**VARIABLE LEGEND HYPER-PARAMETER LEGEND**

**=** timestep  **= =** degree of exploration vs exploitation

**=** selected action at time  **= =** budget for max iterations of algorithm

**=** estimated value of action at time  **= =** threshold must reach to converge

**=** uncertainty of action at time  **= =**  threshold used to increment

**=** number of times action has been selected

**=** reward generated by  **,** and

**=** binary of based on and  **=** user’s self-scored confidence in response

**=** state inferred by the user  **=** convergence counter

**=** real state of the robot  **=** difference in -value between and

**INITIALIZE** for all values in Q-value table # Uninformed or Informed

**INITIALIZE** # Time and convergence counter

**SET** hyper-parameter # Defined in hyper-param legend

**FOR** iterations in budget ***:*** # Number times we probe user

**INCREMENT**  # Increment time step counter

**CALCULATE** uncertainty for each action # Calculate all action uncertainties

**SELECT** action with max value  = # Select action based on argmax

**INCREMENT** # Increment action a counter

**EXECUTE** action # Present sound to user

**PROBE**  user for feedback and  # Ask users two questions

**CALCULATE** # Check if state inferred correctly

**CALCULATE** reward signal # Generate reward signal

**UPDATE**  # Update value for selected action

**CALCULATE** # Increment time step

**CALCULATE** # Adjust convergence counter

**IF :** # Check for convergence

**BREAK**

**END** loop