Study of backdoor signal for beacon design

Haofan Lu
ZJU-UIUC Institute

Abstract
This report study the low-frequency shadow produced by backdoor signals. The goal of this research is to find backdoor signal that is applicable for the beacon designed of EAAR indoor localization system. The frequency band, duration of the signal, the power of transmitter, distance of receiver and the angular bias between the speaker and microphone are examined. Furthermore, the condition that mobile phone is in pocket is studied. In this study, I assume the ideal signal is known. Therefore, the cross-correlation between the received signal and ideal signal is used as the metric to decide if the beacon is identifiable. Besides, a criterion to decide whether a signal is detectable is proposed and applied to measure the detectability for certain conditions.

1 Experiment Setting

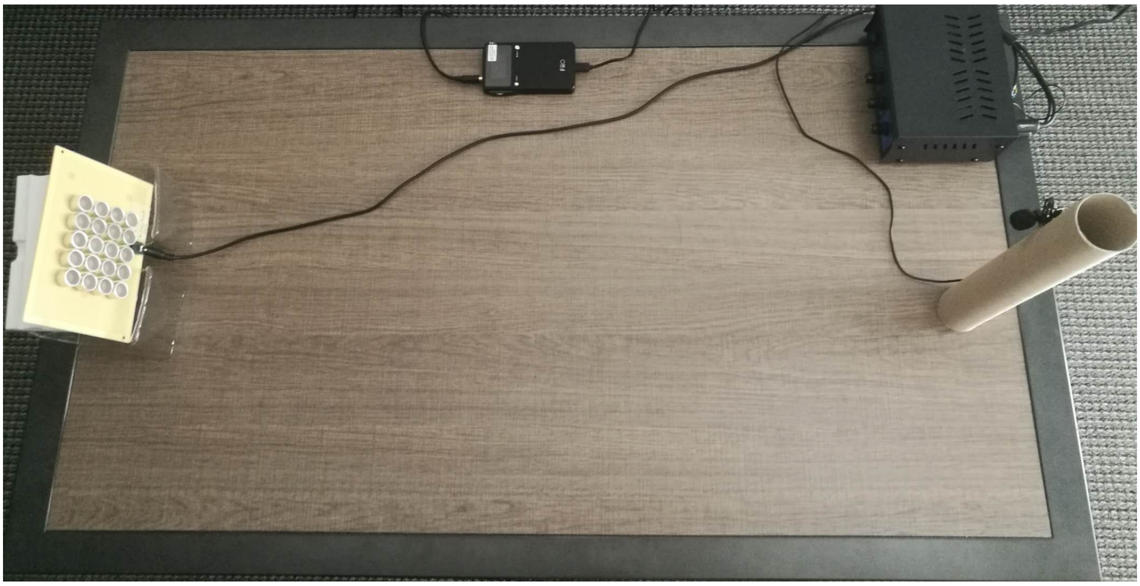

Figure 1: The experiment setting

The setting of the hardware is shown in figure 1. The primary components include ultrasound speaker array (left), microphone (on the paper tube), amplifier (right) and soundcard (middle). The distance between speaker array and microphone is 80cm in this setting (the effect of distance is studied in section 5).

The software is running on MATLAB, which produces the backdoor signal and record the sound through microphone for analysis. The backdoor signal consists of two carrier signals and the low frequency shadow is produced by the difference of carrier frequencies. In the experiment, I fixed one carrier frequency to be 40kHz and adjust the other carrier frequency. In the following discussion, frequency refers to the varying one.

To produce ultrasound signal, the soundcard is set to sample at 96kHz to avoid aliasing. The microphone is sampled at 48kHz, which is the upper bound of sample rate for most of the mobile phones.

Since the frequency of ultrasound is above the Nyquist rate, the aliasing on the microphone side should be taken good care of. The approach to deal with aliasing will be discussed in section 2.1.

In the experiment, I started with a large ultrasound signal domain and narrowed the scope step by step by eliminating unsuitable signals. The initial domain include,

- start frequency $\in \{25, 26, 27, ..., 38, 39, 40\}$,
- bandwidth $\in \{0, 1, ..., m\}$, where m is the maximum values, such that the sum of start frequency and bandwidth is below 40kHz,
- duration $\in \{0.1, 0.2, 0.4, 0.5, 0.6, 0.8, 1.0\}$

In the settings, there are two equipment affect the power of ultrasound: amplifier and soundcard. The amplifier is used for crude adjustment, while the soundcard serves for the fine tuning. According to the specification of the amplifier one level is 8W, the maximum power output is 80W. The soundcard can adjust the volume from 0 to 60. The concrete study of signal power will be elaborated in section 6

2 Bandwidth and Start Frequency

In this section, I present the cross-correlation with respect to start frequency and bandwidth of backdoor signals. The experiment is conducted under the condition:

