Understanding Ambiguous Effects of Multi-Modal Cueing on Task Performance Using Bayesian Sensory Integration

Tobias Michael BENZ, Verena NITSCH

Human Factors Institute, University of the Bundeswehr Munich
Werner-Heisenberg-Weg 39, D-85579 Neubiberg

Abstract: It has been suggested that multi-modal cues may improve task performance in teleoperation applications. Nevertheless, the added value of multi-modal feedback is not yet clear. The present research investigates the value of Bayesian Sensory Integration (BSI) in predicting the effects of multi-modal cues on localization performance. We use secondary data (n = 48) from an experiment in which participants adjusted uni-, bi- and tri-modal cue intensities and indicated the direction of the cue origin. Results show a significantly decreased variance in localization performance for combinations of haptic (focal) with visual and auditory (focal) with visual cues. Overall, the data suggest that BSI can provide insight into the differential effects of multi-modal signals on localization performance in a teleoperation setting.

Keywords: teleoperation, multimodal, integration, haptic, auditory, visual

1. Introduction

Task performance (TP) in teleoperated applications is frequently inferior to performance in non-mediated applications. Oftentimes, this shortcoming is addressed by using multi-modal feedback cues. Nevertheless, the added value of multi-modal feedback is not yet clear (Hancock et al., 2015; Vitense, Jacko, & Emery, 2003). Currently, two underlying concepts of multi-modal feedback in teleoperation dominate: (1) redundant information representation enhances task performance (Lewkowicz & Kraebel, 2004) (2) multi-modal feedback increases immersion which increases task performance (Minsky, 1980; Pongrac, 2011). Both concepts fail to explain unchanging or decreasing task performance with multi-modal feedback cues. A critical step in the processing of multi-modal cueing may be the sensory integration. Utilizing the well-established Bayesian Sensory Integration (BSI) model (Ernst & Bülthoff, 2004) may help to better explain and understand the involved processes. Specifically, BSI would explain why multi-modal cues can improve task performance in some circumstances but not in others. Moreover, BSI can account for the effects of other variables on task performance, such as time delay. As of yet, process models of multi-modal cueing effects that take early cue processing into account, are exceedingly rare.

2. Literature Review

So far, it is unclear which perception and/or cognitive mechanisms might account for the effects of multi-modal interfaces in increasing teleoperation performance, and under which circumstances multi-modal feedback or cueing might be helpful at all

(Hancock et al., 2015; Vitense et al., 2003). In fact, research with both humans and animals has identified many cases of cross-modal interaction in which the interpretation of data in one sensory modality is influenced by the data that are available in another sensory modality (Bertelson & Gelder, 2004). Hence, the available literature suggests that task performance is influenced by the perception and accuracy of information from the environment (presented in different modalities), the process of which may be described by sensory integration.

BSI is a commonly used theorem of multi-sensory perception that assumes integration occurs on the basis of reliability (= inverse variance) estimates (Ernst, Rohde, & van Dam, 2016). According to this model, multi-sensory perception does not necessarily benefit from increased information density of added modalities or a winner-takes-all rule (i.e. the “best” sensory input is perceived, others neglected). Several sources in the literature describe the merging of senses in a statistically optimum fashion according to Bayes' Rule, with the goal to increase the reliability of estimates (Ernst et al., 2016; Pouget, Beck, Ma, & Latham, 2013). The Bayes' Rules state the probability of an estimated position S given the input of the modalities is the product of the modalities’ likelihood to occur for that position multiplied by the prior probability that these modalities occur in that combination for this position. If the sensory integration is described in a bottom-up process the BSI Model reduces to the Maximum Likelihood Estimate (MLE) but also assumes a complete fusion (Ernst & Bülthoff, 2004). Eqn. 1 shows the MLE and that the integrated estimate is the weighted sum of the individual estimates. These weights (w) are proportional to their inverse variance (= reliability) (Eqn. 2) (Ernst & Bülthoff, 2004) and sum up to 1 for complete fusion (Bresciani, Dammeier, & Ernst, 2006). The index refers to each sensory signal.

