Investigating the Effect of Polarity in Auditory and Vibrotactile Displays Under Cognitive Load

Jamie Ferguson
University of Glasgow
Glasgow, Scotland
jamie.f.ferguson@glasgow.ac.uk

Euan Freeman
University of Glasgow
Glasgow, Scotland
euan.freeman@glasgow.ac.uk

Stephen Brewster
University of Glasgow
Glasgow, Scotland
stephen.brewster@glasgow.ac.uk

ABSTRACT

When users are undertaking mentally demanding visuals tasks, it can be beneficial to convey information through the auditory or tactile modality instead. A fundamental problem when mapping information to sound or vibration is establishing which polarity the mapping should use. Magnitude estimation is a popular method of establishing polarity preferences, however the effectiveness of this approach remains unclear, especially in more ecologically valid contexts. We investigate what impact the polarity of a data-sound or data-vibration mapping has on how well users can interpret these mappings, under two different levels of mental workload. Our results show that polarity does not affect error rate or cognitive workload, although may affect response time. We also found that induced cognitive load may influence usability. An implication of this is that commonly used methods of establishing data mappings need to be revisited, with cognitive load in mind, to help designers create more usable auditory and vibrotactile displays.

CCS CONCEPTS

• **Human-centered computing → Auditory feedback**; **Haptic devices**.

KEYWORDS

Auditory Display, Vibrotactile Display, Polarity, Sonification, Vibrotactile, Cognitive Load

ACM Reference Format:
Jamie Ferguson, Euan Freeman, and Stephen Brewster. 2021. Investigating the Effect of Polarity in Auditory and Vibrotactile Displays Under Cognitive Load. In *Proceedings of the 2021 International Conference on Multimodal Interaction (ICMI '21), October 18–22, 2021, Montréal, QC, Canada.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3462244.3479911

1 INTRODUCTION

When users are engaged in visually-demanding tasks, it can be beneficial to present information through non-visual modalities instead. This can be achieved by mapping information onto other sensory modalities, e.g., as audio or vibration. This can make information more noticeable whilst driving [35, 36, 41], during complex tasks like air-traffic control [1, 8, 9, 37], when users have limited vision [21], or when using devices with little screen space [4, 17, 18]. A fundamental design challenge when doing this is choosing suitable parameter mappings for each dimension of information. A poorly designed mapping may not only reduce information accuracy [47], but may have severe consequences. This can be seen from many documented cases of aircraft pilots, nuclear power system operators, and military personnel turning off auditory alarms which are unpleasant, obstructive, or inaccurate [40].

A widely used method for informing the design of a data-sound or data-vibration mapping is magnitude estimation [2, 14, 15, 44, 45, 47]. This technique scales the relationship between an audio or vibrotactile stimulus and an information dimension (e.g., speed, temperature), helping designers establish the *polarity* and *scale* of the mapping. Polarity is the direction of the relationship and scale determines how much the perceived value changes as the actual stimulus parameter changes [44]. User consensus for polarity helps establish which mapping design is most 'intuitive'. These studies have shown the importance of polarity in perception of a parameter mapping; however, it is unclear if this technique also leads to designs that are usable in their intended usage context.

Designing a mapping with an appropriate and *usable* polarity is critical in high-stakes safety-critical contexts where users are under high cognitive load (e.g., use of data-sound or data-vibration mappings in process monitoring, aircraft cockpits, in cars, etc). Usability is vital because users may need to react quickly and instinctively. Conversely, in low-stakes contexts with low-cognitive load, the appropriateness of the mapping polarity is less critical, because users have time to process and learn a mapping, and understand that the information being conveyed is not safety-critical. Conducting a magnitude estimation experiment to establish polarity consensus is costly; if designers understand how the usability of a system may be affected by mappings created using this method, then they can decide if gathering this data is worth the time and cost.

We investigate the effect of polarity on usability in data-sound and data-vibration mappings, under varying levels of cognitive load. Our goal is to establish if polarity mappings based on consensus from magnitude estimation studies lead to increased performance, versus seemingly unintuitive polarity designs, and to see how these perform under cognitive load. We conducted an experiment with the N-back task to induce controlled levels of mental workload [32]. This was used alongside a secondary task where users responded to acoustic or vibrotactile stimuli with opposing polarities. Our results have implications for the design and evaluation of data-sound and data-vibration mappings, which are increasingly being used in safety-critical usage contexts where good usability is crucial.

ICMI '21, October 18–22, 2021, Montréal, QC, Canada

2 BACKGROUND

2.1 Data-Sound Mappings

Walker and Kramer's early studies [46, 47] of data-sound mappings found that those that were predicted to perform well (e.g., mapping temperature to pitch) were not always better than those that were predicted to perform poorly, or even mappings that were randomly designed. They posited the explanation that a listener's mental models of data-sound relationships had a large effect on their results. For example, *slow* acoustic changes, like slow tempos and onsets, were perceived to be representative of *larger* objects, as would be the case with larger objects in the physical world, e.g., as larger objects move more slowly due to inertia.

Continuing this, Walker [44, 45] proposed magnitude estimation as a method for evaluating data-sound mappings. This is a standard method of psychophysical scaling, which maps the relationship between the magnitude of a sensory stimulus and its associated perceived intensity [43], resulting in a power function between the actual stimulus magnitude and the perceived magnitude.

