Sounds of the Woods

Angela He and Kyle Qian

Stanford University Music 257

Motivation

- In order to perceive the world as accurately as possible, our brains must successfully integrate information from multiple sensory modalities
- As a result of this multimodality, our brains must be able to identify and resolve stimuli that correspond to the same or different external entities, as described by the Bayesian model of cue integration[5]
- Successful sensory integration, however, is contextual—for instance, while we may normally favor visual stimuli for localization, that strategy losses effectiveness in the dark
- We want to train individuals to be more flexible in how they perform Bayesian cue integration, specifically for favoring audio cues for localization over visual cues in low visibility conditions

Background

Duplex Theory of Sound Localization

- Interaural time difference (ITD)
  - A measure of the difference in arrival times of sounds between two ears
  - Effective for localization of lower frequency sounds, particularly on the azimuthal plane (i.e. sounds that differ in their horizontal location)
  - Can be digitally reproduced without much user-specific HRTF data
- Interaural level difference (ILD)
  - A measure of the difference in loudness and frequency distribution between two ears
  - Effective for localization of higher frequency sounds, particularly for elevation-varied sounds when ITDs are identical
  - Difficult to effectively implement digitally without personalized HRTF data

Bayesian Inference

- A statistical framework for modeling uncertainty about sensory inputs, combining sensory estimation, prior data, and a decision-making process[2]
- A Bayesian approach to modeling multimodal cues requires a prior describing the expected interaction between different sensory modalities (coupling prior)[1]
- The *ventriloquist effect* is a prominent example of this:
  - Visual input dominates other sensory inputs for the purposes of localization (a ventriloquist's voice is of course coming from their mouth, but an audience incongruently perceives it as coming from the puppet's)
  - A Bayesian explanation would take into account the spatial reliability of visual input[4]

Figure 1: Day and night difference

Neural Representation

Convolution Coding[3]

- To represent Bayesian probability distributions, neurons must somehow encode continuous variables
- Probability density functions can be sparsely and discretely encoded by sampling
- Gaps between the samples can be resolved by filtering the probability density function
- This way, Bayesian inferences can be calculated via a product operation, much like multiplying the underlying functions

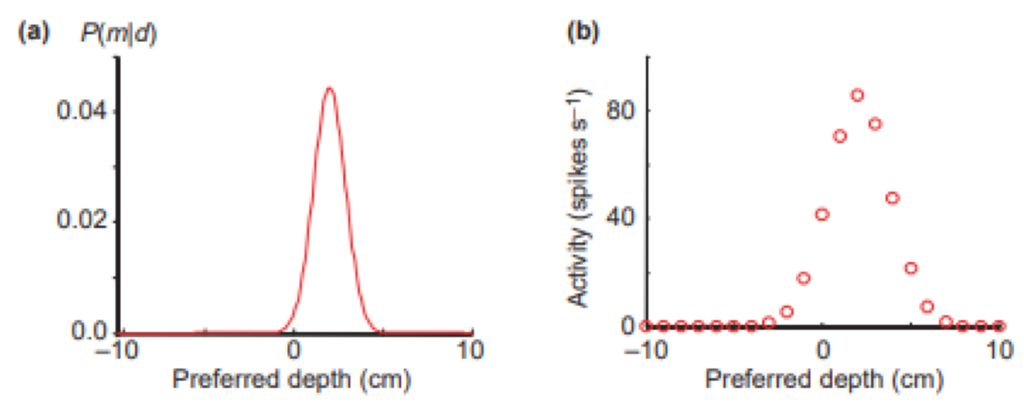

Figure 3: A probability density function and associated convolution code for representing depth perception

Gain Encoding[3]

- The gain encoding scheme is an alternative to convolution coding
- Models neural noise as a Poisson distribution, thus allowing simultaneous coding of both mean and variance
- In the case of orientation preference, plotting adjacent orientation-selective cells against their respective neural response to a stimulus yields a probability density curve for that input

