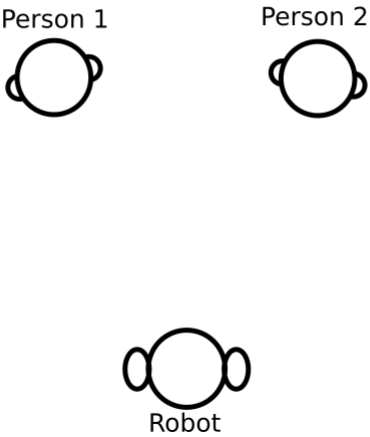
Dynamic neural fields and manifold learning for audiovisual fusion in psychophysics and robotics

Simon Forest[1, 2], Jean-Charles Quinton[1], Mathieu Lefort[2]

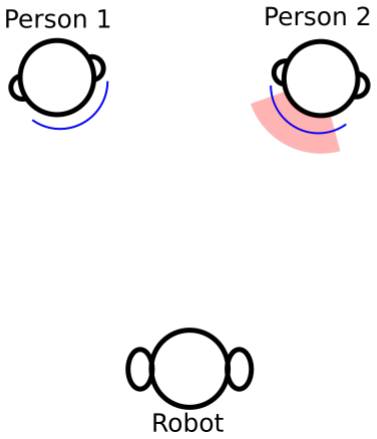
[1]LJK, Univ. Grenoble Alpes
[2]LIRIS, Univ. Claude Bernard Lyon 1

23/08/2022
Workshop on Methods and Tools for Audio-Visual Processing and Human Robot Interaction

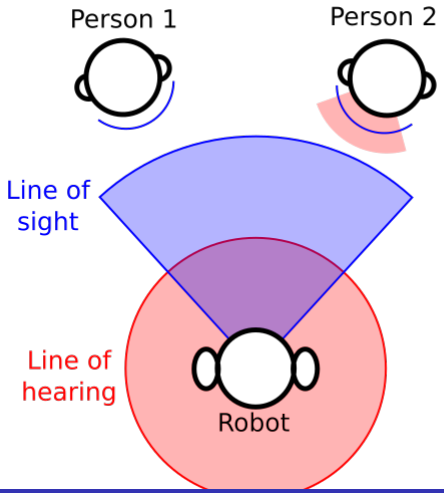
Example: stimulus localization and selection



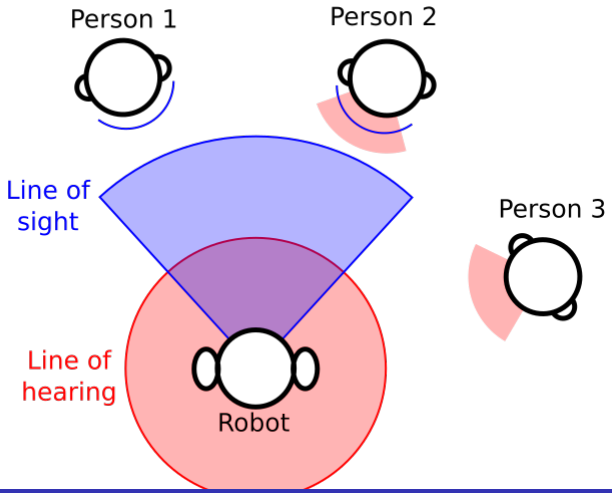
Example: stimulus localization and selection