Magnitude estimation is now often used when studying the perception of auditory stimuli in a wide array of contexts, e.g., ergonomics [27] and medicine [29]. Walker's results showed that magnitude estimation provides reliable measurement of both the *scaling* function and *polarity* of a given data-sound mapping. Walker suggested that consensus over polarity will indicate how 'natural' or 'intuitive' a mapping is.

Recent research [14] has further investigated the importance of mental models in listener preference of polarity in data-sound mappings. In that work, a study was conducted using the same approach as Walker [44, 45], but instead focused on data-sound mappings where the data dimension had a negative valence (error, danger, and stress) and the acoustic parameters were 'undesirable' in terms of western musical tradition (roughness/disharmony and noise). Polarity preference results showed that the majority of participants perceived these mappings in a positive polarity, meaning that as the 'undesirable' acoustic parameter increased, participants perceived an increase in the negative valence information parameter.

These works give insight into how data-sound mappings are interpreted and hint at the role of polarity in their perception. If mental models are consistent, then magnitude estimation studies will likely lead to consensus about the 'best' mapping of scale and polarity. It seems reasonable to expect such mappings to be more usable if they are inherently 'intuitive', but this is not known.

2.2 Data-Vibration Mappings

Vibration is often used to attract attention, e.g., for notifications, or to give confirmatory feedback about interactions. However, vibration also can be used to encode information. Structured vibrations can manipulate properties like frequency, amplitude and rhythm to encode and convey a variety of data dimensions [5, 6, 19, 20, 24, 25], including multidimensional mappings [7].

Such data-vibration mappings are often abstract, mapping properties of vibration to arbitrary categories, e.g., types of fruit [11] and notification categories [6]. Vibration can effectively convey affective qualities and emotions [6, 35, 36, 39, 49, 51], suggesting that users have mental models to aide their understanding vibration cues, similar to data-sound mappings. For example, rough or strong vibrations were perceived as more alarming or urgent [6, 36, 39] (analogous to rougher or more intense tactile sensations being indicative of dangers in the physical world) and increasingly intense vibrations were associated with unpleasantness [49, 51].

Research has shown that the magnitude estimation approach used in auditory research [44, 45] can provide similar insights into how a data-vibration mapping is perceived [15], informing the selection of scale and polarity. Again, the assumption is that mental models will drive consensus towards an 'intuitive' mapping.

2.3 Data Mapping Polarity and Usability

How the polarity of a mapping between a data and sensory dimension is perceived is a significant design problem. For example, individual differences in how sounds are perceived are a persistent challenge in selecting polarities for data-sound mappings [33]. In a visual display, *up* represents *more* in most cases, but in the auditory or tactile modality, this is less clear; humans have less metaphors in the auditory or tactile modalities.

Similarly, the context in which a mapping is situated can impact how it is perceived. Consider a Geiger counter, a simple metaphor often used in data-sound and data-vibration mapping designs, where the repetition rate of a signal is used to convey information. This can be perceived differently in different contexts, with both positive or negative valence. For example, in a Geiger counter, faster 'clicking' means the user is getting closer to a source of ionising radiation; as something inherently dangerous, this would be considered to have *negative* valence. Conversely, the Geiger counter metaphor is used successfully in human-computer interaction contexts like gesture interaction [18], navigation [26, 42], and reaching [48], where faster auditory or vibrotactile 'clicking' means the user is doing the correct action (e.g., interacting correctly [18] or moving towards a target [26, 42, 48]); here, correctness implies *positive* valence. Thus it is important to consider the context in which a data-sound or data-vibration will be used, as this will impact how 'intuitive' it is.

There is limited research investigating the effect of polarity on the usability of a data-sound or data-vibration mapping. One study suggested that for a basic auditory display task, polarity did not affect task performance in simple circumstances [12]. Auditory and vibrotactile displays are being increasingly used in more demanding contexts, where non-visual data mappings can reduce visual demand. We address this gap in the literature by investigating the relationship between polarity and usability in a cognitively demanding context, to see how this affects perception and usability.

3 EXPERIMENT

3.1 Aims

We ran an experiment to investigate the impact the polarity of a data-sound and data-vibration mapping has on its ability to convey information, when the user is under different degrees of cognitive load. We used a simple working memory task with two difficulties, to induce low and high levels of cognitive load. At the same time, participants were asked to classify auditory and vibrotactile alarms. These alarms were simple parameter mappings where the sensory parameter (e.g., acoustic noise) was mapped to four levels of danger; when participants are presented with an alarm, they must respond with the danger level it represents. This dual task design allows us to investigate the mapping usability in a demanding context.

We chose danger as the information parameter for auditory and vibrotactile stimuli for two reasons. First, there are numerous potential use cases where 'danger' could conceivably be represented through an auditory or vibrotactile alarm, e.g., in hospitals, surgery, in-car, in-aircraft and process monitoring. Second, the results from the previous works that our mappings were based on [14, 15] also used danger as the information parameter, with polarity data for several data-sound and data-vibration mappings.

3.2 Manipulating Cognitive Load

To induce cognitive load, we used a visual N-back task [32] with two levels of difficulty (N=0 and N=2). This is a common protocol for inducing cognitive load in neuroimaging [34] and has been used successfully in human-computer interaction studies [22, 30]. In the N-back task, participants are presented with a series of symbols (usually numbers), one at a time. On each presentation, they must respond with whether the current stimulus matches a stimulus that they saw N presentations back in the sequence.

