

---

**Algorithm 1: Generation of Targeted Adversarial Audio Files with High SNR**

**Inputs:** Original benign example  $\mathbf{x}$   
target command **command**  
save path **save**  
compare target command **command\_out**

**Output:** Targeted attack example  $\mathbf{x}_{adv}$

```
logits  $\leftarrow$  getLogits( $\mathbf{x}$ )
/* Identify adversarial audio commands from a given model */
AE_Command  $\leftarrow$  model(logits)
Threshold = 0.2
/* Find the subsequence that matches the corresponding target at the current
threshold */
indexSets  $\leftarrow$  getIndexSet(Threshold, logits)
diterNum = 0
While Len(indexSets) < Len(command) do
    diterNum++
    indexSet  $\leftarrow$  getIndexSet(Threshold, logits)
    Threshold /= (1.1 + math.exp(-diterNum))
/* Find all subsequences under the current threshold */
indexSets  $\leftarrow$  getAllSubsequences(Threshold, logits)
maxIndexSet  $\leftarrow$  argmax(indexSets)
cl = 0
While AE_Command != command_out do
    for i  $\leftarrow$  1 to logits do
        if i belongs to maxIndexSet then
            /*Don't update characters that are already correctly recognized */
            if logits[maxIndexSet[c]] != command[cl] then
                loss_adv += CrossEntropyLoss()
                cl++
            end
        else
            if logits[i] != 0 then
                loss_adv += CrossEntropyLoss()
            end
        end
    end
    loss_ori = MSE( $\mathbf{x}_{adv}$ ,  $\mathbf{x}$ )
    loss_sum = (1 -  $\alpha$ ) * loss_adv +  $\alpha$  * loss_ori
    update  $\mathbf{x}_{adv}$  by Adam optimization
return  $\mathbf{x}_{adv}$ 
```

---