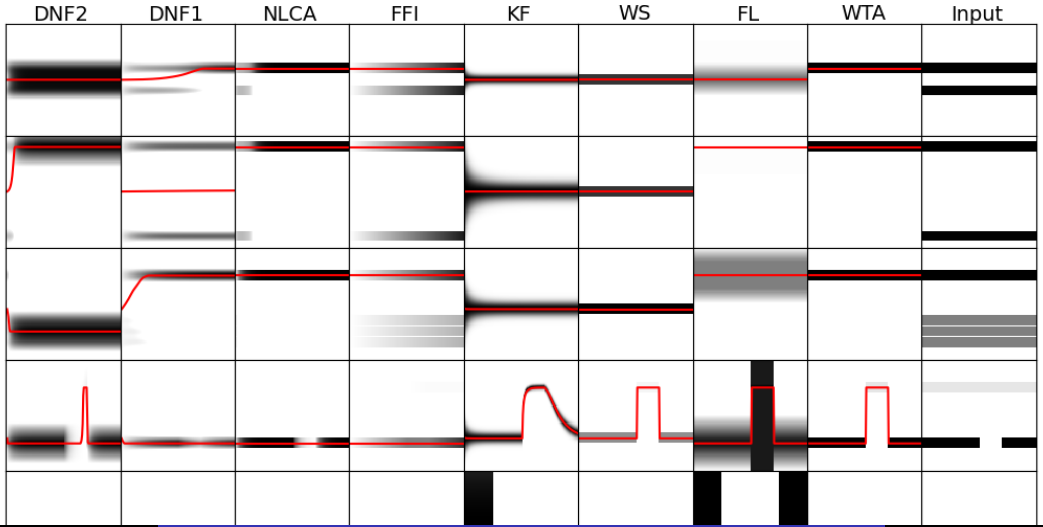
Example: stimulus localization and selection



Example: stimulus localization and selection



Decision-making algorithms (review article in preparation)

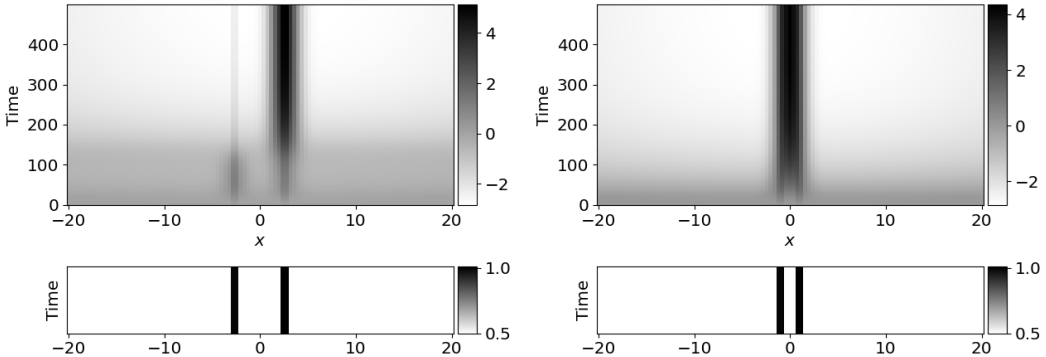


Dynamic neural fields (DNF) [Amari, 1977]

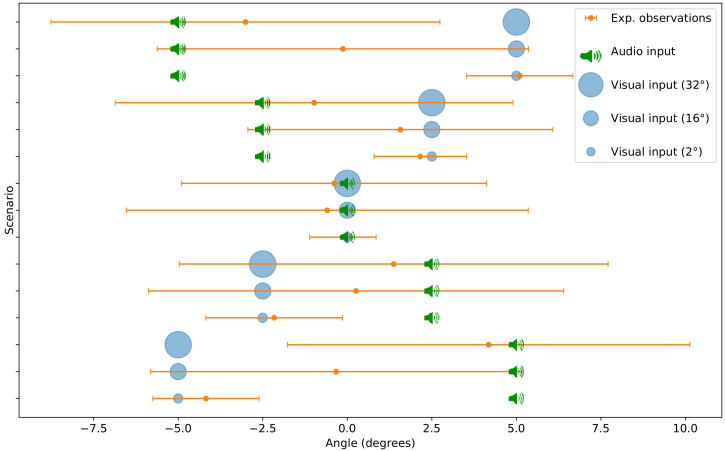
$$\tau \frac{\partial U}{\partial t}(x,t) = -U(x,t) + I(x,t) + \int_{y \in X} W\left(\|x-y\|\right) f\left(U(y,t)\right) \mathrm{d}y + h$$