For the 2-back task, participants must indicate if the current number they see matches the one they saw two trials ago. For the 0-back task, participants need to compare each stimulus with the one that they saw first in the sequence; this is a matching task, unlike the 2-back task where working memory is being constantly updated. The advantage of the N-back task is that the perceptual and motor demand remains constant across difficulty levels.

3.3 Design

We used a within-subjects design with three independent variables: (1) the audio or vibration parameter (acoustic roughness, acoustic noise, vibration pulse tempo and vibration duration); (2) polarity of the data-sound or data-vibration mapping (aligned or inverted); and (3) the level of the N-back task (0-back or 2-back). This gave 16 conditions (4 parameters × 2 polarities × 2 N-back levels).

There was one block of trials for each condition. Participants completed both levels of the N-back task before continuing to the next combination. For example, if the current combination was {roughness × aligned polarity}, the participant would complete the 0-back and 2-back tasks for that combination together. There were eight trials in each block: two each for the four danger levels.

The N-back order was randomised. The order in which polarities were presented could have a confounding effect, as participants may be biased towards whichever they experienced first. Therefore, the order of each polarity × parameter pairing was counterbalanced.

3.4 Procedure

At the beginning of each block, participants were given an explanation of the acoustic or vibrotactile parameter for that block, and were told how it mapped to danger level. They could press buttons to hear/feel the cues for each level. They needed to hear/feel each level at least once before moving to the next step, but could repeat the cues as many times as they wished. They were also told which N-back level they were about to undertake. Finally, there were separate practice trials for the N-back task and classification task, followed by a combined practice task.

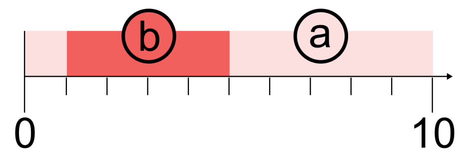

Figure 1: Audio/vibration alarms were presented within a 10 s window (a); there was at least 1 s gap at the start and at least 5 s gap at the end, leaving a 4 s window (b) in which the stimuli would be presented.

For each trial, participants carried out the N-back task whilst classifying auditory or vibrotactile stimuli in terms of the 'danger' they represent. Each task started after participants indicated they were ready. We measured accuracy of classification response, accuracy of N-back task response, and reaction times. Immediately after the final task in each condition, participants completed the NASA Task-Load Index (TLX) survey [23], as a subjective measure of perceived workload.

3.4.1 N-back Task. Numbers from 1–9 were presented in random order for the symbol sequence. Each number was displayed in the centre of the screen for one second, then the screen went blank for three seconds before the next number appeared (as per [22]). Participants indicated if a given number on-screen was a match or non-match via a keypress. Each N-back task sequence had 21 numbers, with the constraint that one third of these were matches.

3.4.2 Audio/Vibration Classification Task. Concurrently with the N-back task, the audio or vibration alarms were presented every ten seconds. The time at which they were presented within that ten second window was randomised, so participants would not anticipate them. There was at least one second gap at the start, and at least five seconds after the onset so there was sufficient time to respond. Thus, there was a four second interval in which the stimulus could be randomly presented. When an alarm stimulus was presented, participants responded with the appropriate number key to indicate what level they believed the alarm to represent.

3.5 Stimuli

Our acoustic and vibrotactile stimuli were based on magnitude estimation studies that obtained polarity preferences for a number of data-sound [14] and data-vibration [15] mappings. Each parameter had four levels for the four 'danger' levels. The mappings as taken from prior were used as the aligned polarities. We also inverted these, giving the other dimension of the polarity variable.

For example, in work by Ferguson et al. [14], the most popular polarity choice for a mapping between acoustic noise and danger was a *positive* polarity, meaning the more noise present in an audio signal, the higher the perceived value of danger. The inverse of this polarity would be the *less* noise present in an audio signal, the higher the perceived value of danger. The reason we investigated the inverse polarity was to see how a 'badly designed' parameter mapping (in terms of the magnitude estimation method) affects its usability – i.e., how well it can be understood and used in a demanding task context.

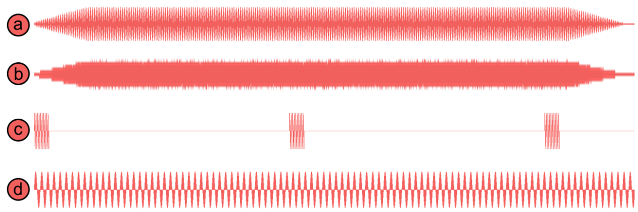

Figure 2: Example waveforms for the roughness (a), noise (b), tempo (c) and duration (d) stimuli, for one stimuli level each. All stimuli can be found the supplementary material.

Roughness and noise were chosen as acoustic parameters, based on results from Ferguson et al. [14], who evaluated the use of these parameters in an auditory display. Roughness is the subjective perception of rapid amplitude modulation, ranging from modulation frequencies of 15 Hz to 300 Hz, with peak perceived roughness around 70 Hz [52]. More commonly, acoustic roughness is related to the sensation of dissonance, or a sound being 'out of tune' [28].

Vibration duration and vibration pulse tempo were chosen as vibrotactile parameters, based on results from another study by Ferguson et al [15]. We adapted their three vibration cues to give four levels for this experiment.

