Beyond L_p clipping: Equalization-based Psychoacoustic Attacks against ASRs

Hadi Abdullah hadi10102@ufl.edu
Muhammad Sajidur Rahman rahmanm@ufl.edu
Christian Peeters cpeeters@ufl.edu
Cassidy Gibson c.gibson@ufl.edu
Washington Garcia w.garcia@ufl.edu
Vincent Bindschaedler vbindsch@cise.ufl.edu
Thomas Shrimpton teshrim@ufl.edu
Patrick Traynor traynor@ufl.edu
University of Florida

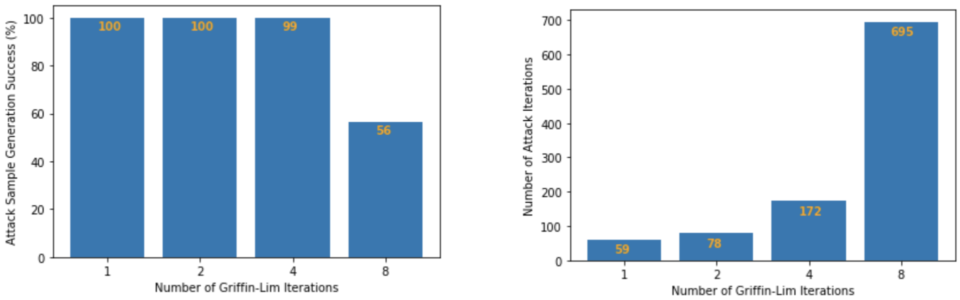
Appendix A. Supplementary Material

A.1. Attack Parameters

The attack success decreases if we increase the number of iterations beyond 2 (Figure 1(a)). In other use cases, the Griffin-Lim algorithm is executed for 50 iterations **?**. However, these number of iterations are too high for our use case, as we see a significant drop in accuracy (approximately 50%) at just 8 iterations (Figure 1(a)). Similarly, increasing the number of Griffin-Lim iterations also results in an increases the number of required iterations of our attack (Figure 1(b)), hence a longer required time to generate an attack audio sample. Clearly, increasing the number of Griffin-Lim iterations is unhelpful for the attacker. As a result, it is worth discussing why this is the case.

As discussed in Section 3.1, the Griffin-Lim algorithm provides an approximate reconstruction of time-domain audio, since a perfect reconstruction is not possible. Increasing the number of iterations results in the finer perturbations being discarded. This decreases the likelihood of the adversarial audio moving in the direction of the target transcription in the decision space.

Moreover, we found that increasing the number of Griffin-Lim iterations did not actually improve the quality of the adversarial audio samples. Manual listening tests by the authors revealed that the perceived quality of adversarial audio samples was consistent across the number of Griffin-Lim iterations. As a result, an attacker will benefit from setting the number of Griffin-Lim iterations to 1. This will allow for a high success rate of adversarial audio with fewer attack iterations, hence faster attack audio generation. Specifically, increasing the Griffin Lim algorithm from 1 to 8 will increase the total attack iterations (and consequently the compute time) by 11 times.



(a) Attack Success vs Griffin-Lim Iterations

(b) Attack Iterations vs Griffin-Lim Iterations

Figure 1: The plot shows the iterations of the Griffin-Lim algorithm that would be ideal for an attacker. (a) Increasing the Griffin-Lim iterations leads to a reduction in attack success. (b) Higher Griffin-Lim iterations requires a higher number of attack iterations to produce an attack audio sample. Manual listening tests showed that audio quality remained consistent across all Griffin-Lim iterations.

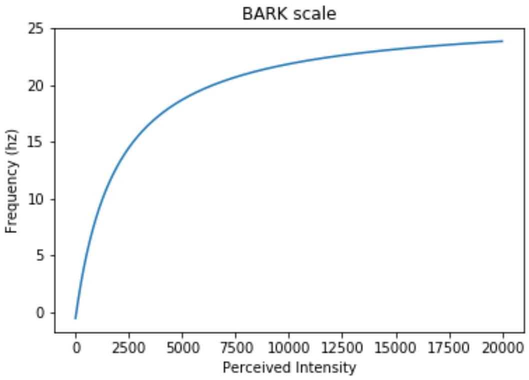


Figure 2: The BARK scale is a psycoacoustic model of the human perception of loudness in relation to frequency. The scale suggests that humans perceive high frequencies as louder than lower frequencies. The perceived intensity values associated with each frequency correspond to the first 24 critical bands of hearing.