U: membrane potential I: input f: activation function $h < 0$: resting level

$W(d) = \lambda_+ \exp\left(-\frac{d^2}{2\sigma_+^2}\right) - \lambda_- \exp\left(-\frac{d^2}{2\sigma_-^2}\right)$: interaction kernel

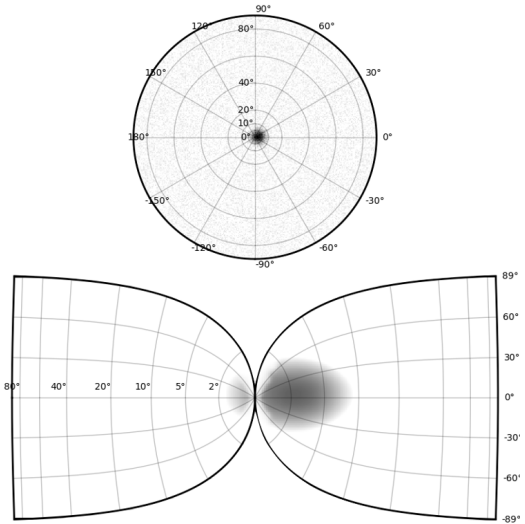


Application to the ventriloquist effect [Forest et al., 2022]



- Inspired by superior colliculus
- Qualitative fit to exp. data
- Parameter sensitivity analysis
- To be applied to new data (+ saccades)

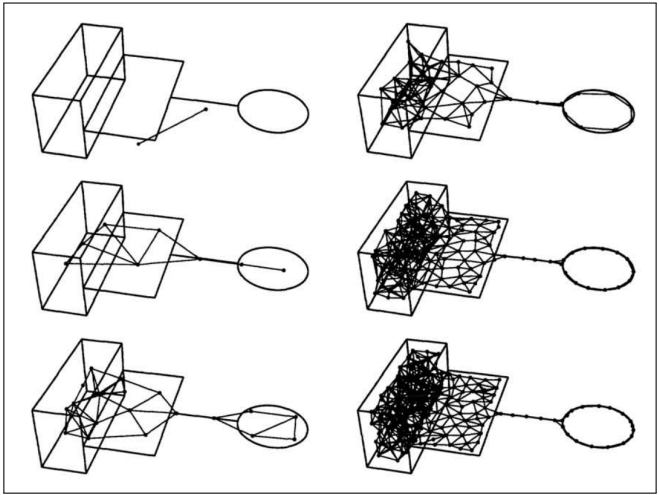
New problem



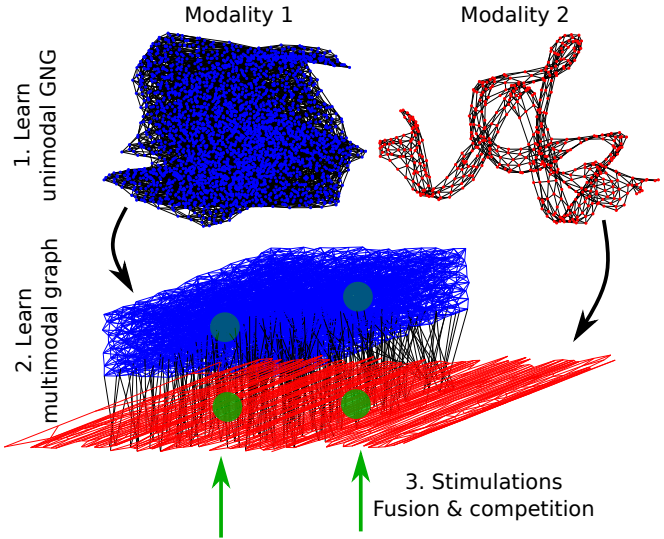
- Usually assuming merging in aligned, regular spaces
- Not faithful to actual perception
  e.g. fovea (top) → superior colliculus (bottom)
- Existing models of selection/fusion in arbitrary topologies
- Contribution: selection/fusion in learned topologies

