Selective enhancement of object representations through multisensory integration

David A. Tovar (david.tovar@vanderbilt.edu)
Vanderbilt Brain Institute, 7110 MRB III
Nashville, Tennessee 37240, United States

Micah M. Murray (micah.murray@chuv.ch)
University of Lausanne Radiology Research Center, RC7, 04/021
Rue du Bugnon 46, 1011 Lausanne, Switzerland

Mark T. Wallace (mark.wallace@vanderbilt.edu)
Vanderbilt Brain Institute, 7110 MRB III
Nashville, Tennessee 37240, United States

Abstract:

Objects are the fundamental building blocks of how we represent the external world. These objects come in a variety of forms, with one major distinction being between those that are animate versus inanimate. Many objects are specified in a multisensory manner, yet the nature by which multisensory objects are represented by the brain, particular those that are animate versus not, remains poorly understood. Using representational similarity analysis of human EEG signals, we show that the often-found advantages for the processing of animate objects are no longer evident when they are presented in a multisensory context. Neural decoding was found to be enhanced asymmetrically for inanimate objects, which were more weakly decoded under unisensory conditions. A distance-to-bound analysis provided critical links between neural decoding and behavior. Improved neural decoding for visual and audiovisual objects was associated with faster behavior, and decoding differences between visual and audiovisual objects predicted reaction time differences between them. Collectively, these findings show that neural representational space and the encoding of objects is flexible and distinct under unisensory and more real-world multisensory contexts.

Keywords: Multisensory integration, Representational Similarity Analysis, EEG, Decoding, Object Recognition

Introduction

The brain is constantly bombarded with sensory information from a host of different sources. In extracting relevant information that can guide behavior, choosing which sensory information is valuable and which should be discarded is critical. This feat can be accomplished through top-down processes such as attention (Corbetta et al., 1990; Posner, 1980) or may also be done in a more automatic fashion by using the stimulus features processed along each step of the feedforward processing cascade (Alais & Burr, 2004; Angelaki et al., 2009; Körding et al., 2007). In this bottom up schema, some stimulus features may receive more weight than others given their ecological importance in guiding behavior or due to their regularity in the environment(Laws, 2000; Patten et al., 2017). Furthermore, an emerging body of literature in audition and vision suggests that these biases in perceptual weighting may propagate to the level of object representations (Carlson et al., 2014; Murray, 2006; Ritchie et al., 2015). Given that many objects in the world are specified not only on the basis of their unisensory features, but also by the fact that they are comprised of multisensory signals, an open question is how these multisensory cues are assembled in order to build object representations. Furthermore, how might biases in abstract object categories leads to differences in perceptual gains from multisensory integration?

In this study, we used representational similarity analysis to study the nature of neural representations for animate and inanimate objects presented in both the visual and auditory modalities, as well as the manner in which these representations change when objects are presented in a multisensory context. To this end, we presented subjects with auditory, visual and semantically congruent audiovisual animate and inanimate objects and had them perform a go/no-go animacy categorization task while we recorded high-density EEG. Our overarching hypothesis was that greater behavioral benefits would be seen for objects presented in a multisensory context, and that these benefits would be accompanied by greater changes in representational space. A secondary hypothesis was that given evidence that the benefits of multisensory integration increase when it is more difficult to process a stimulus using one modality alone, we would see greater multisensory benefits for the category (animate

or inanimate) that was categorized worse under unisensory conditions.

Behavioral Results

We first examined behavioral differences across sensory modalities and categories, as shown in Figure 1. Figure 1A shows mean reaction times (RTs) split by animate and inanimate category to investigate the effects of animacy on RTs.