- Amplifier power: 16W,
- Soundcard volume: 60,
- distance between speaker and microphone: 80cm,
- the speaker and microphone are confronted without angular bias and obstacle.
- the duration of the signal is kept as 0.5s

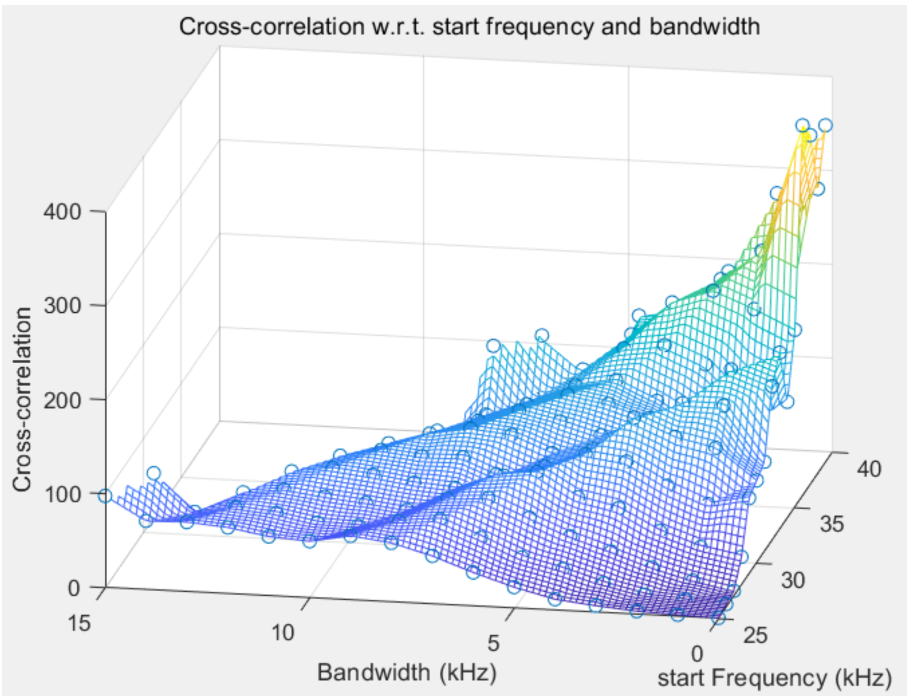

Figure 2: Plot of cross-correlation with respect to bandwidth and start frequency

From figure 2, the cross-correlation increases as start frequency increases. For lower frequencies, the cross-correlation increases as bandwidth increases. For higher frequencies, the cross-correlation does not vary much in terms of the variation of bandwidth. The effect of start frequency is more conspicuous. The explanations to the above observations are:

1. The non-linearity of the microphone is not prominent for lower frequencies. On the other hand, higher frequencies undergo greater non-linearity and the shadow signal produced at lower frequencies are more identifiable.
2. For lower frequencies, longer bandwidth helps the signal reach high-frequency band, where is signal is more identifiable. Besides, longer bandwidth provide more information to identify the signal.
3. For higher frequencies, the signal can reach relatively high cross-correlation. Increasing bandwidth does not help enhance the identifiability of the signal .