Growing neural gas (GNG) [Fritzke, 1995]

A simple manifold learning algorithm:



Creating multimodal topologies



Modality 1: 3D

Modality 2: 100D

Modalities linked by co-activations

Projected in 2D for visualization

DNF adaptation

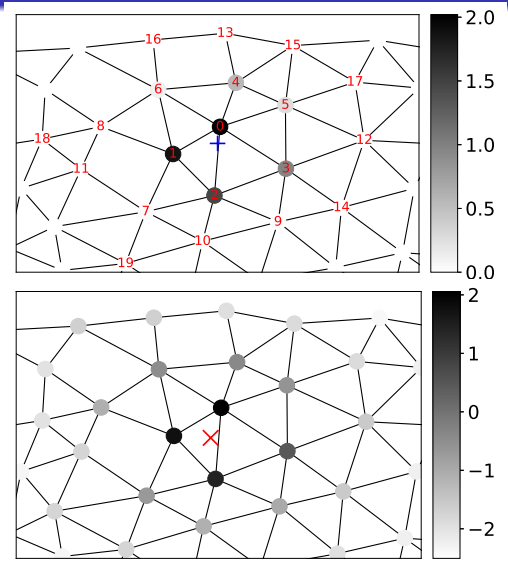
For each node k:

$$I_k = \lambda_{modality,s}\, \mathrm{e}^{\frac{-r_{k,s}^2}{2\sigma^2}}$$

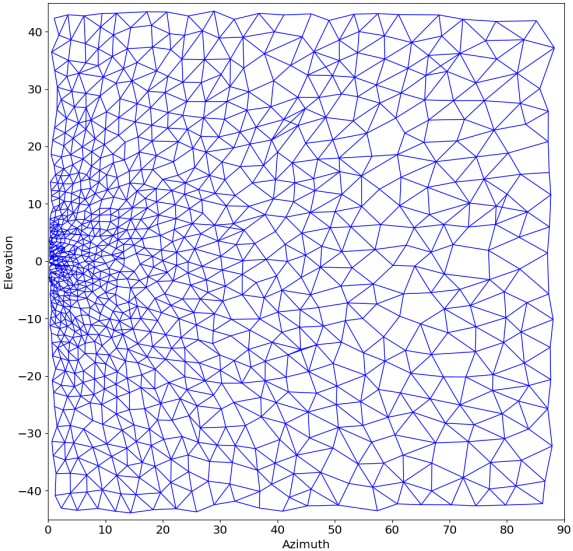
$r_{k,s}$: rank of k by proximity to stimulus s

$$\Delta U_k = \frac{\Delta t}{\tau}\left(-U_k + I_k + \sum_{k'} W(<k,k'>)\, f(U_{k'}) + h\right)$$