Stimuli were two seconds long. All had an amplitude envelope with a 0.2 second linear onset (attack) and offset (release). All stimuli were created using SuperCollider[1] as follows (Figure **??** shows an example waveform for each condition):

3.5.1 Audio Roughness. Stimuli were 1 kHz sine tones, amplitude modulated at 0, 7, 23 and 70 Hz. These frequencies were chosen based on results from [13] that established ten levels of roughness for use in an auditory display. Here the first, fourth, seventh and tenth levels of that range were used to ensure a maximum difference between cues. It is worth noting again that the roughness levels which have been used in all experiments so far are based on Zwicker & Fastl's work [52] where they established a range of *perceptually equally distributed* levels for roughness. So, even though 0, 7, 23 and 70 Hz are unevenly distributed numerically, they are equally distributed in terms of listener perception of roughness.

3.5.2 Audio Noise. Stimuli consisted of a 1 kHz sine tone and white noise combined to varying degrees. These waveforms were combined as {100% sine, 0% noise} (level 1), {60% sine, 30% noise}, {30% sine, 60% noise}, and {0% sine, 100% noise} (level 4). Like the auditory roughness stimuli, these were extracted from the range of ten evaluated in [13].

3.5.3 Vibration Duration. Stimuli were 200 Hz waveforms that were 100, 500, 1000 and 2000 ms in duration.

3.5.4 Vibration Tempo. Stimuli were 200 Hz waveforms which pulsed at varying intervals. The *on* period for all cues was 50 ms and the *off* periods were: 15, 150, 400 and 800 ms.

[1]SuperCollider: https://supercollider.github.io/

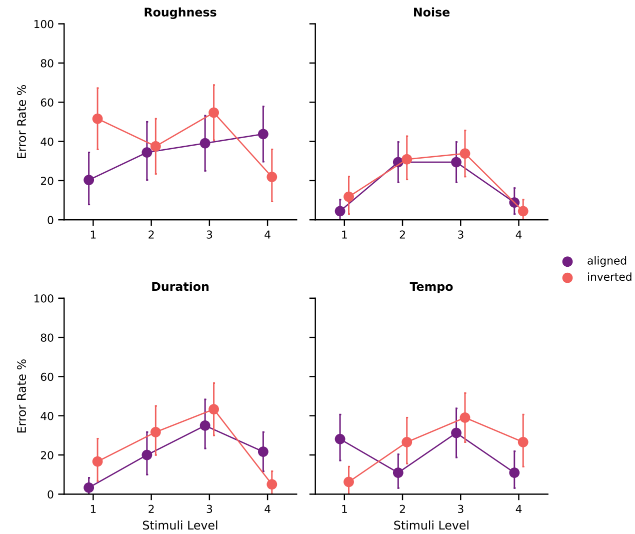

Figure 3: Mean classification error rates.

3.6 Apparatus

Audio stimuli were presented using a pair of wired Beyerdynamic DT-100 headphones. Vibrotactile stimuli were presented using a Tactile Labs Haptuator MkII[2], attached to the wrist (in a similar position to a watch). The Haptuator MkII response time is less than 1 ms [10] so hardware latency is negligible, orders of magnitude lower than our mean response times. The *N*-Back task stimuli were shown on a monitor placed on a table in front of the participant.

3.7 Participants

Sixteen participants took part in the study (8 female, 7 male and 1 non-binary; mean = 28.4 years, SD = 5 years. Fifteen were right-handed and one was left-handed, but all indicated they used their right hand for input, so we presented stimuli to their left hand and they provided input with their right hand. All participants reported no uncorrected vision impairment and no hearing impairment.

4 RESULTS

4.1 Classification Error Rate

Mean classification error rate was 25.3% (SD 35.8%). Figure 3 shows the mean classification error rate for each level of each parameter mapping. Results are aggregated across *N*-back level for the plot. Colour shows polarity and error bars show 95% CIs.

We applied the Aligned-Rank Transform [50] before analysis. We used a repeated-measures ANOVA to investigate the effect of parameter, polarity, and *N*-back level on error rate. Table 1 shows the ANOVA results.

There was a significant effect of parameter on error rate ($p < .001$). Post hoc comparisons of estimated marginal means found higher error rates for roughness than all others ($t \geq 4.40$, $p < .001$). No other comparisons were significantly different ($p \geq .78$).

[2]Haptuator: http://tactilelabs.com/products/haptics/haptuator-mark-ii-v2

Table 1: ANOVA results for error rate.

Effect	F	p
Parameter	12.25	<.001
Polarity	1.78	.18
N-Back	2.89	.09
Parameter:Polarity	4.80	.003
Parameter:N.Back	2.85	.04
Polarity:N.Back	1.33	.25
Parameter:Polarity:N.Back	.46	.71

Table 2: ANOVA results for response time.

Effect	F	p
Parameter	5.11	.002
Polarity	15.85	<.001
N.Back	17.80	<.001
Parameter:Polarity	2.64	.05
Parameter:N.Back	0.79	.50
Polarity:N.Back	0.62	.43
Parameter:Polarity:N.Back	1.37	.25

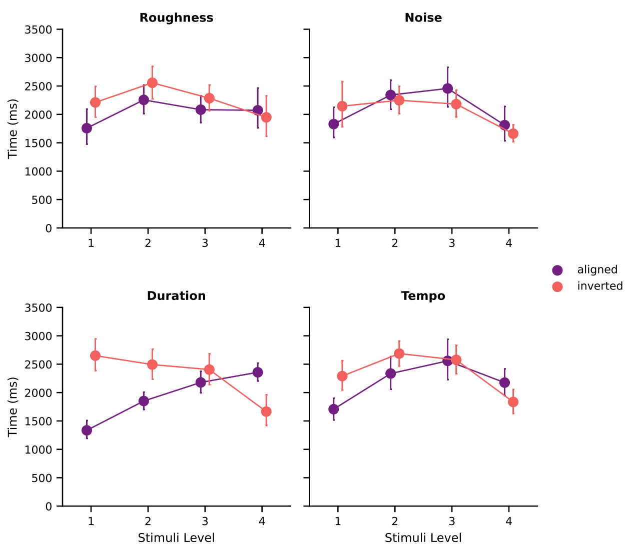

Figure 4: Mean classification response time.

