10.1145/2470654.2466265
12. Julie S. Downs, Mandy B. Holbrook, Steve Sheng, and Lorrie Faith Cranor. 2010. Are your participants gaming the system?. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*. ACM Press, New York, New York, USA, 2399. DOI: http://dx.doi.org/10.1145/1753326.1753688
13. Mohamad Eid, Atif Alamri, and Abdulmotaleb Saddik. 2006. MPEG-7 Description of Haptic Applications Using HAML. In *2006 IEEE International Workshop on Haptic Audio Visual Environments and their Applications (HAVE 2006)*. IEEE, 134–139. DOI: http://dx.doi.org/10.1109/HAVE.2006.283780
14. Mohamad Eid, Sheldon Andrews, Atif Alamri, and Abdulmotaleb El Saddik. 2008. HAMLAT: A HAML-Based Authoring Tool for Haptic Application Development. In *LNCS 5024 - Haptics: Perception, Devices and Scenarios*. Vol. 5024. 857–866. DOI: http://dx.doi.org/10.1007/978-3-540-69057-3
15. Adrienne Porter Felt, Serge Egelman, and David Wagner. 2012. I've got 99 problems, but vibration ain't one: A survey of smartphone User's concerns. In *Proceedings of the second ACM workshop on Security and privacy in smartphones and mobile devices - SPSM '12*. ACM Press, New York, New York, USA, 33. DOI: http://dx.doi.org/10.1145/2381934.2381943
16. Steve Guest, JeanMarc Dessirier, Anahit Mehrabyan, Francis McGlone, Greg Essick, George Gescheider, Anne Fontana, Rui Xiong, Rochelle Ackerley, and Kevin Blot. 2011. The development and validation of sensory and emotional scales of touch perception. *Attention, Perception, & Psychophysics* 73, 2 (2011), 531–550. DOI: http://dx.doi.org/10.3758/s13414-010-0037-y
17. Vincent Hayward. 2008. A brief taxonomy of tactile illusions and demonstrations that can be done in a hardware store. *Brain research bulletin* 75, 6 (April 2008), 742–52. DOI: http://dx.doi.org/10.1016/j.brainresbull.2008.01.008
18. Jeffrey Heer and Michael Bostock. 2010. Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 203–212. DOI: http://dx.doi.org/10.1145/1753326.1753357
19. Eve Hoggan, Craig Stewart, Laura Haverinen, Giulio Jacucci, and Vuokko Lantz. 2012. Pressages: augmenting phone calls with non-verbal messages. (Oct. 2012), 555–562. DOI: http://dx.doi.org/10.1145/2380116.2380185
20. M. Hollins, R. Faldowski, S. Rao, and F Young. 1993. Perceptual dimensions of tactile surface texture: A multidimensional scaling analysis. *Perception & Psychophysics* 54 (1993), 697–705.
21. Inwook Hwang, K E MacLean, Matthew Brehmer, Jeff Hendy, Andreas Sotirakopoulos, and Seungmoon Choi. 2011. The Haptic Crayola Effect: Exploring the Role of Naming in Learning Haptic Stimuli. In *WorldHaptics '11*. IEEE, Istanbul, 385–390.

22. Inwook Hwang, Jongman Seo, Myongchan Kim, and Seungmoon Choi. 2013. Vibrotactile Perceived Intensity for Mobile Devices as a Function of Direction, Amplitude, and Frequency. *IEEE Transactions on Haptics* 6, 3 (July 2013), 352–362. DOI: http://dx.doi.org/10.1109/TOH.2013.2

23. Ali Israr and Ivan Poupyrev. 2011. Tactile brush: drawing on skin with a tactile grid display. In *CHI '11*. ACM Press, Vancouver, BC, 2019–2028. DOI: http://dx.doi.org/10.1145/1978942.1979235

24. Ali Israr, Siyan Zhao, and Oliver Schneider. 2015. Exploring Embedded Haptics for Social Networking and Interactions. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '15*. ACM Press, New York, New York, USA, 1899–1904. DOI: http://dx.doi.org/10.1145/2702613.2732814

25. Ali Israr, Siyan Zhao, Kaitlyn Schwalje, Roberta Klatzky, and Jill Lehman. 2014. Feel Effects: Enriching Storytelling with Haptic Feedback. *Transactions on Applied Perception (TAP)* 11, 3 (2014).

26. Cathrine V. Jansson-Boyd. 2011. Touch matters: exploring the relationship between consumption and tactile interaction. *Social Semiotics* 21, 4 (Sept. 2011), 531–546. DOI: http://dx.doi.org/10.1080/10350330.2011.591996

27. T. Kaaresoja and J. Linjama. 2005. Perception of Short Tactile Pulses Generated by a Vibration Motor in a Mobile Phone. In *First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*. IEEE, 471–472. DOI: http://dx.doi.org/10.1109/WHC.2005.103

28. Yeongmi Kim, Jongeun Cha, Ian Oakley, and Jeha Ryu. 2009. Exploring Tactile Movies: An Initial Tactile Glove Design and Concept Evaluation. *IEEE Multimedia* PP, 99 (2009), 1. DOI: http://dx.doi.org/10.1109/MMUL.2009.63

29. Aniket Kittur, Ed H. Chi, and Bongwon Suh. 2008. Crowdsourcing User Studies with Mechanical Turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. ACM, New York, NY, USA, 453–456. DOI: http://dx.doi.org/10.1145/1357054.1357127

30. Emilia Koskinen, Topi Kaaresoja, and Pauli Laitinen. 2008. Feel-good Touch: Finding the Most Pleasant Tactile Feedback for a Mobile Touch Screen Button. In *Proceedings of the 10th International Conference on Multimodal Interfaces (ICMI '08)*. ACM, New York, NY, USA, 297–304. DOI: http://dx.doi.org/10.1145/1452392.1452453

31. Karon E MacLean. 2008. Foundations of Transparency in Tactile Information Design. *IEEE Transactions on Haptics* 1, 2 (July 2008), 84–95. DOI: http://dx.doi.org/10.1109/TOH.2008.20

32. Karon E MacLean and Mario Enriquez. 2003. Perceptual design of haptic icons. In *Proceedings of Eurohaptics Conference*. 351–363.