$<k,k'>$: distance between k and k' in multimodal graph

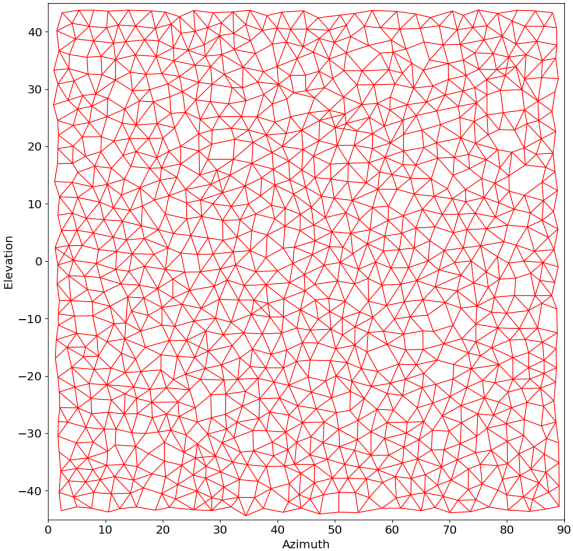


Example inspired from superior colliculus



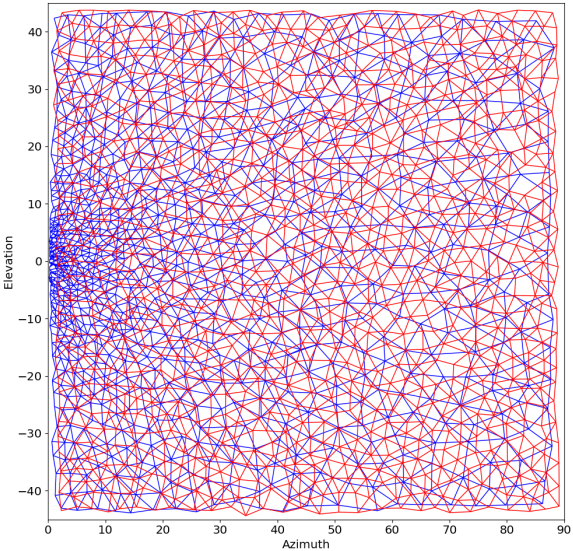
Vision: 2D+logpolar (blue)

Example inspired from superior colliculus



Audio: regular 2D (red)

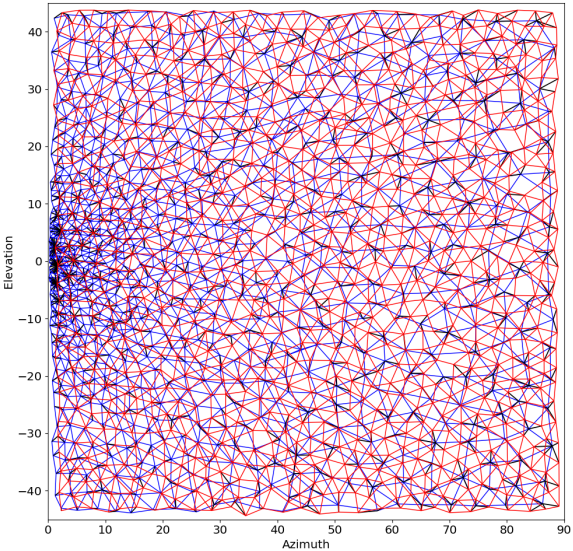
Example inspired from superior colliculus



Vision: 2D+logpolar (blue)

Audio: regular 2D (red)

Example inspired from superior colliculus

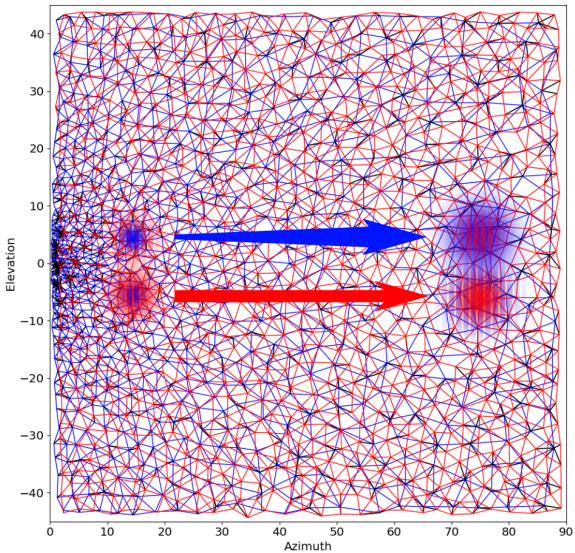


Vision: 2D+logpolar (blue)

Audio: regular 2D (red)

Added crossmodal edges (black)

Example inspired from superior colliculus



Vision: 2D+logpolar (blue)

Audio: regular 2D (red)