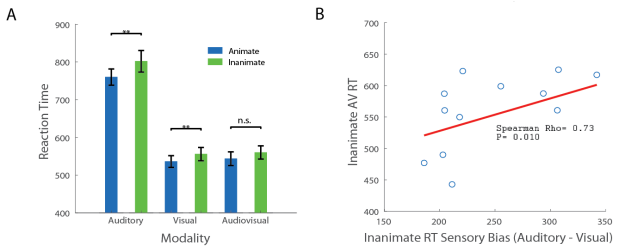

Figure 1: Animate and Inanimate Categorization Reaction Times by Modality and Sensory Bias

The bias towards animate objects is consistent with the results from previous studies (Carlson et al., 2014; Murray, 2006; Vogler & Titchener, 2011). However, we note that in the audiovisual conditions, there is no longer a RT bias. We further investigated this result by examining sensory bias, operationalized as the difference in audio and visual reaction times, and correlating this bias score to audiovisual RTs on a subject-by-subject basis using a Spearman correlation. Figure 1B shows a significant correlation between sensory bias and audiovisual RTs present for inanimate objects. The positive correlation indicates that subjects whose RTs for visual and auditory stimuli were most similar had the fastest multisensory RTs. There was no correlation between sensory bias and audiovisual RTs for animate objects (not shown).

Representational Similarity Analysis: Modality

To investigate the neural correlates of the behavioral differences between modalities and animacy categories, we used representational similarity analysis. First, we explored the effect of modality on the distinction between animate and inanimate exemplars by calculating the mean pairwise decoding for between category pairs (e.g., dog vs. bell, dog vs. cannon). As can be seen in figure 2A, before stimulus onset, decoding is at chance (i.e., 50%) because the classifier does not have any meaningful neural data that will distinguish between category pairs. However, shortly after stimulus onset, decoding performance becomes significant above chance (FDR corrected, $p<0.025$) across all three modalities. The latency of these decoding differences, defined as at least 20ms of sustained significant decoding (see Carlson, Tovar, Alink, & Kriegeskorte, 2013), were 88ms for auditory, 95ms for visual, and 60ms for audiovisual stimulus presentations. Visual and audiovisual decoding peaked at 163ms and 154ms, respectively, with higher peak decoding for audiovisual presentations, while auditory presentations showed comparatively lower decoding peaking at 190ms. To statistically compare decoding performance across modalities, we computed the mean decoding from shortly before the onset of significant decoding at 50ms post-stimulus to 600ms post-stimulus (figure 2B). We found that mean audiovisual decoding was significantly higher than both visual and auditory decoding (Wilcoxon signed rank test, $p<0.001$), and visual decoding was higher than auditory decoding. These results suggest that the audiovisual presentation of an object creates a more discernible distinction between animate and inanimate objects when compared to the corresponding unisensory presentations.

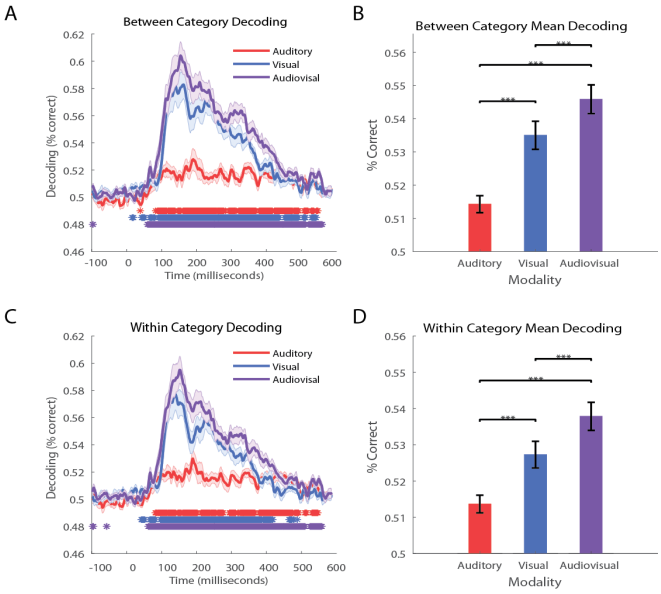

Figure 2: RSA Between and Within Animate Category Decoding across Modalities

We further explored whether audiovisual presentations expanded exemplar distinctions within animacy categories by calculating the mean within category pairwise decoding accuracies (Figure 2C). In this analysis, onset latencies for significant decoding for audio, visual, and audiovisual stimuli were 89ms, 99ms, and 62ms. The corresponding peak decoding latencies were 190ms, 140ms, and 152ms. The modality specific comparisons for within-category decoding (Figure 2D) mirrored that seen for between-category decoding with

higher audiovisual decoding when compared with visual and auditory decoding, and higher visual decoding than auditory decoding (Wilcoxon signed rank test, p<0.001). A comparison of between-category decoding and within-category decoding demonstrated higher decoding for between category decoding for visual and audiovisual stimulus presentations (Wilcoxon signed rank test, p<0.001) but only a marginally significant difference for auditory presentations (p=. 09). In sum, when compared to unisensory presentations, audiovisual stimulus presentations not only expand the representational space between animacy categories, but also make exemplars within the animacy categories easier for a classifier to distinguish.