33. Winter Mason and Siddharth Suri. 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods* 44, 1 (2012), 1–23. DOI: http://dx.doi.org/10.3758/s13428-011-0124-6

34. Sile O Modhrain and Ian Oakley. 2001. Touch TV : Adding Feeling to Broadcast Media. (2001).

35. Marianna Obrist, Sue Ann Seah, and Sriram Subramanian. 2013. Talking about tactile experiences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM Press, New York, USA, 1659–1668. DOI: http://dx.doi.org/10.1145/2470654.2466220

36. Shogo Okamoto, Hikaru Nagano, and Yoji Yamada. 2013. Psychophysical Dimensions of Tactile Perception of Textures. *IEEE Transactions on Haptics* 6, 1 (Jan. 2013), 81–93. DOI: http://dx.doi.org/10.1109/TOH.2012.32

37. Jerome Pasquero, Joseph Luk, Shannon Little, and Karon Maclean. 2006. Perceptual Analysis of Haptic Icons: An Investigation into the Validity of Cluster Sorted MDS. In *14th IEEE Haptic Symposium (HAPTICS '06)*. IEEE, Alexandria, Virginia, 437–444.

38. Helena Pongrac. 2008. Vibrotactile perception: examining the coding of vibrations and the just noticeable difference under various conditions. *Multimedia Systems* 13, 4 (2008), 297–307. DOI: http://dx.doi.org/10.1007/s00530-007-0105-x

39. J. Raisamo, R. Raisamo, and V. Surakka. 2013. Comparison of Saltation, Amplitude Modulation, and a Hybrid Method of Vibrotactile Stimulation. *Haptics, IEEE Transactions on* 6, 4 (Oct 2013), 517–521. DOI: http://dx.doi.org/10.1109/TOH.2013.25

40. Oliver S. Schneider and Karon E. MacLean. 2014. Improvising Design with a Haptic Instrument. In *HAPTICS '14*. Houston, USA.

41. D.J. Schuirmann. 1981. On hypothesis testing to determine if the mean of a normal distribution is contained in a known interval. *Biometrics* 37, 3 (1981), 617–617.

42. Hasti Seifi, Chamila Anthonypillai, and Karon E. MacLean. 2014. End-user customization of affective tactile messages: A qualitative examination of tool parameters. In *HAPTICS '14*. IEEE, 251–256. DOI: http://dx.doi.org/10.1109/HAPTICS.2014.6775463

43. Hasti Seifi, Kailun Zhang, and Karon MacLean. 2015. VibViz: Organizing, visualizing and navigating vibration libraries. In *World Haptics '15*.

44. Jongman Seo and Seungmoon Choi. 2013. Perceptual analysis of vibrotactile flows on a mobile device. *IEEE transactions on haptics* 6, 4 (Jan. 2013), 522–7. DOI: http://dx.doi.org/10.1109/TOH.2013.24

45. Pao Siangliulue, Kenneth C. Arnold, Krzysztof Z. Gajos, and Steven P. Dow. 2015. Toward Collaborative Ideation at Scale - Leveraging Ideas from Others to Generate More Creative and Diverse Ideas Pao. In *CSCW '15*. ACM Press, New York, New York, USA, 937–945. DOI:http://dx.doi.org/10.1145/2675133.2675239

46. Diane Tam, Karon E. MacLean, Joanna McGrenere, and Katherine J. Kuchenbecker. 2013. The Design and Field Observation of a Haptic Notification System for Timing Awareness During Oral Presentations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 1689–1698. DOI: http://dx.doi.org/10.1145/2470654.2466223

47. David Ternes and Karon E. MacLean. 2008. Designing Large Sets of Haptic Icons with Rhythm. In *Haptics: Perception, Devices and Scenarios*, Manuel Ferre (Ed.). Lecture Notes in Computer Science, Vol. 5024. Springer Berlin Heidelberg, 199–208. DOI: http://dx.doi.org/10.1007/978-3-540-69057-3_24

48. Jan BF van Erp and Michiel MA Spapé. 2003. Distilling the underlying dimensions of tactile melodies. In *Proceedings of Eurohaptics*, Vol. 2003. 111–120.

49. Anbang Xu, Shih-Wen Huang, and Brian Bailey. 2014. Voyant: Generating Structured Feedback on Visual Designs Using a Crowd of Non-Experts. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing - CSCW '14*. ACM Press, New York, New York, USA, 1433–1444. DOI:http://dx.doi.org/10.1145/2531602.2531604

50. Yongjae Yoo, Taekbeom Yoo, Jihyun Kong, and Seungmoon Choi. 2015. Emotional responses of tactile icons: Effects of amplitude, frequency, duration, and envelope. In *IEEEWorld Haptics Conference (WHC'15)*. 235–240. DOI: http://dx.doi.org/10.1109/WHC.2015.7177719

51. Siyan Zhao, Oliver Schneider, Roberta Klatzky, Jill Lehman, and Ali Israr. 2014. FeelCraft: Crafting Tactile Experiences for Media Using a Feel Effect Library. In *Proceedings of the Adjunct Publication of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST'14 Adjunct)*. ACM, New York, NY, USA, 51–52. DOI: http://dx.doi.org/10.1145/2658779.2659109