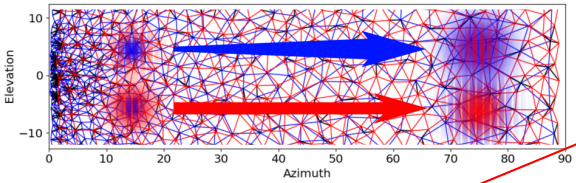
Added crossmodal edges (black)

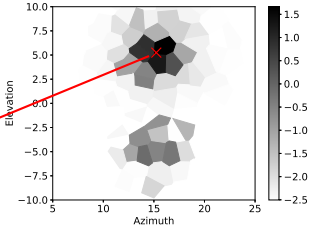
Tests $\forall$ azimuths:
2 bimodal stimuli:

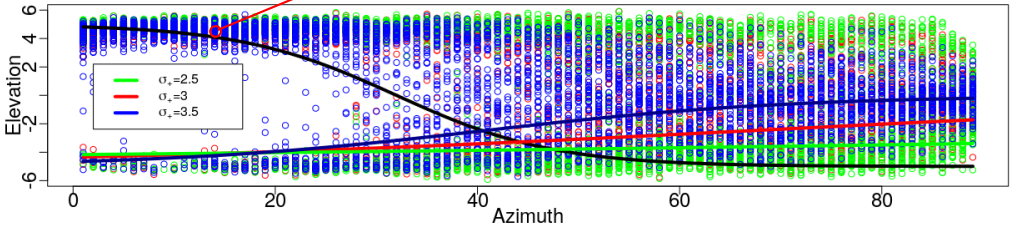
- Top: stronger visually
- Bottom: stronger auditorily

$\rightarrow$ multiple DNF runs

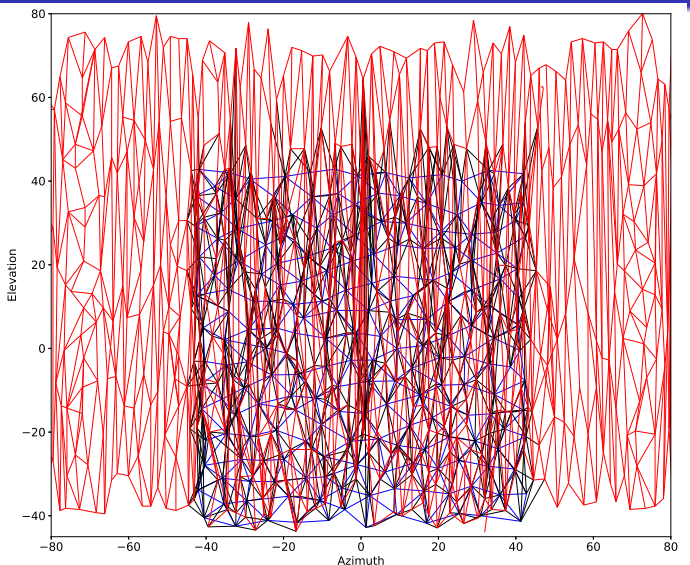
Model analysis (50 simulations × 90 azimuths × 3 DNF kernels)