Representational Similarity Analysis: Animacy

We further investigated representational space broken down by animacy categories to study the neural underpinnings for the reaction time differences between animate and inanimate presentations (Figure 3).

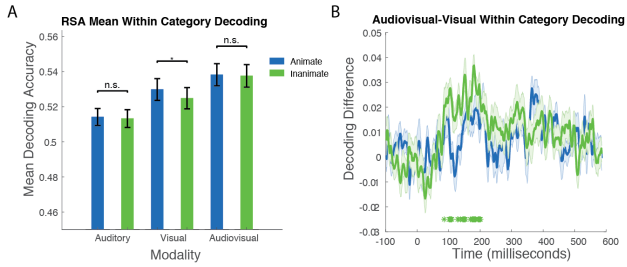

Figure 3: Within Animate Category Decoding and Audiovisual Enhancement by Animacy Category

We quantified the difference between animacy categories by using the mean decoding performance during the stimulus period [50ms to 600ms] and confirmed that there was a significant animacy category difference for visual presentations (p<0.05), but not the other modalities. Since audiovisual presentations had overall higher within category pairwise decoding than visual presentations (Figure 2D), we probed whether the lack of animate and inanimate within category decoding difference for audiovisual presentations was due to visual inanimate objects incurring a special benefit from audiovisual presentation. Figure 3B shows the difference between audiovisual decoding and visual decoding for animate and inanimate exemplars. The audiovisual/visual decoding difference is significantly above zero across several timepoints between 100ms to 200ms post stimulus onset for inanimate objects (Wilcoxon signed rank p<0.025, FDR corrected), but not for animate objects. Furthermore, a comparison of mean decoding performance difference across the 100ms to 200ms time period reveals a significant difference between animate and inanimate exemplars (Wilcoxon signed rank, p=.001). Together, these results suggest that audiovisual presentations may asymmetrically enhance the neural representations of inanimate objects.

Distance to bound: Relating neural data to reaction times

Having found both behavioral and neural differences between sensory modalities and animacy categories, we next considered whether the two measures were associated with one another. To do this, we computed the distance to the classifier decision boundary for all exemplars and correlating these distances back to reaction times. A negative correlation would denote that exemplars that are farthest away from the classifier decision boundary are those that are fastest categorized. Figure 4A shows a significant negative Spearman correlation (Wilcoxon signed-rank test, FDR threshold = 0.025) between representational distance and reaction time at several timepoints between 100 and 200ms post-stimulus and between 270 and 400ms post-stimulus for visual and audiovisual presentations. Auditory presentations did not show any significant timepoints.

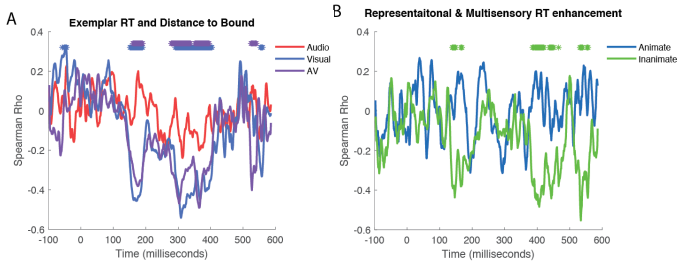

Figure 4: (A) Distance to Bound Analysis across modalities and (B) Distance to Bound Audiovisual Enhancement Analysis.