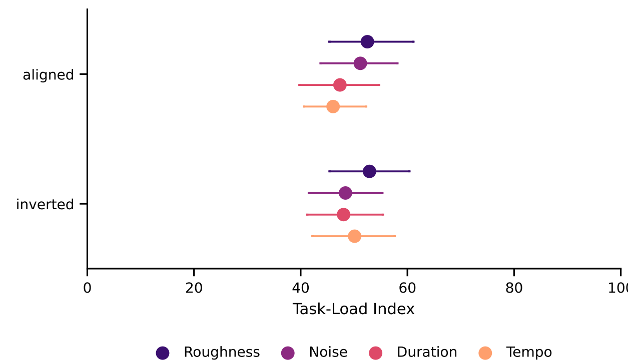

Figure 5: TLX scores for each parameter and polarity.

There was a significant interaction effect between parameter and polarity (p = .003). Post hoc comparisons only found previously known differences between parameters (e.g., roughness vs tempo). There were no significant differences between polarities within individual parameters: roughness (t = 1.48, p = .82), noise (t = 1.05, p = .97), duration (t = .66, p = .99), and tempo (t = .4, p = .99).

There was a significant interaction effect between parameter and *N*-back task level (p = .04). Post hoc comparisons only revealed previously found differences between parameters (e.g., roughness vs noise), with higher error rates in the 2-back task.

4.2 Classification Response Time

Mean response time in the classification task was 2.15s (SD .83s). Figure 4 shows the mean response time for each level of each parameter mapping. Results are aggregated across *N*-back level for the plot. Colour shows polarity and error bars show 95% CIs.

Times were not normally distributed (Shapiro-Wilk W = .92, p <.001) so we applied the aligned-rank transform [50] to the time data. We used a repeated-measures ANOVA to investigate the effect of parameter, polarity, and *N*-back level on the response time time. Table 2 shows the ANOVA results.

There was a significant effect of parameter on time (p = .002). Post hoc comparisons of estimated marginal means found one sig. difference: noise was faster than tempo (2.08s vs 2.27s, t = 3.84, p <.001). No other comparisons were significant (p ≥ .06).

There was a significant effect of polarity on time (p <.001). The post hoc comparison found that aligned stimuli were classified more quickly (2.07s vs 2.24s, t = 3.98, p <.001).

There was a significant effect of *N*-back task level on time (p <.001). The post hoc comparison found lower times for the 0-back task (2.04s vs 2.27s, t = 4.22, p <.001).

4.3 Task-Load Index

Task-Load Index (TLX) survey responses were aggregated into 'raw' task workload scores [23]. Mean task workload score was 49.6 out of 100 (SD 19.3). Figure 5 shows the mean scores for each parameter and polarity. Error bars show 95% CIs.

We applied the aligned-rank transform [50] to workload scores. We then used a repeated-measures ANOVA to investigate the effect of parameter, polarity, and *N*-back level on workload. Table 3 shows the ANOVA results.

There was a significant main effect of *N*-back level on workload (p <.001). A post hoc comparison found lower workload scores for the 0-back task than the 2-back task (41.9 vs 57.3, t = 9.82, p <.001).

5 DISCUSSION

Our aim was to investigate the effect of polarity on the usability of data-sound and data-vibration mappings. It is reasonably assumed that the most 'intuitive' mapping will also be the most usable, but

Table 3: ANOVA results for TLX workload scores.

Effect	F	p
Parameter	2.15	.09
Polarity	.07	.79
N.Back	96.4	<.001
Parameter:Polarity	.72	.54
Parameter:N.Back	.24	.87
Polarity:N.Back	<.001	.98
Parameter:Polarity:N.Back	.70	.55

this was not the case in this experiment. Even under two levels of cognitive demand (induced via the *N*-back task), we did not find much difference between the polarities.

Polarity had a significant effect on response time, which was lower for the aligned mappings. However, there was no difference in error rates between the polarities. There was also no effect of polarity on perceived task workload, which was unexpected. Our aligned mappings were taken from prior magnitude estimation studies that identified these as the most 'natural' polarities for these parameters based on user consensus. We expected the aligned polarity to achieve lower error rates and lower cognitive workload, since the inverted polarities were supposedly less intuitive. This was not the case, however.

Mapping parameter had a significant effect on response accuracy and time. Roughness had the highest error rate (37.9%) and unlike the other mappings, this was fairly constant across all stimuli levels. In contrast, noise and duration had n-shaped curves, showing better performance at the lower and highest magnitude values (i.e., stimuli levels 1 and 4 in Figure 3). This suggests that intermediate values (i.e., levels 2 and 3) were more ambiguous in these conditions, but the extreme values were easily identified (e.g., pure tones and pure white noise for the noise parameter).

