Inaudible Voice Attack Cancellation

Zihui Qian

Introduction

- Voice Controllable Aystems(VCS)
- Non-linearity in Microphones
- Ultrasonic Attack
- Backdoor & DolphinAttack

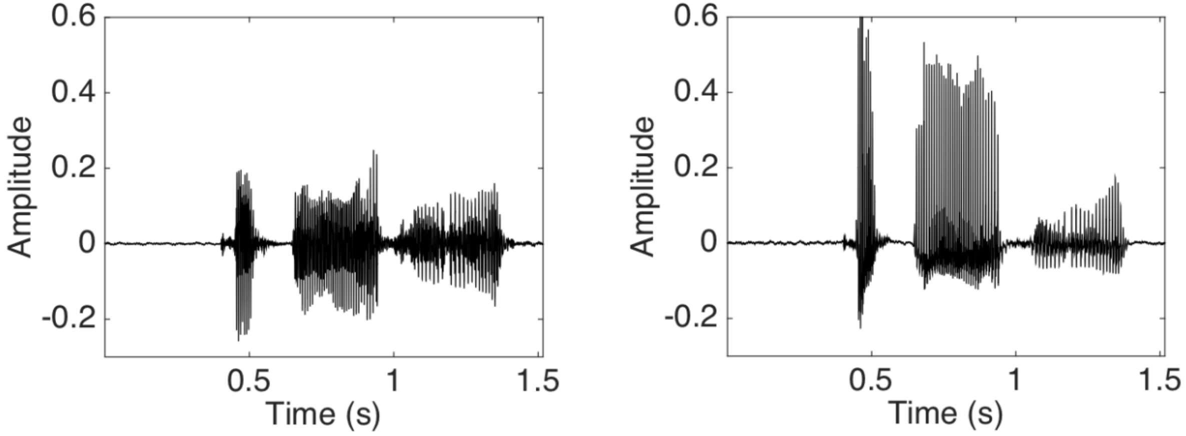
Background

$$\boxed{\begin{aligned} s_{\text{out}}(t) &= \sum_{i=1}^{\infty} A_i s^i(t) = A_1 s(t) + A_2 s^2(t) + A_3 s^3(t) + \dots \\ &\approx A_1 s(t) + A_2 s^2(t) \end{aligned}}$$

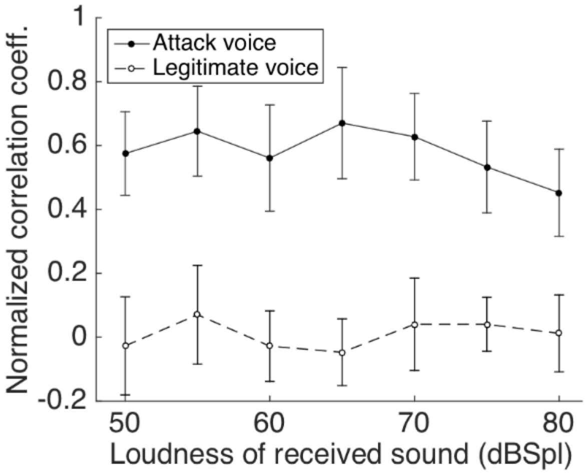
$$s(t) = (m(t) + \alpha) \times \sin(w_c t)$$

$$\begin{aligned} s_{\text{out}}(t) &= A_1 s(t) + A_2 s^2(t) \\ &= \underline{\frac{A_2}{2}\alpha^2 + A_2 \alpha m(t) + \frac{A_2}{2} m^2(t)} - \frac{A_2}{2}\alpha^2 \cos(2w_c t) \\ &- A_2 \alpha m(t) \cos(2w_c t) - \frac{A_2}{2} m^2(t) cos(2w_c t) \\ &+ A_1 m(t) \sin(w_c t) + A_1 \alpha \sin(w_c t) \end{aligned}$$

Previous Scheme



(a) (b)