In Figure 4B, we studied behavioral and neural enhancement by using a Spearman correlation to link the audiovisual-visual reaction time difference with the audiovisual-visual representational difference for animate and inanimate exemplars. A negative correlation denotes: 1) Exemplars that were further away from the decisional boundary for audiovisual presentations when compared with visual presentations (positive AV-V distance value), are also the exemplars that demonstrate either more of an audiovisual RT bias (positive AV-V RT value) or less of a visual bias (negative AV-V RT value). 2) Exemplars that were further away from the decision boundary for visual presentations when compared with audiovisual presentations (negative AV-V distance value), are also

the exemplars that demonstrate less of an audiovisual RT bias (positive AV-V RT value) or more of a visual bias (negative AV-V RT value). Given that this data was relatively noisy, we used a less conservative FDR threshold of 0.10 to identify timepoints where representational distance differences show a marginally significant correlation to reaction time differences. Using this criterion, we found several significant timepoints between 100 and 200 ms post-stimulus and 370 and 450 ms post-stimulus for inanimate exemplars, but no significant timepoints for animate exemplars. If we calculate the mean correlation across the entire stimulus analysis epoch [50ms - 600ms] we find a significant negative correlation for inanimate exemplars (Wilcoxon signed rank, r=-0.1661 p< 0.0001) but not for animate exemplars (Wilcoxon signed rank, r=.0045 p=0.11). Collectively, these results show associations between that neural decoding and behavioral performance for audiovisual and visual stimulus presentations.

Conclusion

The results of our study provide new insights into how perceptual unisensory differences affects their subsequent integration as well as establishes critical links between neural activity and behavior. Furthermore, we show that neural representational space and object encoding is flexible. Understanding these mechanisms is the first step towards understanding a number of pathologies which show deficits in sensory integration(Burnett, Panis, Wagemans, & Jellema, 2015; Loth, Gómez, & Happé, 2010).

Acknowledgments

This work was supported by a NIGMS of the National Institutes of Health (T32GM007347) and Swiss National Science Foundation (Grants 149982 and 169206).

References

Alais, D., & Burr, D. (2004). Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current Biology*, *14*(3), 257–262.

Angelaki, D. E., Klier, E. M., & Snyder, L. H. (2009). A Vestibular Sensation: Probabilistic Approaches to Spatial Perception. *Neuron*, *64*(4), 448–461.

Burnett, H. G., Panis, S., Wagemans, J., & Jellema, T. (2015). Impaired identification of impoverished animate but not inanimate objects in adults with high-functioning autism spectrum disorder. *Autism Research*, *8*(1), 52–60.

Carlson, T. A., Ritchie, J. B., Kriegeskorte, N., Durvasula, S., & Ma, J. (2014). Reaction Time for Object Categorization Is Predicted by Representational Distance. *Journal of Cognitive Neuroscience*, *26*(1), 132–142.

Carlson, T. A., Tovar, D., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. *Journal of Vision*, *13*(10), 1–19.

Corbetta, M., Miezin, F., Dobmeyer, S., Shulman, G., & Petersen, S. (1990). Attentional modulation of neural processing of shape, color, and velocity in humans. *Science*, *248*(4962), 1556–1559.

Grootswagers, T., Ritchie, J. B., Wardle, S. G., Heathcote, A., & Carlson, T. A. (2017). Asymmetric Compression of Representational Space for Object Animacy Categorization under Degraded Viewing Conditions. *Journal of Cognitive Neuroscience*, *29*(12), 1995–2010.

Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. *Journal of Cognitive Neuroscience*, *29*(4), 677–697.

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS ONE*, *2*(9).

Loth, E., Gómez, J. C., & Happé, F. (2010). When seeing depends on knowing: Adults with Autism Spectrum Conditions show diminished top-down processes in the visual perception of degraded faces but not degraded objects. *Neuropsychologia*, *48*(5), 1227–1236.

Murray, M. M. (2006). Rapid Brain Discrimination of Sounds of Objects. *Journal of Neuroscience*, *26*(4), 1293–1302.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25.

Ritchie, J. B., Tovar, D. A., & Carlson, T. A. (2015). Emerging Object Representations in the Visual System Predict Reaction Times for Categorization. *PLoS Computational Biology*, *11*(6).

Vogler, J. N., & Titchener, K. (2011). Cross-modal conflicts in object recognition: Determining the influence of object category. *Experimental Brain Research*, *214*(4), 597–605.