2.1 Aliasing

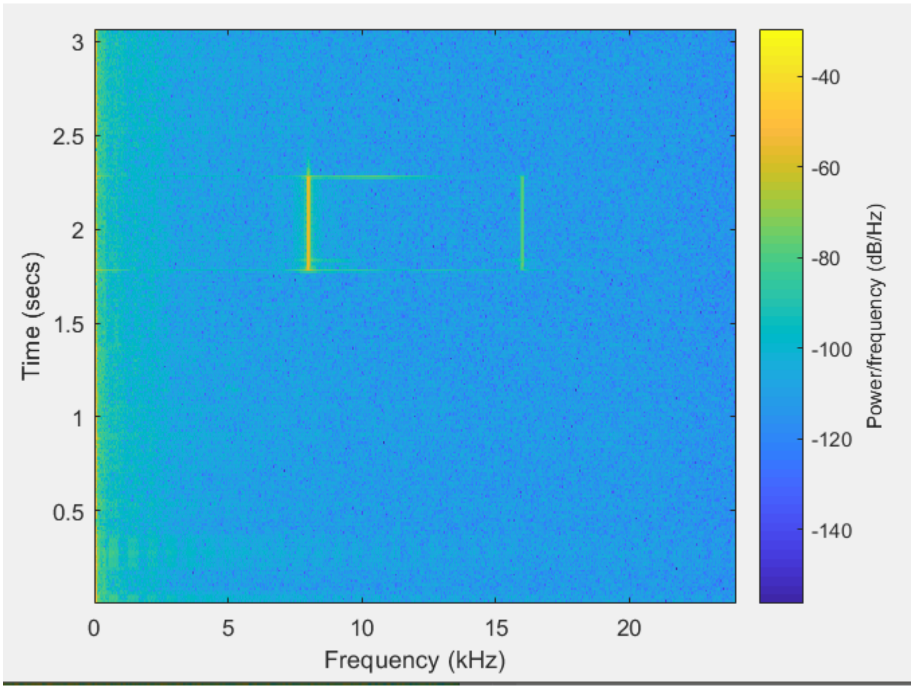

Figure 3: Aliasing in 32kHz carrier frequency

Another observation derives from the spectrogram of received signal. Figure 3 is the spectrogram of the received signal where the carrier frequency is 32kHz and 40kHz. As is shown by the figure, the aliasing of high-frequency carrier signals (40kHz to 4kHz and 32kHz to 16kHz) can overlap with the shadow to affect the evaluation.

For lower carrier frequencies, the shadow signal produced by non-linearity might be affected by the aliasing produced by carrier signals. It is hard to filter it out if the microphone does not have a built-in Low-pass filter applied before sampling. One way to avoid the effect of aliasing is to adopt high-frequency signal (above 32 KHz), so as to keep the shadow frequency in lower band (below 8kHz), while the aliasing signals will stay above 8kHz. Then a low pass filter with cutoff frequency 8KHz can be applied to the sampled signal to filter out the aliasing signals.

The low pass filter can also be replaced by a bandpass filter with passband from 1kHz to 8kHz to filter out some baseband noise.

Through this step, the domain of start frequency is shrunken to $\{33, 34, 35, 36, 37, 38, 39, 40\}$. The bandwidth domain is temporally remain the same.

3 Duration

Theoretically, the cross-correlation increases with the duration of the signal, since there are more samples to be correlated, which contribute to the total cross-correlation. The experiment verifies this assumption, as is shown below.

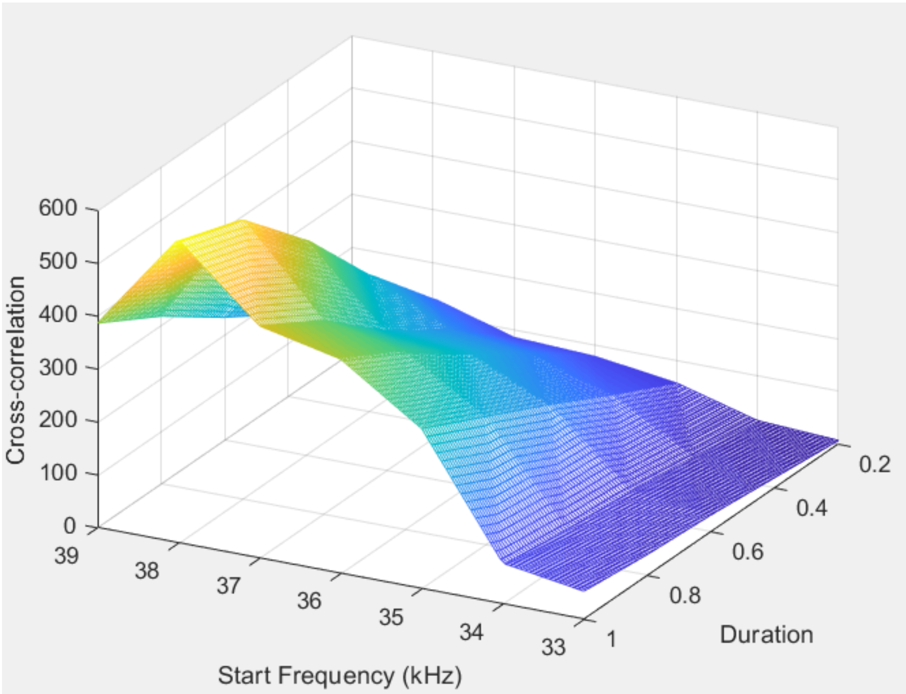

Figure 4: Plot of cross-correlation with respect to start frequency and duration

Figure 4 is plotted under the condition

- Amplifier power: 8W,
- Soundcard volume: 60,
- distance between speaker and microphone: 80 cm
- the speaker and microphone are confronted without bias angle and obstacle.

As the above section has shown, the cross-correlation does not vary a lot across the bandwidth domain. Therefore, start frequencies is varying across the current domain, while the bandwidth is kept to 0 (i.e. a single tone). The duration increases from 0.2s to 1.0s. The result is consistent with the theoretical reasoning. The cross-correlation increases as duration increases.

However, there's a trade-off that worth mentioning. In the real application scene, the users are moving constantly. The time they pass the beacon is limited. Therefore, the duration of the signal cannot be too long, otherwise, the mobile phone or earphone cannot receive the signal completely with a stable amplitude, which increase the difficulty of cross-correlation.