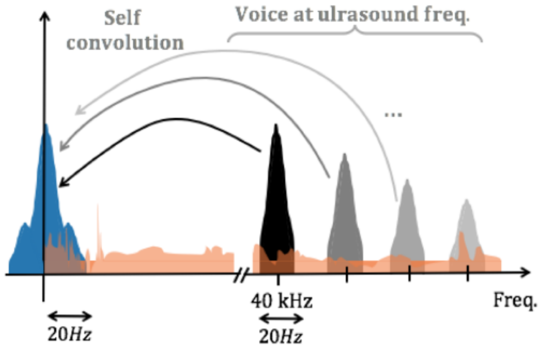
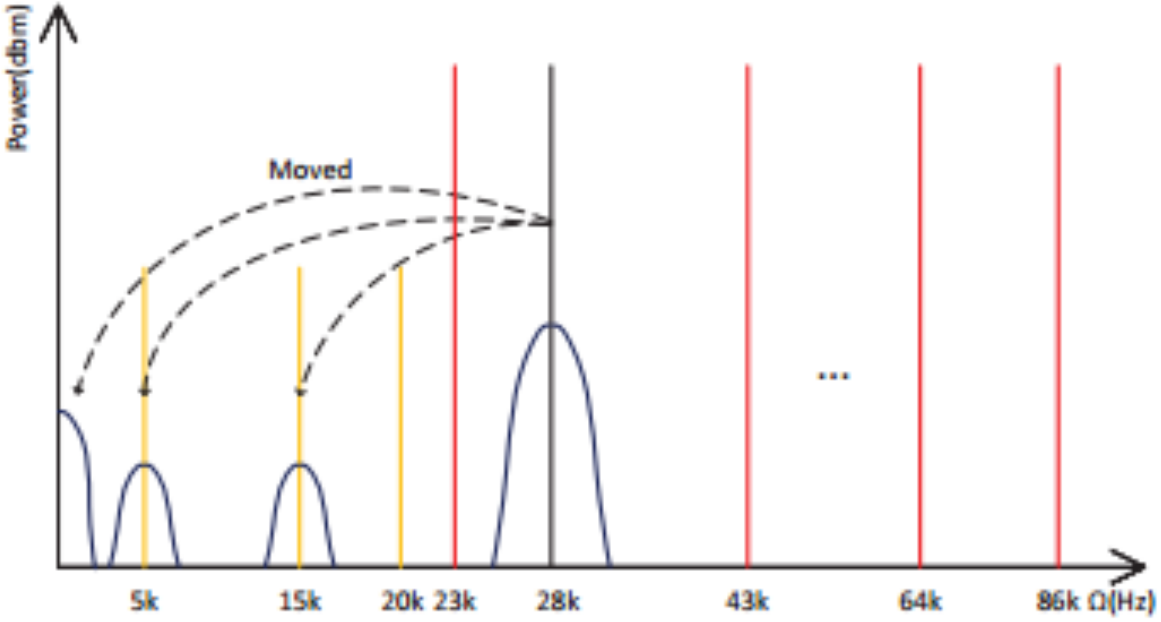
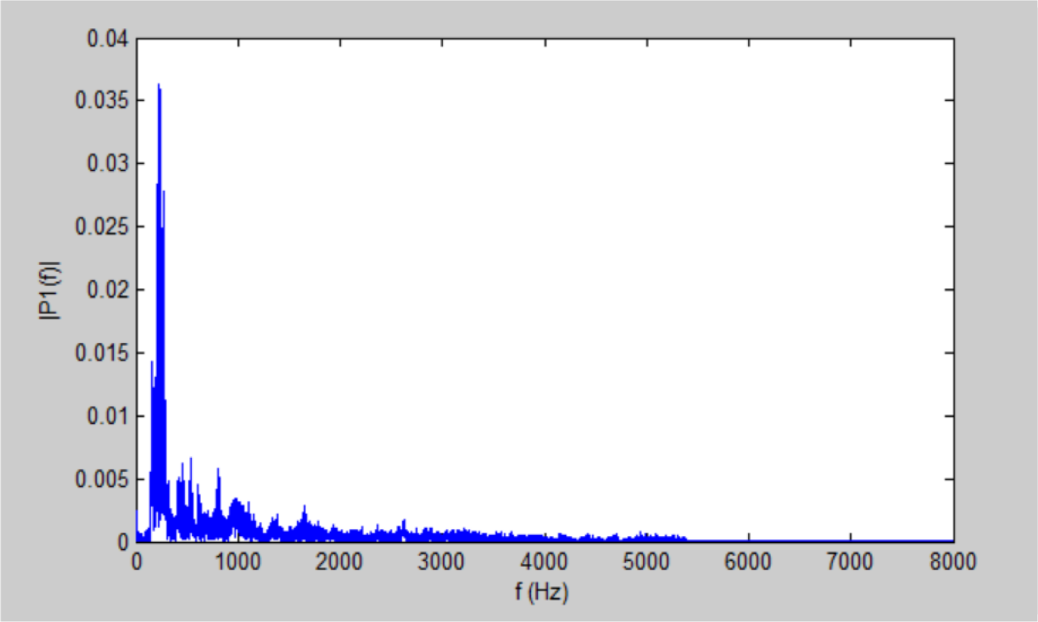


Figure 7: (a) A simplified voice spectrum showing the structure. (b) Voice spectra after non-linear attack.

Our Scheme



Evaluation


0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
|P1(f)|
0
1000
2000
3000
4000
5000
6000
7000
8000
f (Hz)

Q&A

Thanks