Response times also had slightly n-shaped curves for the roughness, noise and tempo parameters, again suggesting that it was easier to interpret the lowest and highest magnitude values. Note that the {duration × aligned} and {duration × inverted} curves change in opposite directions; this reflects the impact of stimuli duration on response time, since users can respond more quickly after the shortest vibrations (aligned level 1, inverted level 4), etc.

Based on our findings, we recommend that designers evaluate their cues under induced cognitive load (e.g., via the *N*-back task or another ecologically valid task), as this can affect usability. We found no evidence that polarity affects usability, although parameter choice itself might. From our four parameters, auditory noise and vibrotactile tempo are most promising, with especially fast and accurate response to the lowest and highest magnitude values.

We used the *N*-back task protocol to induce cognitive load, since data-sound and data-vibration mappings like these are often used in cognitively demanding scenarios, e.g., driving [35, 36, 41], air traffic control [1, 8, 9, 37], and in hospitals [3, 16, 31, 38]. The *N*-back task level had an effect on TLX scores and response time; both values were higher in the 2-back task than the 0-back task. Higher TLX scores show the *N*-back task working as intended. The higher response time shows the importance of evaluating data-sound and data-vibration mappings in more ecologically valid circumstances, under cognitive load, as this may impact cue usability, since users need to divide attention between two tasks.

5.1 Design Implications

The findings from this research have implications relevant to both methodological and design:

For methodology, we questioned the use of consensus-based methods for establishing parameter mappings and found that whilst the so-called 'intuitive' mapping has faster reaction times, there was little effect on cognitive load or response accuracy. We also found that induced cognitive load can influence perception. These findings imply that time and/or resources would be better spent evaluating new audio/vibrotactile cues in cognitively demanding contexts, rather than conducting magnitude estimation studies that establish mappings via end users instead of designers.

For design, our results can be used to inform the selection of audio/vibrotactile mappings, especially in safety or reaction critical usage scenarios. Our error rate and time results can inform the choice of both parameter selection (e.g., auditory roughness vs auditory noise?) and polarity selection (e.g., aligned vs inverted?). Whilst our aim was not to establish a 'best' mapping, our results show the promising performance of the auditory noise mapping, regardless of polarity.

6 CONCLUSION

When mapping information to audio or vibration, designers need to identify an appropriate scale (the rate of change) and polarity (the direction of change) between a design parameter and data range. Consensus-based methods like magnitude estimation can inform design, on the assumption that the most 'intuitive' mapping will also be the most usable. There is a cost associated with running consensus-based studies like magnitude estimation, so our work investigated if such studies actually do result in more effective and usable mappings, especially in mentally demanding scenarios.

We selected four mappings from prior magnitude estimation studies, evaluating their original and opposite polarities (i.e., aligned and inverted). We induced cognitive load via the *N*-back task, to evaluate cue usability in a demanding scenario. Our results suggest that the most intuitive polarity (i.e., aligned with consensus) can lead to faster classification, although error rates and workload were not significantly different. This was unexpected as we anticipated best performance from the aligned mapping. An implication of this is that choosing polarity through consensus may not necessarily yield a more usable polarity mapping.

Our results also characterise the effects of cognitive load on the usability of non-visual information mappings. Designers need to consider cognitive load when creating and evaluating mappings for challenging usage contexts. Induced cognitive load is necessary for reliable insight into cue usability and performance and this can be achieved, e.g., through ecologically valid tasks or standard protocols like the *N*-back task.

REFERENCES

[1] Pietro Aricò, Maxime Reynal, Jean-Paul Imbert, Christophe Hurter, Gianluca Borghini, Gianluca Di Flumeri, Nicolina Sciaraffa, Antonio Di Florio, Michela Terenzi, Ana Ferreira, Simone Pozzi, Viviana Betti, Matteo Marucci, Enea Pavone, Alexandru C. Telea, and Fabio Babiloni. 2018. Human-Machine Interaction Assessment by Neurophysiological Measures: A Study on Professional Air Traffic Controllers. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 4619–4622. https://doi.org/10.1109/EMBC.2018.8513212

[2] Louise Axon, Michael Goldsmith, and Sadie Creese. 2018. Sonification Mappings: Estimating Effectiveness, Polarities and Scaling in an Online Experiment. *Journal of the Audio Engineering Society* 66, 12 (2018), 1016–1032.

[3] David Black, Christian Hansen, Arya Nabavi, Ron Kikinis, and Horst Hahn. 2017. A survey of auditory display in image-guided interventions. *International Journal of Computer Assisted Radiology and Surgery* 12, 10 (2017), 1665–1676.

[4] Stephen Brewster. 2002. Overcoming the lack of screen space on mobile computers. *Personal and Ubiquitous Computing* 6, 3 (2002), 188–205.

[5] Stephen Brewster and Lorna M Brown. 2004. Tactons: structured tactile messages for non-visual information display. In *Proceedings of the fifth conference on Australasian user interface-Volume 28 – OzCHI*. Australian Computer Society, Inc., 15–23.

[6] Lorna M. Brown, Stephen A. Brewster, and Helen C. Purchase. 2005. A first investigation into the effectiveness of Tactons. In *First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics Conference*. 167–176. https://doi.org/10.1109/WHC.2005.6

[7] 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 – Mobile HCI*. ACM, 231–238.

[8] Densil Cabrera and Sam Ferguson. 2006. Considerations arising from the development of auditory alerts for air traffic control consoles. In *Proceedings of the 12th International Conference on Auditory Display – ICAD*.