$$\hat{s} = \sum_i w_i \hat{s}_i \ \ with \sum_i w_i = 1 \qquad \text{Equation 1}$$

$$w_j = \frac{1/\sigma_j^2}{\sum_{i=1\ldots j\ldots N} 1/\sigma_i^2} \qquad \text{Equation 2}$$

The described combining of redundant sensory information in a statistically optimum way leads to a minimization of the estimated noise with an optimum bi-modal variance (Ernst et al., 2016). Combining e.g. haptic and auditory cues would result in a statistically optimum estimate of the position, when taking the respective uni-modal variance and estimates into account. By reducing variance, this would lead to more reliable information source for the teleoperation. Even though the variance is minimized through integration, the estimation error may be increased if the more reliable (i.e. higher weighted) cue is error prone and is combined with a less reliable (i.e. lower weighted) accurate estimate. As Eqn. 2 indicates, combining sensory information with a high ratio of variance does not improve perception of the more precise sensory cue to a greater extent. This might explain, why we often find a visual dominance compared to haptic or auditory cues in teleoperation. It follows that the added value of multi-modal feedback is highly dependent on the uni-modal sensory variance and therefore signal reliability, which in turn strongly depends on the transmitted information. This is supported by results found in (Ernst, Banks, & Bulthoff, 2000), which indicate that if the variance / noise of a visual estimate is increased, the weight of the haptic estimate increases accordingly. Overall, BSI seems to offer a plausible model of human sensory perception and may help to understand and predict performance differences with different multi-modal feedback cues. It can be used in different scenarios to understand the impact of time delay (Rohde & Ernst, 2016) the use of natural and non-natural cue combinations (Ernst, 2007) and even the weighting of modalities according to attentional mechanisms or task demands (Quak, London, & Talsma, 2015; Roach, Heron, & McGraw, 2006).

As of yet, research on the influence of multi-sensory cue processing on task performance in teleoperation is virtually non-existent., Hence, data from a teleoperation experiment, which investigated the effects of multi-modal cues on direction localization, were examined for evidence of BSI. Specifically, it was investigated whether bi-modal representations would effect a variance reduction in performance compared to uni-modal representations, as that would be a strong indicator, of sensory integration according to BSI (Ernst et al., 2016).

3. Method

The analyzed data were acquired originally for the analysis of localization accuracy and intensity adjustments of uni- as well as multi-modal cues. For a detailed description of the data collection and results see (Benz & Nitsch, 2017).

An opportunity sample of 48 participants took part in the experiment (n = 11 female). The mean age is 28.4 yrs. (SD = 9.56 yrs.). All participants had normal or corrected-to-normal vision and no hearing impairment.

Three different directional cues that are supposed to inform the operator of a target location, are presented in the visual, auditory and haptic modalities. A visual directional cue is presented in form of an arrow in the middle of screen that is pointing towards the target. The auditory cue is a beep tone emitted from the target presented to five loudspeakers (Edifier C6XD). The haptic cue is provided with a Microsoft Sidewinder Force Feedback 2 fixed on the table and portrays an attraction force into a given direction. The overall teleoperation setup is simulated using V-Rep. The target is located outside of vision and the robot rotates with a constant speed of 4 RPM.

All participants received standardized instructions for the experiment, in which they were asked to adjust the intensities of the respective uni-modal stimuli until they perceived it to be most informative with respect to target location. After the uni-modal stimulus presentations, the other two stimuli would be added, first on their own (bi-modal presentation) or combined (tri-modal presentation). This results in a 2 (adjustment direction) x 12 (modalities) within-subject design. In the multi-modal conditions, the intensities of the respective stimuli were to be adjusted, until they matched in intensity to the reference stimulus, that had been presented uni-modally. The cue that is adjusted is henceforth referred to as *focal cue*.

When participants were satisfied with each adjustment, the respective trial would end and participants were to indicate via a point-and-click interface the direction of the target. Each adjustment and localization was made twice with a different start intensity (0 % vs. 100 %). The start intensity was randomized; the order of stimuli was semi-randomized with the uni-modal presentation of the reference stimulus always starting first. The reference modality is varied systematically, so that each combination is presented to participants. After finishing all trials, participants were compensated with 10 EUR for their effort. The experiment took each participant app. 1 hour to complete.

