Models of Bimodal Duration Discrimination

Reliability based duration estimation assuming fully fusion

This model assumes participants optimally fuse auditory and visual durations when AV conflict is negligible or ignored, as in the low-conflict AV condition.

We will go further to explain the case when the conflict between the modalities are too large to ignore and we cant simply assume fusion but for now lets start with simpler case. When there is no conflict or the conflict is o small, we can safely assume fusion. If there is fusion in the perception (perceiving the stimulus is originated from same source), we can apply the rules of optimal integration where each cue is weighted by its reliability.

$$\hat{S}_{av} = \frac{\sigma_a^{-2} s_a + \sigma_v^{-2} s_v}{\sigma_a^{-2} + \sigma_v^{-2}}$$

$$\hat{S}_{av} = \frac{J_a s_a + J_v s_v}{J_a + J_v}$$

$$= w_a s_a + w_v s_v \tag{21}$$

Where σ_a^{-2} and σ_v^{-2} are the inverse variances of the auditory and visual modalities, respectively.

In the task of 2-IFC, let AV be the combined auditory and visual stimulus, and let s_{av} be the duration of the AV stimulus. The internal estimate of the duration of the AV stimulus can be expressed as:

Then the duration of each modality in test interval can be defined as:

$$s_{a,t} = s_{a,s} + \Delta s$$
$$s_{v,t} = s_{v,s} + \Delta s + \mu_{bias}$$

Where Δs is the difference in duration between the test and standard intervals based on the staircase method.

In the standard interval S_s the duration of auditory modality is simply the standard duration of experiment (0.5s). But the visual modality is biased by the prior bias duration and also by the conflict duration between the modalities depending on the condition.

$$s_{a,s} = 0.5$$
$$s_{v,s} = 0.5 + \mu_{bias} + c_i$$

Where c_i is the conflict duration between the modalities in i-th trial.

The weights for each modality can be defined as:

$$w_a = \frac{\sigma_a^{-2}}{\sigma_a^{-2} + \sigma_v^{-2}} \quad \text{and} \quad w_v = \frac{\sigma_v^{-2}}{\sigma_a^{-2} + \sigma_v^{-2}} \tag{22}$$

Where σ_a^{-2} and σ_v^{-2} are the inverse variances (precision) of the auditory and visual modalities, respectively.

The variance of the integrated estimate can be defined as:

$$\sigma_{av}^2 = \frac{1}{\sigma_a^{-2} + \sigma_v^{-2}}$$

Then we can formulate the decision rule as probability of perceiving the test interval longer than the standard interval:

$$D = \hat{S}_t^{AV} - \hat{S}_s^{AV} = w_a(m_a^t - m_a^S) + w_v(m_v^t - m_v^s)$$
$$= w_a \Delta S_a + w_v \Delta S_v$$

If the true durations for test and standard are S_t and S_s , and their measurements are:

$$m_a^t = s_{a,t} \sim N(s_{a,t}, \sigma_a^2)$$
$$m_a^s = s_{a,s} \sim N(s_{a,s}, \sigma_a^2)$$
$$m_v^t = s_{v,t} \sim N(s_{v,t}, \sigma_v^2)$$
$$m_v^s = s_{v,s} \sim N(s_{v,s}, \sigma_v^2)$$

Then the difference in the measurements can be defined as:

$$m_a^t - m_a^s \sim N(\Delta s_a, \sigma_a^2)$$
$$m_v^t - m_v^s \sim N(\Delta s_v, \sigma_v^2)$$

The difference in the integrated estimate can be defined as:

$$D = w_a(m_a^t - m_a^s) + w_v(m_v^t - m_v^s) \sim N(w_a \Delta s_a + w_v \Delta s_v, 2\sigma_{av}^2)$$
$$D \sim N(S_t - S_s, 2\sigma_{av}^2)$$

Where w_a and w_v are the weights for the auditory and visual modalities, respectively.

The decision rule can be defined as the probability of perceiving the test interval longer than the standard interval:

$$P(\Delta_{t-s} > 0) = P(w_a \Delta S_a + w_v \Delta S_v > 0)$$

$$P(\text{test longer}) = \Phi(\frac{S_t - S_s}{\sqrt{2\sigma_{av}^2}})$$

$$P(\text{test longer}) = \Phi\left(\frac{w_a \Delta S_a + w_v \Delta S_v}{\sqrt{2}\sigma_{av}}\right)$$

Where (Phi) is the cumulative standard normal CDF.

Causal inference model for auditory-visual duration discrimination

When we perceive a multimodal stimulus, observer has to also deal with the problem of common source assumption.

In our experiment this is crucial for giving accurate estimates of auditory duration, because if there is a huge conflict between the auditory and visual duration,

the observer should ignore the visual duration and rely on the auditory duration as that would be the bayesian optimal solution.

The model assumes that the observer has a prior belief about the likelihood of the auditory and visual stimuli being caused by the

same source. This is represented by a prior probability $(P(\mathrm{common}))$.