[9] Densil Cabrera, Sam Ferguson, and Gary Laing. 2005. Development of Auditory Alerts for Air Traffic Control Consoles. In *Audio Engineering Society Convention 119*. http://www.aes.org/e-lib/browse.cfm?elib=13300

[10] Landry Chapwouo Tchakouté, Louis Tremblay, and Bob-Antoine Menelas. 2018. Response Time to a Vibrotactile Stimulus Presented on the Foot at Rest and During Walking on Different Surfaces. *Sensors* 18, 7 (Jun 2018), 2088. https://doi.org/10.3390/s18072088

[11] Mario Enriquez, Karon MacLean, and Christian Chita. 2006. Haptic phonemes: basic building blocks of haptic communication. In *Proceedings of the 8th International Conference on Multimodal Interfaces – ICMI*. ACM, 302–309.

[12] Jamie Ferguson and Stephen Brewster. 2019. Evaluating the magnitude estimation approach for designing sonification mapping topologies. In *Proceedings of the International Conference on Auditory Displays – ICAD*. http://hdl.handle.net/1853/61521

[13] Jamie Ferguson and Stephen A Brewster. 2017. Evaluation of psychoacoustic sound parameters for sonification. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction – ICMI*. ACM, 120–127.

[14] Jamie Ferguson and Stephen A Brewster. 2018. Investigating Perceptual Congruence Between Data and Display Dimensions in Sonification. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems – CHI*. ACM, 611.

[15] Jamie Ferguson, John Williamson, and Stephen Brewster. 2018. Evaluating mapping designs for conveying data through tactons. In *Proceedings of the 10th Nordic Conference on Human-Computer Interaction*. ACM, 215–223.

[16] Loris Fichera, Claudio Pacchierotti, Emidio Olivieri, Domenico Prattichizzo, and Leonardo S Mattos. 2016. Kinesthetic and vibrotactile haptic feedback improves the performance of laser microsurgery. In *Proceedings of the 2016 IEEE Haptics Symposium – HAPTICS*. IEEE, 59–64.

[17] Euan Freeman, Stephen Brewster, and Vuokko Lantz. 2014. Tactile Feedback for Above-Device Gesture Interfaces: Adding Touch to Touchless Interactions. In *Proceedings of the 16th International Conference on Multimodal Interaction* (Istanbul, Turkey) *(ICMI '14)*. Association for Computing Machinery, New York, NY, USA, 419–426. https://doi.org/10.1145/2663204.2663280

[18] Euan Freeman, Stephen Brewster, and Vuokko Lantz. 2016. Do that, there: an interaction technique for addressing in-air gesture systems. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems – CHI*. ACM, 2319–2331.

[19] Euan Freeman and Graham Wilson. 2021. Perception of Ultrasound Haptic Focal Point Motion. In *Proceedings of the 23rd ACM International Conference on Multimodal Interaction (ICMI '21)*. ACM. https://doi.org/10.1145/3462244.3479950

[20] Euan Freeman, Graham Wilson, Dong-Bach Vo, Alex Ng, Ioannis Politis, and Stephen Brewster. 2017. *Multimodal Feedback in HCI: Haptics, Non-Speech Audio, and Their Applications*. Association for Computing Machinery and Morgan and Claypool, 277–317. https://doi.org/10.1145/3015783.3015792

[21] Nicholas A Giudice, Hari Prasath Palani, Eric Brenner, and Kevin M Kramer. 2012. Learning non-visual graphical information using a touch-based vibro-audio interface. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, 103–110.

[22] David Grimes, Desney S Tan, Scott E Hudson, Pradeep Shenoy, and Rajesh PN Rao. 2008. Feasibility and pragmatics of classifying working memory load with an electroencephalograph. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems – CHI*. ACM, 835–844.

[23] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in Psychology*. Vol. 52. Elsevier, 139–183.

[24] Eve Hoggan and Stephen Brewster. 2007. New Parameters for Tacton Design. In *CHI '07 Extended Abstracts on Human Factors in Computing Systems* (San Jose, CA, USA) *(CHI EA '07)*. Association for Computing Machinery, New York, NY, USA, 2417–2422. https://doi.org/10.1145/1240866.1241017

[25] Eve Hoggan, Roope Raisamo, and Stephen A Brewster. 2009. Mapping information to audio and tactile icons. In *Proceedings of the 2009 International Conference on Multimodal Interfaces – ICMI*. ACM, 327–334.

[26] Simon Holland, David R Morse, and Henrik Gedenryd. 2002. AudioGPS: Spatial audio navigation with a minimal attention interface. *Personal and Ubiquitous Computing* 6, 4 (2002), 253–259.

[27] Yu Huang and Weikang Jiang. 2017. The effect of exposure duration on the subjective discomfort of aircraft cabin noise. *Ergonomics* 60, 1 (2017), 18–25.

[28] Phil N Johnson-Laird, Olivia E Kang, and Yuan Chang Leong. 2012. On musical dissonance. *Music Perception: An Interdisciplinary Journal* 30, 1 (2012), 19–35.

[29] Alan Kan, Ruth Y Litovsky, and Matthew J Goupell. 2015. Effects of interaural pitch-matching and auditory image centering on binaural sensitivity in cochlear-implant users. *Ear and hearing* 36, 3 (2015), e62.