4. Results

Only uni- and bi-modal conditions are analyzed as in tri-modal conditions, the focused modality is not entirely clear. Previous research shows that integrated cues result in a smaller variability than their uni-modal / non-integrated counterparts (Bresciani & Ernst, 2007; Drewing & Ernst, 2006). Additionally, the more reliable cue is less susceptible to bias if it is presented as focal cue and has a stronger influence if presented as background cue (Bresciani & Ernst, 2007). Consequently, according to the BSI concept, a reduction in variability and the influence of the focal stimulus should also be observable in the obtained data.

To test this, mean values and standard deviations of the angle error of each condition were calculated for the repeated measures of the same cue. The variance of these within-subject repetitions does not vary significantly ($t(551) = -0.04$, $p = 0.97$), thus indicating that both trials were not influenced by external noise. First, the variability of estimates for uni-modal cues was compared. Figure 1 (left) presents the frequencies of the angle error for uni-modal cues. The more spread out the values are, the higher is the variability. The distributions of the responses show a large difference, with the lowest variability for visual cues and the highest variability for auditory cues. We modelled the angle error with mixed-effect models including random intercepts only. The model including uni-modal cues confirms that subjects were significantly more variable in estimating the location based on auditory ($t(90) = 4.17$, $p < .001$) compared to visual direction cues. However, variability did not differ between position estimates in the haptic and visual conditions ($t(90) = 1.86$, $p = 0.7$). Additionally, post-hoc tests using Games-Howell correction did not show a difference between auditory and haptic localization variability ($t(83.81) = -8.79$, $p = 0.15$). The visual-haptic variance ratio (1.42), the haptic-auditory variance ratio (1.32), and the visual-auditory ratio (1.88) are within the expected and ideal range of [0.5, 2] (Ernst et al., 2016).

Provided that the variance ratio of all cue combinations is within the predicted range, integration should take place in all conditions leading to a decreased variance. To test this, the standard deviation of each multi-modal condition is compared to the individual standard deviations of uni-modal cues. Figure 1 (right) shows the expected pattern of a reduced variability in combined conditions. The only exception is the combination of auditory and haptic cues. The differences of adding a visual to the focal haptic cue ($t(90) = -2.72$, $p < .01$) and the auditory cue ($t(90) = -3.58$, $p < .000$) are significant using mixed-effect models. Furthermore, the reduction depends on the

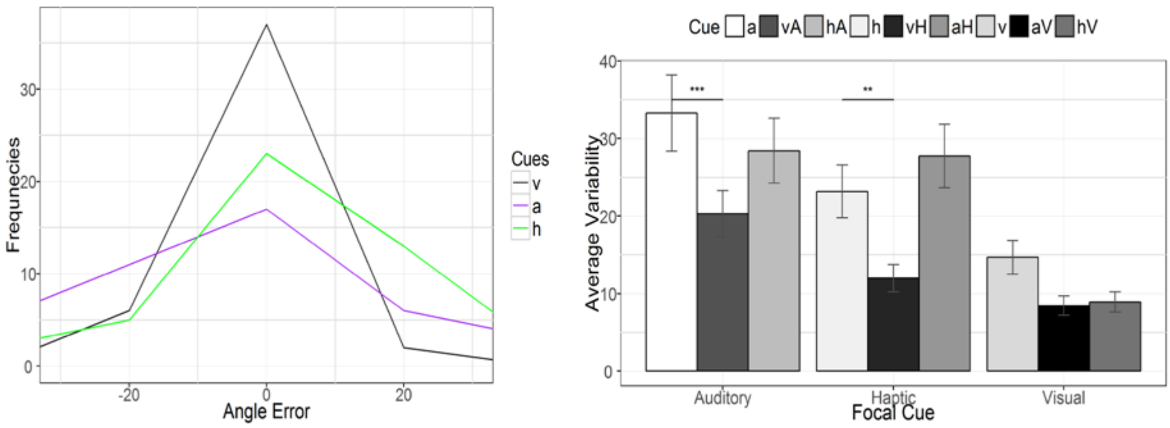

Figure 1. *(left) frequencies of the angle error showing the different variabilities for uni-modal cues; (right) average variability of the individual estimates for all cues; capital letters indicate the focal cue.*

variability of the focal stimulus as, for example, adding visual to the focal auditory cue (vA) shows a smaller decrease than adding auditory to the focal visual cue (aV). The increasing variability of a focal haptic cue with an auditory (aH) cue seems to be surprising. Overall, however, the reduction of variability decreases if the more variable cue is used as focal cue, which is in line with previous research (Bresciani et al., 2006).