4 Angular Bias

Since ultrasound is highly directional compared to audio signal in baseband, the signal strength is strongest when the speaker and microphone are confronted. Theoretically, the cross-correlation curve should be a convex when the user pass before the beacon. The cross-correlation reaches its maximum when the user is directly in front of the beacon. The experiment verifies the phenomenon. In this section, the angle between the normal direction and the direct channel path is referred to as angular bias. The angular bias is sampled to increase linearly.

The experiment is conducted under the condition:

- Amplifier power: 8W,
- Soundcard volume: 60,
- distance between speaker and microphone: 80 cm
- The angular bias $\in \{0, 15, 30, 45, 60, 75, 90\}$ in degree.

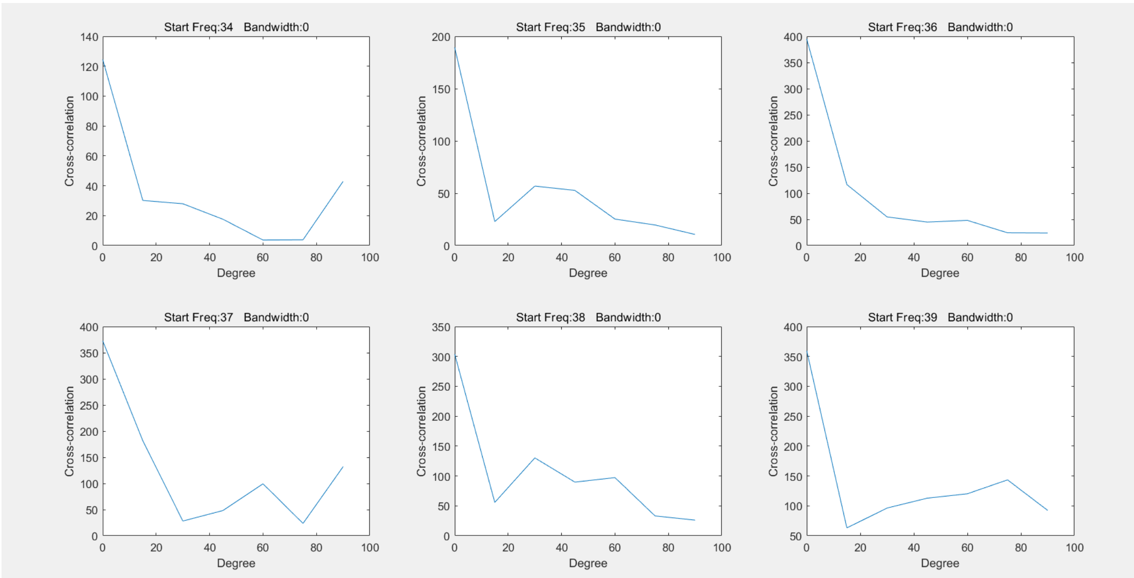

Figure 5: Cross-correlation with respect to angular bias for signal tone

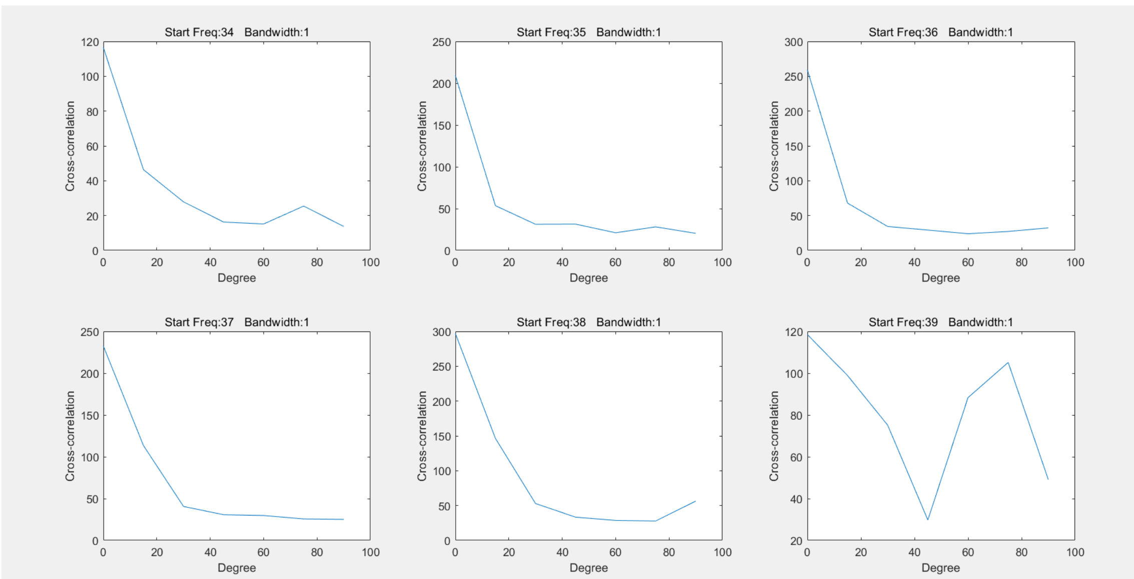

Figure 6: Cross-correlation with respect to angular bias for chirp with bandwidth 2kHz

