The influence of noise on multimodal spatial localization tasks

People's conscious experience appears to be consisting of unified percepts made of complex information coming from multiple sensory systems. Although stimuli come to the processing system (a brain) with different speed, they get integrated to produce a robust experience. In spatial localization tasks, on top of different speeds of particular types of signals, stimuli may also come from different locations. Current research on multimodal integration in spatial localization tasks mostly explores relationships between the modalities of vision, hearing and touch. One of the most famous effects related to sensory integration is called ventriloquism effect, where the perceived location of the sound is manipulated by the location of the visual stimulus in a way that people perceive auditory stimulus as biased towards the location of the visual one (e.g. [1]). This points out the vision as the dominant stimulus in multimodal perception. Other studies, however, concluded that when subjects are presented with an event related more to temporal processing, rather than spatial processing, as in the sound-induced flash illusion [3], there is a bias towards an acoustic information. Both of these examples point to the model called Maximum Likelihood Estimation (MLE) that, depending on a situation a subject is in, takes the modality which is the most reliable one (has the lowest variance) and assigns it a higher weight than to other more senses with more varied information.

According to the MLE model of spatial localization, the location of a signal L is given by the weighted average of the visual and the auditory location estimates, i.e.

$$L = w_v L_v + w_a L_a.$$

Where w_v, w_a are the visual and auditory weights summing to one. The value of the weights is assumed to be proportional to the reliability of the signals. This means that, for example, a reliable visual location estimate would have a weight closer to one, and a more unreliable, noisy auditory signal would have a weight closer to zero. Consequently, the respective senses would be more or less relevant to the final location estimate.

An alternative theory to the MLE model is a Bayesian model, where priors are added to these weights. In addition to the situational stability of a signal, more general, 'top-down' information on reliability is encoded. The assumption, based on the so-called visual capture theory, is that visual information is generally more reliable and thus has a higher prior.

The current paper will present results from a replication of a study performed by Battaglia et al. [2] with a regard to other study by Alais and Burr [1]. Both of them investigated spatial localization with multimodal (visual and auditory) stimuli. Both of the studies noticed change of the weights after a noise was introduced to visual stimuli. The results we obtained are in line with the original study by Battaglia et al

[2]. The weights were obtained using the variance of the location estimates based on visual and auditory signals alone, then tested in a multimodal setting. More noisy signals resulted in smaller weights, whereas more reliable signals yielded weights closer to one. As in the original work, the visual modality seemed to be systematically underestimated, supporting the potential of introducing a prior towards vision. Additionally, it has been shown that in the block with visual stimuli, participants were less accurate when the noise was higher, but accuracy in the block with combined stimuli was not damaged.

In addition to discussion about the meaning of the successful replication of the experiment by Battaglia et al. [2], we aim to discuss possible adjustments to the introduced mathematical model towards the bayesian one including priors and practical as well as theoretical implications of obtained results. Regarding the bayesian model, an interesting question remaining is where these priors on sensory reliability originate. Are they learned during lifetime or encoded at an even deeper level: evolutionary.

Current research will be evaluated particularly from the perspective of multimodal communication. Furthermore, suggestions for further research will be presented, including methods allowing for automatic recognition of dominating modality from the activation pattern in the brain.

References:

[1] Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. Current biology, 14(3), 257-262.

[2] Battaglia,P.W.,Jacobs,R.A.,&Aslin,R.N.(2003) .Bayesian integration of visual and auditory signals for spatial localization.JOSAA, 20(7), 1391-1397.

[3] Shams, L., Ma, W. J., & Beierholm, U. (2005). Sound-induced flash illusion as an optimal percept. Neuroreport, 16(17), 1923-1927.