5. Discussion and Conclusion

Employing secondary data in which three modalities were used to convey direction, it was found that: (1) BSI is applicable in the context of teleoperation which could be used to explain behavior that cannot be explained by sheer redundancy or immersion, (2) variability is significantly lower when the cues are presented in two modalities, (3) the reduction varies depending on which cue – focal or background – is less variable.

The results suggest that BSI rather than a winner-take-all mechanism takes place in teleoperated scenarios, as the latter approach would predict that the more appropriate cue is used and therefore the variability would not change. In contrast, the BSI model predicts a mutual influence of both integrated cues, with the less variable (= more reliable) cue having a stronger effect. This effect is diminished depending on the focal cue, with the effect being largest if the focal cue is also the less variable / more reliable. The data confirm this pattern as the difference in variability is significantly lower when adding visual to focal auditory and focal haptic cues compared to the uni-modal cue. In the analyzed data, the least variable cue is the visual cue and the reduction is smallest in combination with the auditory cue, which is the most variable. Furthermore, the influence on the focal visual cue is smaller compared to the other focal cues.

Overall, visual cues seem to be integrated quite well. Inversely, variability of haptic and auditory cue combinations did not change or even increase compared to uni-modal haptic cues. An explanation can be the lack of a prior that relates auditory and haptic cues for direction estimates and, therefore, no integration would take place. Consequently, a winner-take-all mechanism in favor of auditory cues, that are typically used for direction estimates in real life, took place. However, this would also apply for haptic-visual combinations. To provide further explanations, it should be investigated whether participants tend to switch their attentional focus between the focal cue and background cue. In this case, a participant would randomly (unconsciously) select from one of the two distributions, which would result in a higher variability than the single cues (Bresciani et al., 2006; Bresciani & Ernst, 2007). The reason for this switching behavior could be the high variability within both cues, thus resulting in higher uncertainty of the estimate.

Overall, BSI seems to be applicable and helps to understand different performance levels with multi-modal feedback in teleoperation. Hence, BSI has the potential to inform the evaluation of HMI with multi-modal feedback according to the added value of modalities. Future studies are planned which include more repetitions for each modality and a manipulation of integration quality to evaluate the applicability of BSI in predicting task performance and situation awareness measures in teleoperation applications.

6. References

Benz, T. M., & Nitsch, V. (2017). Using Multisensory Cues for Direction Information in Teleoperation: More Is Not Always Better. In *Proc. IEEE International Conference on Robotics and Automation.* Singapore.

Bertelson, P., & Gelder, B. de. (2004). The psychology of multimodal perception. *Crossmodal Space and Crossmodal Attention*, 141–177.

Bresciani, J.-P., Dammeier, F., & Ernst, M. O. (2006). Vision and touch are automatically integrated for the perception of sequences of events. *Journal of Vision*, *6*(5), 554–564. https://doi.org/10.1167/6.5.2

Bresciani, J.-P., & Ernst, M. O. (2007). Signal reliability modulates auditory-tactile integration for event counting. *Neuroreport*, *18*(11), 1157–1161. https://doi.org/10.1097/WNR.0b013e3281ace0ca

Drewing, K., & Ernst, M. O. (2006). Integration of force and position cues for shape perception through active touch. *Brain Research*, *1078*(1), 92–100. https://doi.org/10.1016/j.brainres.2005.12.026

Ernst, M. O., Banks, M. S., & Bulthoff, H. H. (2000). Touch can change visual slant perception. *Nature Neuroscience*, *3*(1), 69–73. https://doi.org/10.1038/71140

Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision*, *7*(5), 7.1-14. https://doi.org/10.1167/7.5.7

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*(4), 162–169. https://doi.org/10.1016/j.tics.2004.02.002

Ernst, M. O., Rohde, M., & van Dam, L. C. J. (2016). Statistically Optimal Multisensory Cue Integration: A Practical Tutorial. *Multisensory Research*, *29*(4-5), 279–317. https://doi.org/10.1163/22134808-00002510

Hancock, P. A., Lawson, B., Cholewiak, R., Elliott, L. R., van Erp, J. B. F., Mortimer, B. J. P., Redden, E. S. (2015). Tactile Cuing to Augment Multisensory Human-Machine Interaction. *Ergonomics in Design: The Quarterly of Human Factors Applications*, *23*(2), 4–9. https://doi.org/10.1177/1064804615572623

Lewkowicz, D. J., & Kraebel, K. S. (2004). The Value of Multisensory Redundancy in the Development of Intersensory Perception. In B. E. Stein, C. Spence, & G. Calvert (Eds.), *The Handbook of Multisensory Processes* (pp. 655–678). Cambridge, Mass: A Bradford Book.

Minsky, K. P. (1980). Telepresence. *Omni Magazine*, 40–52.

Pongrac, H. (2011). *Gestaltung und Evaluation von virtuellen und Telepräsenzsystemen: Aufgabenleistung und Präsenzempfinden in Abhängigkeit von haptischen und visuellen Ein- und Ausgabeparametern* (neue Ausg). Saarbrücken: Südwestdeutscher Verlag für Hochschulschriften.

Pouget, A., Beck, J. M., Ma, W. J., & Latham, P. E. (2013). Probabilistic brains: knowns and unknowns. *Nature Neuroscience*, *16*(9), 1170–1178. https://doi.org/10.1038/nn.3495

Quak, M., London, R. E., & Talsma, D. (2015). A multisensory perspective of working memory. *Frontiers in Human Neuroscience*, *9*(197). https://doi.org/10.3389/fnhum.2015.00197

Roach, N. W., Heron, J., & McGraw, P. V. (2006). Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. *Proceedings. Biological Sciences*, *273*(1598), 2159–2168. https://doi.org/10.1098/rspb.2006.3578

Rohde, M., & Ernst, M. O. (2016). Time, agency, and sensory feedback delays during action. *Current Opinion in Behavioral Sciences*, *8*, 193–199. https://doi.org/10.1016/j.cobeha.2016.02.029

Vitense, H. S., Jacko, J. A., & Emery, V. K. (2003). Multimodal feedback: an assessment of performance and mental workload. *Ergonomics*, *46*(1-3), 68–87. https://doi.org/10.1080/00140130303534

Acknowledgement: This work was supported by the German Research Foundation (DFG).

Gesellschaft für
Arbeitswissenschaft e.V.

ARBEIT(S).WISSEN.SCHAF(F)T

Grundlage für Management & Kompetenzentwicklung

64. Kongress der
Gesellschaft für Arbeitswissenschaft

FOM Hochschule für
Oekonomie & Management gGmbH

21. – 23. Februar 2018

GfA Press

Bericht zum 64. Arbeitswissenschaftlichen Kongress vom 21. – 23. Februar 2018

FOM Hochschule für Oekonomie & Management

Herausgegeben von der Gesellschaft für Arbeitswissenschaft e.V.
Dortmund: GfA-Press, 2018
ISBN 978-3-936804-24-9

NE: Gesellschaft für Arbeitswissenschaft: Jahresdokumentation

Als Manuskript zusammengestellt. Diese Jahresdokumentation ist nur in der Geschäftsstelle erhältlich.

Schriftleitung: Matthias Jäger
im Auftrag der Gesellschaft für Arbeitswissenschaft e.V.

USB-Print:
Prof. Dr. Thomas Heupel, FOM Prorektor Forschung, thomas.heupel@fom.de

Screen design und Umsetzung
© 2018 fröse multimedia, Frank Fröse
office@internetkundenservice.de · www.internetkundenservice.de