[30] Thomas Kosch, Markus Funk, Albrecht Schmidt, and Lewis L Chuang. 2018. Identifying Cognitive Assistance with Mobile Electroencephalography: A Case Study with In-Situ Projections for Manual Assembly. *Proceedings of the ACM on Human-Computer Interaction* 2, EICS (2018), 11.

[31] Robert G Loeb, Birgit Brecknell, and Penelope M Sanderson. 2016. The sounds of desaturation: a survey of commercial pulse oximeter sonifications. *Anesthesia & Analgesia* 122, 5 (2016), 1395–1403.

[32] Bruce Mehler, Bryan Reimer, and Jeffery A Dusek. 2011. MIT AgeLab Delayed Digit Recall Task (N-Back). *Cambridge, MA: Massachusetts Institute of Technology* (2011).

[33] John G Neuhoff. 2019. Is sonification doomed to fail?. In *Proceedings of the International Conference on Auditory Displays – ICAD*. Georgia Institute of Technology.

[34] Adrian M Owen, Kathryn M McMillan, Angela R Laird, and Ed Bullmore. 2005. N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping* 25, 1 (2005), 46–59.

[35] Ioannis Politis, Stephen Brewster, and Frank Pollick. 2017. Using multimodal displays to signify critical handovers of control to distracted autonomous car drivers. *International Journal of Mobile Human Computer Interaction (IJMHCI)* 9, 3 (2017), 1–16.

[36] Ioannis Politis, Stephen A. Brewster, and Frank Pollick. 2014. Evaluating Multimodal Driver Displays under Varying Situational Urgency. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 4067–4076. https://doi.org/10.1145/2556288.2556988

[37] Maxime Reynal, Jean-Paul Imbert, Pietro Aricò, Jérôme Toupillier, Gianluca Borghini, and Christophe Hurter. 2019. Audio Focus: Interactive spatial sound coupled with haptics to improve sound source location in poor visibility. *International Journal of Human-Computer Studies* 129 (2019), 116–128.

[38] Sebastian Schwarz and Tim Ziemer. 2019. A psychoacoustic sound design for pulse oximetry. In *Proceedings of the International Conference on Auditory Displays – ICAD*.

[39] Hasti Seifi and Karon E Maclean. 2013. A first look at individuals' affective ratings of vibrations. In *Proceedings of the IEEE World Haptics Conference – WHC*. IEEE, 605–610.

[40] Robert D Sorkin. 1988. Why are people turning off our alarms? *The Journal of the Acoustical Society of America* 84, 3 (1988), 1107–1108.

[41] Charles Spence and Cristy Ho. 2008. Multisensory warning signals for event perception and safe driving. *Theoretical Issues in Ergonomics Science* 9, 6 (2008), 523–554. https://doi.org/10.1080/14639220701816765

[42] Delphine Szymczak, Kirsten Rassmus-Gröhn, Charlotte Magnusson, and Per-Olof Hedvall. 2012. A Real-World Study of an Audio-Tactile Tourist Guide. In *Proceedings of the 14th International Conference on Human-Computer Interaction with Mobile Devices and Services* (San Francisco, California, USA) *(MobileHCI '12)*. Association for Computing Machinery, New York, NY, USA, 335–344. https://doi.org/10.1145/2371574.2371627

[43] Robert Teghtsoonian, SS Stevens, and Geraldine Stevens. 1975. Psychophysics: Introduction to Its Perceptual, Neural, and Social Prospects. *The American Journal of Psychology* 88, 4 (1975), 677.

[44] Bruce N. Walker. 2002. Magnitude Estimation of Conceptual Data Dimensions for Use in Sonification. *Journal of Experimental Psychology: Applied* 8, 4 (2002), 211–221.

[45] Bruce N Walker. 2007. Consistency of magnitude estimations with conceptual data dimensions used for sonification. *Applied Cognitive Psychology* 21, 5 (2007), 579–599.

[46] Bruce N Walker and Gregory Kramer. 1996. Mappings and Metaphors in Auditory Displays: An Experimental Assessment. In *Proceedings of the International Conference on Auditory Display – ICAD*. Georgia Institute of Technology.

[47] Bruce N Walker and Gregory Kramer. 2005. Mappings and metaphors in auditory displays: An experimental assessment. *ACM Transactions on Applied Perception (TAP)* 2, 4 (2005), 407–412.

[48] Graham Wilson and Stephen A. Brewster. 2016. Using Dynamic Audio Feedback to Support Peripersonal Reaching in Young Visually Impaired People. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility* (Reno, Nevada, USA) *(ASSETS '16)*. Association for Computing Machinery, New York, NY, USA, 209–218. https://doi.org/10.1145/2982142.2982160

[49] Graham Wilson and Stephen A Brewster. 2017. Multi-Moji: Combining Thermal, Vibrotactile & Visual Stimuli to Expand the Affective Range of Feedback. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems – CHI*. ACM, 1743–1755.

[50] Jacob O Wobbrock, Leah Findlater, Darren Gergle, and James J Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems – CHI*. ACM, 143–146.

[51] Yongjae Yoo, Taekbeom Yoo, Jihyun Kong, and Seungmoon Choi. 2015. Emotional responses of tactile icons: Effects of amplitude, frequency, duration, and envelope. In *Proceedings of the 2015 IEEE World Haptics Conference – WHC*. IEEE, 235–240.

[52] Eberhard Zwicker and Hugo Fastl. 2013. *Psychoacoustics: Facts and models*. Vol. 22. Springer Science & Business Media.