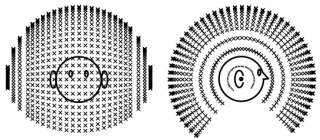




Using data from robotics



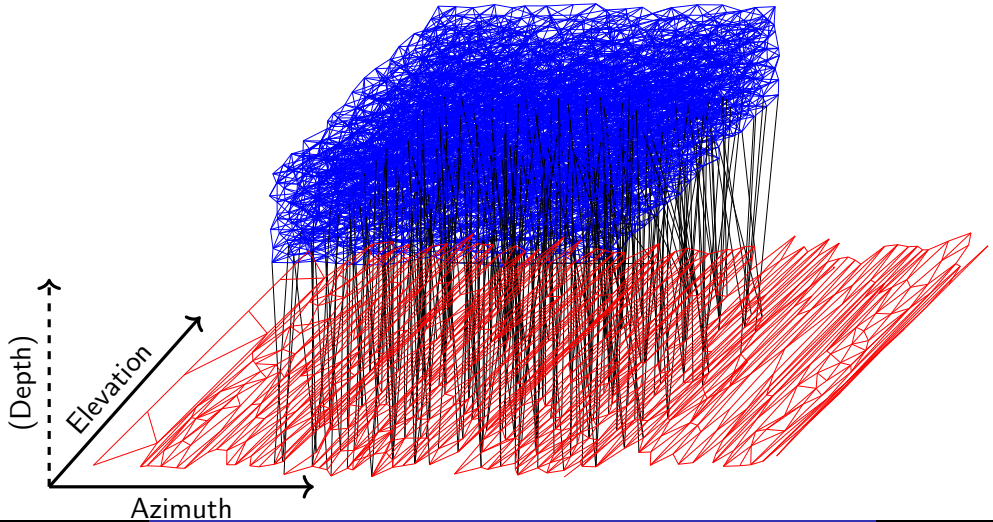
- Vision in blue: regular 2D
- Audition in red: 100D HRTF data



[Algazi et al., 2001]

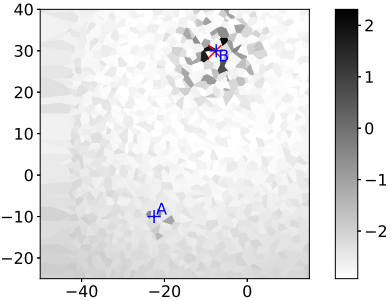
(in 2D for visualization)

Same with 3D vision



Results

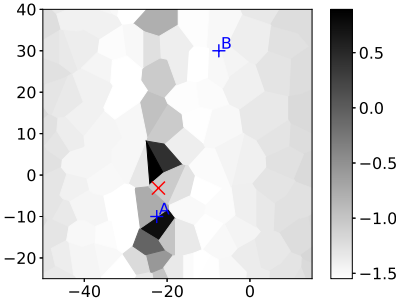
Visual: 3D, regular



A and B visible (B > A)

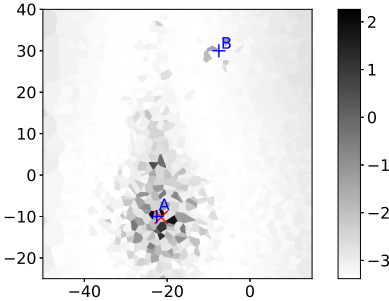
⇒ B selected
(high precision)

Auditory: 100D, HRTF



A audible
⇒ A selected
(low precision)

Bimodal



A and B visible (B > A)
A audible
⇒ A selected
(high precision)

Conclusion

Contributions and results

- Adaptation of neuro-inspired model (DNF) to irregular topologies
- Learning and combination of manifolds of different sensory space
- Consistent results in selection tasks

Perspectives

- High-dimensional data using deep learning
- Integration in robotics
- Implementation of eye movements

PS: PhD defense on September 16th, 10 AM in Villeurbanne (+ video retransmission)

References

- Amari, S.-I. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27(2):7787.
- Forest, S., Quinton, J.-C., and Lefort, M. (2022). A dynamic neural field model of multimodal merging: application to the ventriloquist effect. *Neural Computation*, 34(8).
- Fritzke, B. (1995). A growing neural gas network learns topologies. In Tesauro, G., Touretzky, D., and Leen, T., editors, *Advances in Neural Information Processing Systems*, volume 7. MIT Press.
- Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (2001). The CIPIC HRTF database. In *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No. 01TH8575)*, pages 99102. IEEE.
- + Forest, S., Quinton, J.-C., and Lefort, M. (2022). Combining manifold learning and neural field dynamics for multimodal fusion. *2022 International Joint Conference on Neural Networks (IJCNN)*.

Appendix: eye movements