Dolphin Attack on Smart Home AI systems

There is a growing trend of utilizing smart home systems like Google Home, Amazon's Echo and the recently launched Apple's Home Pod for several purposes ranging from asking it to turn on the lights in the home to making orders from e-commerce websites. These smart home speakers have become an essential part of the user's life and have been controlled by smart Artificial Intelligent systems designed by Google, Apple and Amazon. They have a speech based system running atop their hardware which follow the speaker's commands. These speakers are generally activated by user's voice and engage in a two way conversation. There are innumerable benefits for these systems but they are exposed to lot of vulnerabilities too. The Dolphin attack is based on exploiting the higher inaudible frequencies of sound which can still trigger the smart home systems to perform any action, but without the user's knowledge. Guoming Zhang et.al performed Dolphin Attack on the AI speech assistants that come on board smart phones. They inject malicious inaudible voice commands into the smart phones and similarly we plan to attack the smart home system through the Text To Speech conversion method and modulating it on high frequency carrier wave. The problem of these commands being intelligible to the systems which are trained to respond to human tonal features is circumvented by amplitude modulating an attack command on an ultrasonic carrier. The problem of Low Pass Filtering to remove frequencies higher than the audible range is solved utilizing the nonlinear nature of amplifiers to produce demodulated signals in low frequency range and retrieve the command at the receiver end before it is passed through the low pass filter. There are two ways in which the assistant could work : speaker dependent and speaker independent. In case of Siri, there is an activation phase which is basically speaker dependent and once trained the recognition phase moves on to speaker independent algorithms. The activation phase in Apple's home assistant was hacked based on the fact that it could be unlocked by any voice with similar tonal features. Google Home Mini responds to anyone who says the keyword 'Ok Google' and so does Amazon's Echo which takes up the keyword 'Alexa' spoken by anyone and utilizes speaker independent algorithms. Our proposed defence involves comparing the differences between the original signal and demodulated signal in the high frequency range by means of Machine Learning Techniques.