Figure 5 and 6 show the variation of cross-correlation with respect to the angular bias in degree. Figure 5 shows the case of signal tone. As we can see from the plots, in general, the cross-correlation decreases as angular bias increases, as expected. However, the trend is not monotonous in most of the subplots in figure 5. In other words, the sometime, the cross-correlation is greater for larger angular bias compared to the smaller ones. One explanation to this phenomenon is the interference of the source signal. Since the source signal only has one frequency component, it is easier to be affected by the interference.

On the other hand, the chirp signal does not have such problem. As figure 6 shows, in almost all the subplots, the cross-correlation decrease monotonously, with little exception. However, the last subplot is wired. there's a unexpected peak at 60^o and 75^o. The explanation to the exception is that the shadow signal reaches the band below 1kHz in this case. There might be noises in this band that can correlate with the signal to produce this abnormally great cross-correlation.

To find out what happened when the shadow signal is below 1kHz, I plot the cross-correlation in this case, shown in figure 7.

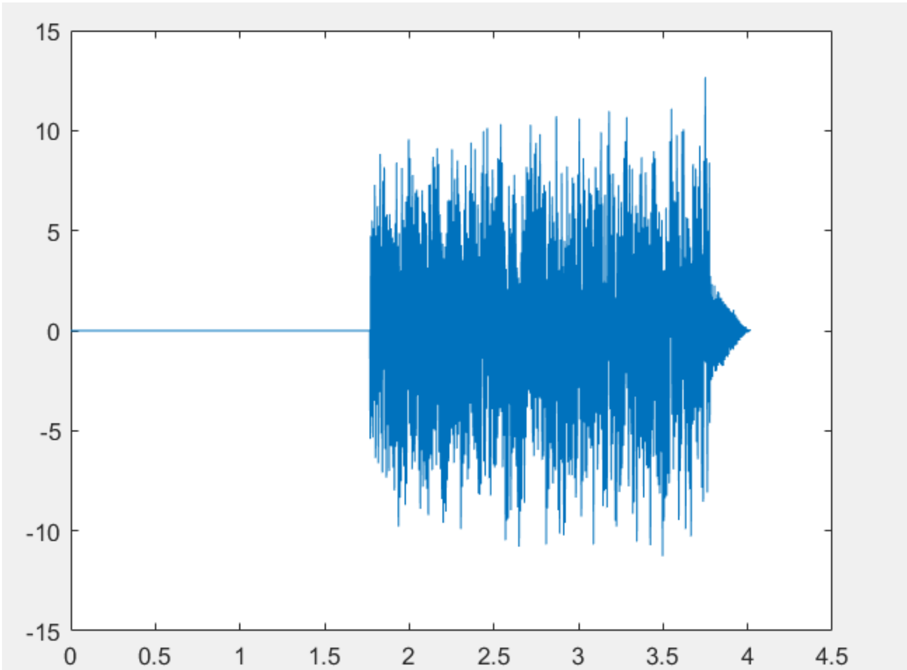

Figure 7: Plot of cross-correlation when the chirp reaches 40KHz

In this plot, the shape of cross-correlation curve is noisy, which implies that the signal is not being found. It probably due to the environment noise below 1kHz, which lowers the SNR. On the contrary, if the carrier signal does not reach 40kHz, the shape of cross-correlation plot is quite clear, as shown in figure 8 and 9. There's a prominent peak in these plots, which indicates where the signals overlap.

Given there's a perceivable pattern in the cross-correlation curve if the signal is identifiable, it is possible to train the machine to identify this signal by analyzing the cross-correlation data. I will discuss this research further in section 7.

To conclude this section, I list some restrictions added to the signal domain.

– The shadow band should not be below 1kHz; therefore, the maximum bandwidth is limited such that the sum of start frequency and bandwidth is below 39kHz.
– The chirp signal is preferred to single tone, since we want the cross-correlation decrease monotonously as angular bias increases.