The observer then uses this prior to update their beliefs about the auditory and visual durations based on the observed durations. The model can be summarized as follows:

In the case of full causal inference, the observer assumes that the auditory and visual durations are caused by the same source. The posterior probability of the common cause is given by:

$$\hat{S}_{av,a} = \hat{S}_{av,v} = \frac{\sigma_a^{-2} m_a + \sigma_v^{-2} m_v}{\sigma_a^{-2} + \sigma_v^{-2}}$$

$$\sigma_{av}^2 = \frac{1}{\sigma_a^{-2} + \sigma_v^{-2}}$$

In the case of no common cause, the observer assumes that the auditory and visual durations are independent. The posterior probability of the independent cause is given by:

$$\hat{S}_{av,a} = s_a, \quad \hat{S}_{av,v} = s_v$$
$$\sigma_{av,a}^2 = \sigma_a^2, \quad \sigma_{av,v}^2 = \sigma_v^2$$

We assume that common cause inference is not a binary decision, but rather a continuous variable that can take any value between 0 and 1. The observer can then use this variable to weight the auditory and visual durations in their final estimate. The final estimate is given by:

$$\hat{S}_{av,a} = \hat{S}_{a,C=1} \cdot P(C = 1|m_a, m_v) + \hat{S}_{a,C=2} \cdot (1 - P(C = 1|m_a, m_v))$$

Where $(P(C = 1|S_a, S_v))$ is the posterior probability of the common cause given the auditory and visual durations, observer has measurements:

- Auditory measurement$ m_a
- Visual measurement: m_v
  Each is a noisy sample::

$$ \quad m_a \sim N(S_a, \sigma^2_{av,a})$$

$$m_v \sim N(S_v, \sigma^2_{av,v})$$

This model allows the observer to flexibly integrate the auditory and visual durations based on their prior beliefs about the common cause, leading to more accurate estimates of the durations.

similarly for the visual duration:

$$\hat{S}_{av,v} = \hat{S}_{v,C=1} \cdot P(C = 1|m_a, m_v) + \hat{S}_{v,C=2} \cdot (1 - P(C = 1|m_a, m_v))$$

At this point to solve the model we need to obtain the probability of common source.

$$ \quad P(C|m_a, m_v) = \frac{P(m_a, m_v|C)}{P(m_a, m_v)}$$

and in this case $p(c = 1|m_a, m_v) + p(c = 2|m_a, m_v) = 1$.

When we expand the the common source equation for C=1 posterior would be:

Final percept:

$$P(C = 1|m_a, m_v) = \frac{p(m_a, m_v|C = 1) \cdot p(C = 1)}{p(m_a, m_v|C = 1)p(C = 1) + p(m_a, m_v|C = 2)(1 - p(}$$

Then the final auditory estimate would be:

$$Final = p(C = 1)\text{Fused} + (1 - p(C = 1))\text{Auditory est}$$

And the posterior probability for not a common source is $1 - P(C = 1|m_a, m_v)$.

Now the problem becomes how to estimate the likelihood of the common source given the auditory and visual measurements.

The likelihood of the common source can be estimated using a Gaussian distribution, assuming that the auditory and visual measurements are independent given the common cause. The likelihood can be expressed as:

$$p(m_a, m_v|C = 1) = \int p(m_a, m_v|s_{a,v})p(s_{a,v})ds_{s,v}$$

$$\boxed{} = \int p(m_a|s_a)p(m_v|s_v)p(s_{a,v})ds_{s,v}$$

Likelihood of the common cause (C=1):

$$P(m_a, m_v | C = 1) = \frac{1}{2\pi\sqrt{\sigma_{av,a}^2\sigma_{av,v}^2 + \sigma_{av,a}^2\sigma_p^2 + \sigma_{av,v}^2\sigma_p^2}}$$

$$\exp\left(-\frac{1}{2}\frac{(m_a - m_v)^2\sigma_p^2 + (m_a - \mu_p)^2\sigma_{av,v}^2 + (m_v - \mu_p)^2\sigma_{av,a}^2}{\sigma_{av,a}^2\sigma_{av,v}^2 + \sigma_{av,a}^2\sigma_p^2 + \sigma_{av,v}^2\sigma_p^2}\right)$$

But in this experiment as we are using interleaved 2-IFC AV duration discrimination task we can ignore the prior in the model ($\sigma_p^2 - \infty$)

This simplifies the likelihood of common-cause to:

$$P(m_a, m_v \mid C = 1) = \frac{1}{\sqrt{2\pi(\sigma_a^2 + \sigma_v^2)}}\exp\left(-\frac{(m_a - m_v)^2}{2(\sigma_a^2 + \sigma_v^2)}\right)$$

Here, **measurement noise ($\sigma_{av,a}$ and σ_{av_v}) is fit directly from AV data**; we do not assume they are identical to unimodal noise($\sigma_{av,a} \neq \sigma_a$).

Likelihood of independent sources (C=2):

$$P(m_a, m_v | C = 2) = P(m_a) \cdot P(m_v)$$

$$= \frac{1}{2\pi\sigma_{av,a}\sigma_{av,v}}\exp\left(-\frac{(m_a - s_a)^2}{2\sigma_{av,a}^2} - \frac{(m_v - s_v)^2}{2\sigma_{av,v}^2}\right)$$

Decision variable:

Finally model should compare the estimated duration for both 'test' and 'standard' intervals decision should be computed as:

$$D = \hat{S}_{CI,t} - \hat{S}_{CI,s}$$

$$P(\text{"test longer"}) = \Phi\left(\frac{D}{\sqrt{2}\sigma_{CI}}\right)$$

$\hat{S}_{CI,t}$ The **final internal estimate** for the duration of the **test** interval and $\hat{S}_{CI,s}$ is the **final internal estimate** for the **standard** interval after causal inference.

σ_{CI} is the **effective sensory noise** (standard deviation) associated with the internal estimate in the causal inference model.