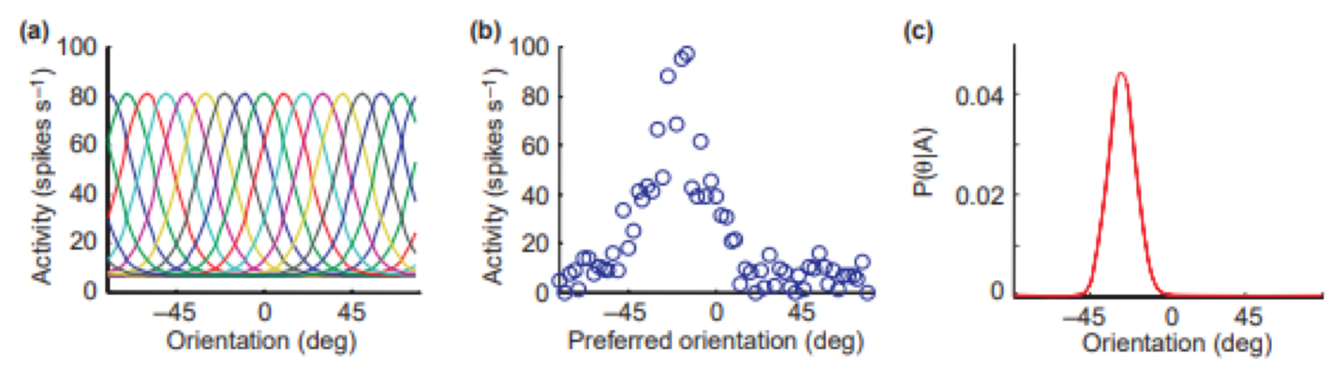

Figure 4: Tuning curves for a set of orientation-selective neurons (left); recorded responses of orientation-selective neurons (middle); predicted hypothetical distribution (right)

Game Implementation

You're lost and injured in the woods, and it's quickly getting dark. All you have with you is a flashlight and a radio. You called for help on the radio, but it looks like it'll take until sunrise until help can arrive. You see friendly woodland creatures as the sun sets, but as it gets darker, you start seeing and hearing other creatures lurking in the night.

- Monsters of the woods are scared of light, so shining your flashlight at them should make them go away
- Each creature will have its own spatialized auditory signature
- The darker it gets, the harder it is to accurately localize them with only visuals

Figure 5: Late night scene with flashlight off and on

Design Choices

- Created for the Oculus Go, a wire-free standalone VR headset that allows players to look around without getting tangled
- Nudge players to rely more on sounds by dimming visibility as the night progresses
- Each enemy has a unique movement pattern and auditory signature to allow for non-visual differentiation
- Limited flashlight battery discourages constantly keeping it on
- Radio fills in player with back story as well as necessary information on how to survive
- Most creatures are at eye level (on the azimuthal plane) because elevation-varied audio localization is not strongly supported
- Flying creatures exist, but the players are not required to localize them to survive (i.e. they only serve as sources of confounding noise)

Testing Methodology

- Have subjects play a simplified localization game where they must locate the source of a spatialized sound
- The sound will be accompanied by a visual stimulus of randomized fidelity—that is, the audio is always spatially accurate, whereas the accompanying visual is not always
- Subjects are told ahead of time that the visuals are not always reliable, and are to play the test game before and after playing the main game

Hypothesis

- Subjects will achieve a higher score on the post-test than the pre-test
- Subjects who play an equally stimulating main game (but that does not train plasticity of multimodal integration) will score lower on the post-test than those who did

Figure 6: Looking up at the night sky

Future

- Implement a weather system that changes the sonic characteristics of the environment (e.g. spatialized pitter-patter of rain on objects in the scene)
- Personalized HRTF calibration to allow for greater use of elevation-varied spatial audio
- Separate into different levels that each have a neurological "theme"
- Incorporate other sensory modalities such as haptics (unavailable on the Oculus Go)
- Incorporate 6 DOF movement for both the player and controller

References

1. Ernst, M. O., & Knoblich, G. (2006). A Bayesian view on multimodal cue integration. Human body perception from the inside out, 131, 105-131.
2. Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. Journal of Vision, 7(5), 7-7.
3. Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. TRENDS in Neurosciences, 27(12), 712-719.
4. Magosso, E., Cuppini, C., & Ursino, M. (2012). A neural network model of ventriloquism effect and aftereffect. PLoS One, 7(8), e42503.
5. Trommershauser, J., Kording, K., & Landy, M. S. (Eds.). (2011). Sensory cue integration. Oxford University Press.