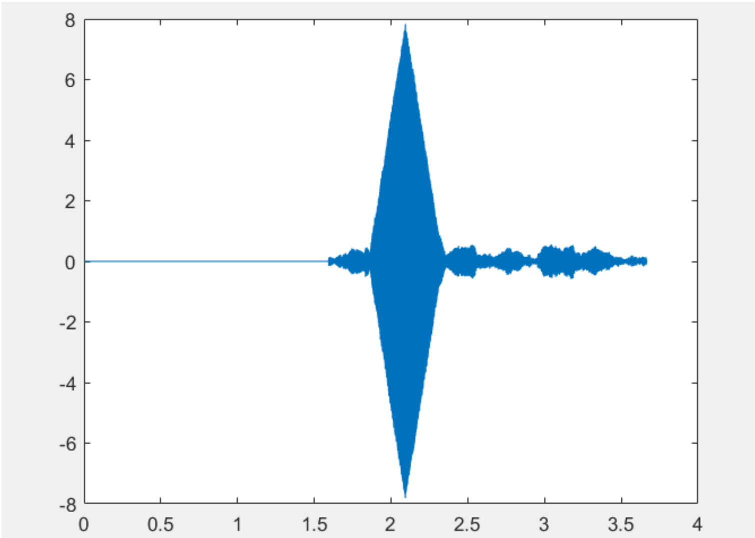

Figure 8: Plot of cross-correlation of signal tone signal

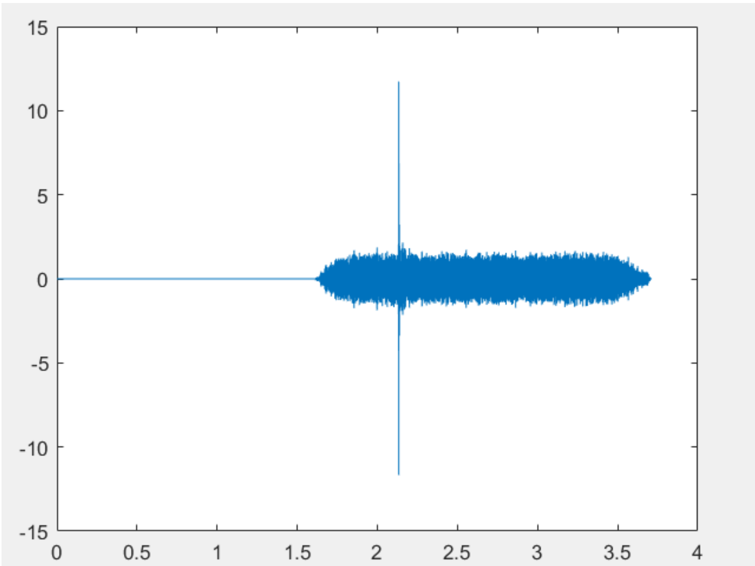

Figure 9: Plot of cross-correlation of chirp signal

5 Distance

Intuitively, the cross-correlation decreases as distance between the speaker and microphone increases, because the ultrasound wave attenuates in air. The experiment verifies this assumption. The experiment was done under the following conditions:

– Amplifier power: 8W
– Soundcard volume: 50
– distance samples $\in \{20, 80, 140, 180\}cm$.
– the speaker and microphone are confronted without angular bias and obstacle.

As figure 10 and 11 shows, generally, the cross-correlation decreases as distance increases. This trend is unexpectedly non-monotonous, which probably dues to the multipath effect in the environment.

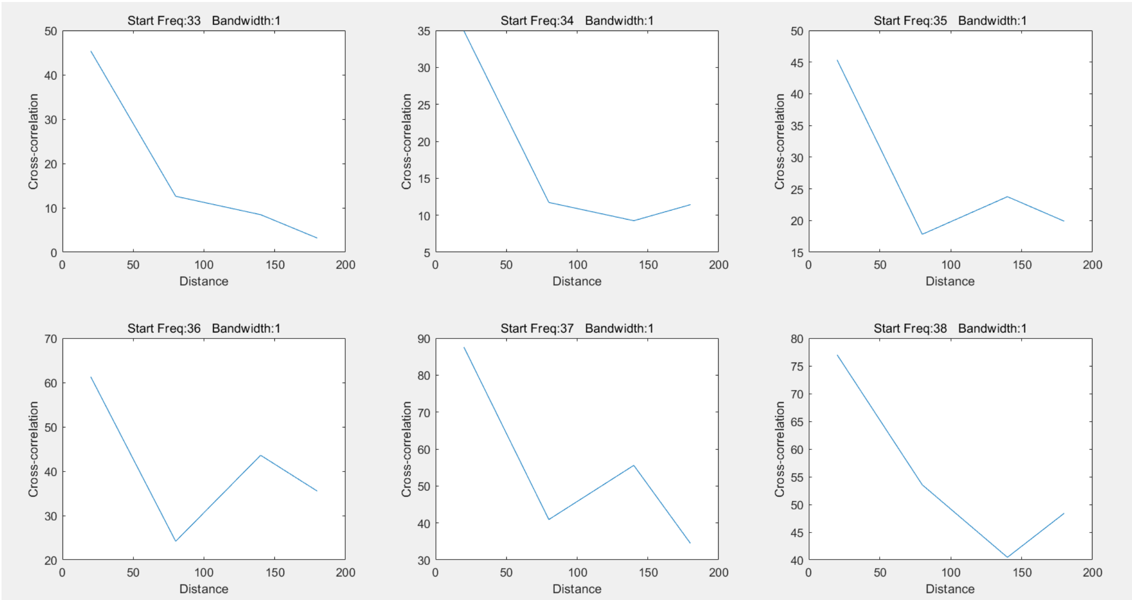

Figure 10: Cross-correlation with respect to distance (bandwidth = 1)

