VARIABLE LEGEND

- t = timestep
- a_t = selected action a at time t
- $Q_t(a)$ = estimated value of action a at time t
- $U_t(a)$ = uncertainty of action a at time t
- $N_t(a)$ = number of times action a has been selected
- R = reward generated by S_{check} , S_{infer} and C
- S_{check} = binary of ± 1 based on S_{infer} and S_{real}
- S_{infer} = state inferred by the user
- S_{real} = real state of the robot
- C = user's self-scored confidence in response $\in [0, 10]$
- F = convergence counter
- ΔQ = difference in Q-value between $Q_t(a)$ and $Q_{t+1}(a)$

HYPER-PARAMETER LEGEND

- $z = 0.5$ = degree of exploration vs exploitation
- $B = 60$ = budget for max iterations of algorithm
- $F_{conv} = 3$ = threshold F must reach to converge
- $\Delta Q_{conv} = 2.0$ = ΔQ threshold used to increment F

INITIALIZE $Q_t(a)$ for all values in Q-value table — # Uninformed or Informed

INITIALIZE $t = 0$, $F = 0$ — # Time and convergence counter

SET hyper-parameter z , B , F_{conv} , ΔQ_{conv} — # Defined in hyper-param table

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

INCREMENT $t = t + 1$ — # Increment time step counter

CALCULATE uncertainty for each action $U_t(a) = z \cdot \sqrt{\frac{2log(t)}{N_t(a)}}$ — # Calculate all action uncertainties

SELECT action with max value $a_t = \arg\max_{a \in A} \left(Q_t(a) + U_t(a)\right)$ — # Select action based on argmax

INCREMENT $N_t(a) = N_t(a) + 1$ — # Increment action a counter

EXECUTE action a_t — # Present sound to user

PROBE user for feedback S_{infer} and C — # Ask users two questions

CALCULATE $S_{check} = \begin{cases} +1 \text{ if infers correct } (S_{infer} = S_{real}) \\ -1 \text{ if infers incorrect } (S_{infer} \neq S_{real}) \end{cases}$ — # Check if state inferred correctly

CALCULATE reward signal $R = S_{check} \cdot C$ — # Generate reward signal

UPDATE $Q_{t+1}(a) = \left[\left(1 - \frac{1}{N_t(a)}\right) \cdot Q_t(a)\right] + \left[\frac{1}{N_t(a)} \cdot R\right]$ — # Update value for selected action

CALCULATE $\Delta Q = Q_{t+1}(a) - Q_t(a)$ — # Increment time step

CALCULATE $F = \begin{cases} F + 1 \text{ if } (S_{infer} = S_{real}) \oplus (a_t = a_{t-1}) \\ F + 1 \text{ if } (S_{infer} = S_{real}) \oplus (\Delta Q = \Delta Q_{conv}) \\ 0 \quad \text{othewise} \end{cases}$ — # Adjust convergence counter

IF $F \geq F_{conv}$**:** — # Check for convergence

BREAK

END loop