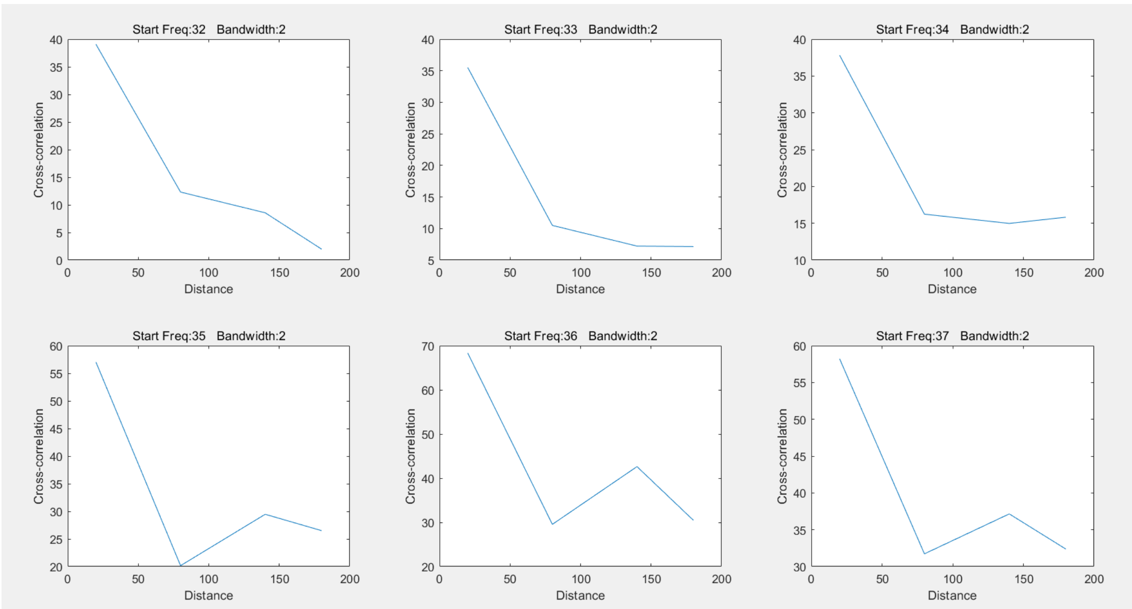

Figure 11: Cross-correlation with respect to distance (bandwidth = 2)

6 Power

It is worthwhile to discuss the power trade-off and restriction before moving forward. In this section, I will mainly present the trade-off regarding the power of transmission. The results of experiment will be presented in section 7 with the case of microphone in pocket. Generally, we believe that ultrasound with greater power is better because the signal can propagate farther and transmit through the clothes. However, there's actually a trade-off on the power of transmission sound.

1. The power cannot be too low, otherwise it cannot propagate too far or transmit though clothes. If the user put his mobile phone in the pocket, we hope our signal has enough power to transmit through the pocket to be heard by the microphones on the phone.

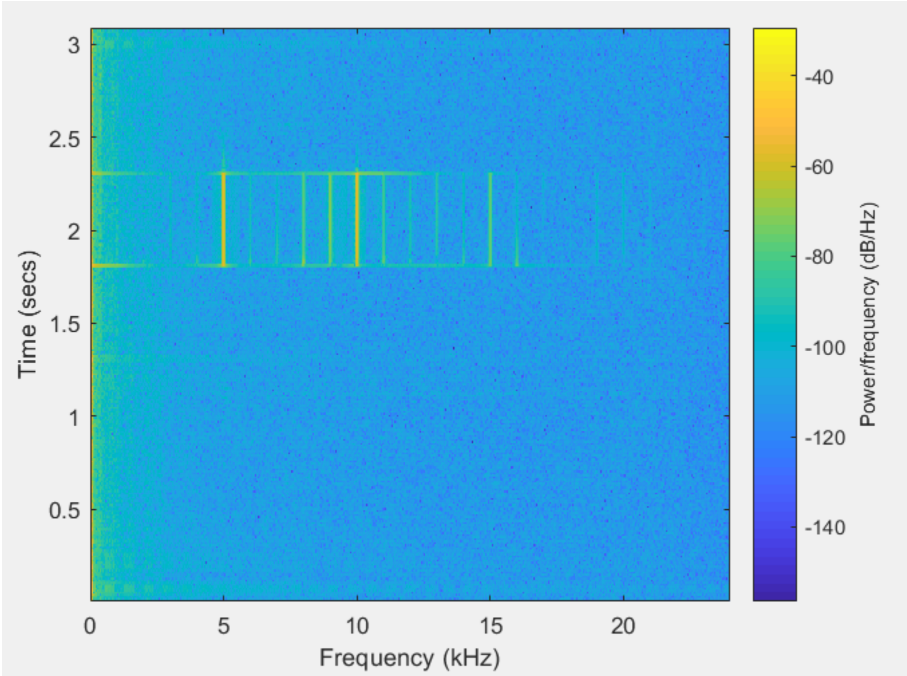

Figure 12: The spectrogram shows the harmonic produced when the power of amplifier is 32W

2. The power cannot be too high, otherwise the sound become audible (might be related to the properties of the speaker). Several harmonic frequencies can be detected in the spectrogram.

Figure 12 shows the the spectrogram of the received signal when the amplifier output is 32W and soundcard volume is set to 60. The transmitted signal is a single tone with carrier frequency 33kHz and 40kHz. There are several unexpected frequency components shown in the figure, which is caused by the overwhelmed amplitude. We do not want these frequency component to affect the evaluation. Therefore, the power of transmission should be limited.

Moreover, when the power is high, it will decrease the life time of speaker as well as produce some unexpected audible sound. Based on my perception, the sound is inaudible when the the power of amplifier is below 8W and the volume of sound card is below 40.

Therefore, the optimal power output is around this point. Below this point, the sound is totally inaudible but also too weak to be detected in 80cm with microphone in the pocket, while above this point, the sound is becoming audible.

However, this experiment was done in a relative quiet environment. The power is allowed to be greater in a more noisy environment to ensure the signal is detectable, while keeping inaudible to human.

Also, based on my experience, the sound is more audible when the frequency band is higher, if the power is kept constant. In other words, there's another trade-off between the power and frequency band. If we keep the frequency band relatively lower (e.g 33 $\sim$ 36kHz) can increase the power of signal, while still keeping the sound inaudible.

7 Microphone in pocket

To study the condition where the microphone is in the pocket, we need to quantify the detectability of the signal. In figure 7, 8 and 9, we can tell whether the signal is detectable by looking at the shape of the cross-correlation curve. We assigned figure 8 and 9 as detectable, because the peak is prominent. This identification approach implies that we can compare the maximum value with the average value in cross-correlation plot to let the machine decide whether the signal is detectable. The formal definition of this decision process is shown below:

$$r = \frac{max(xcorr)}{E[xcorr]} \tag{1}$$

where $xcorr$ denotes the cross-correlation array. $E[\cdot]$ denotes the mean of the cross-correlation array. r is the ratio of maximum value and average value of the cross-correlation array. Then I define a threshold T. When $r > T$, the signal is assigned detectable; otherwise, undetectable.

I finally select $T = 5$, which achieves accuracy of 95% (40/42), when I test at a tricky power level (Amplifier: 8W, soundcard: 35). The correct assignment is decided by myself.

With this quantified metric defined, I took a step further and define the detectability as the proportion of detectable cases in a standard test set. The standard test set consist of the signals in our signal domain.

volume	30	35	40	45	50
out-of-pocket	21.4%	82.1%	100%	100%	100%
in-pocket	14.3%	46.4%	100%	100%	100%

Table 1: Detectability of different soundcard volume

The experiment is conducted under the condition:

- Amplifier power: 8w
- distance between speaker and microphone is 80cm
- In-pocket and out-of-pocket cases are both tested.
- The duration of the signal is fixed at 0.5

The table 1 shows the results. The in-pocket condition does attenuate the signal and lower the detectability. However, when the volume is above 40, the sound is still detectable in pocket. The results of finer measurement of the feasible volume range of the sound card is presented in Appendix.

Moreover, by analyzing the distribution of detectable condition, two observations are concluded.

1. signals on higher frequency band is more likely to be detected.
2. single tone is more likely to be detected than chirp.

8 Mobile phone

Figure 13 and 14 shows the cross-correlation of the received signal when I took the phone, walking pass the speaker array. Figure 13 shows the case where the transmitted signal is a single tone with carrier frequency 36kHz and duration 0.2s. It is identifiable that the height of peaks first increase then decrease. This result is expected. Figure 14 shows the case where the transimitted signal is a chirp starts from 36kHz, ends with 37kHz with duration 0.2s. The central peak is prominent, while the other peaks with non-zero angular bias are not apparent.

9 Conclusion

To conclude, the frequency band of 33kHz $\sim$ 39kHz is recommended for the carrier frequencies. Compared to single tone, chirp is more stable with respect to the angular bias. The power output is recommended to be around 8W and volume 40 to ensure the sound to be inaudible, while retaining the detectability.

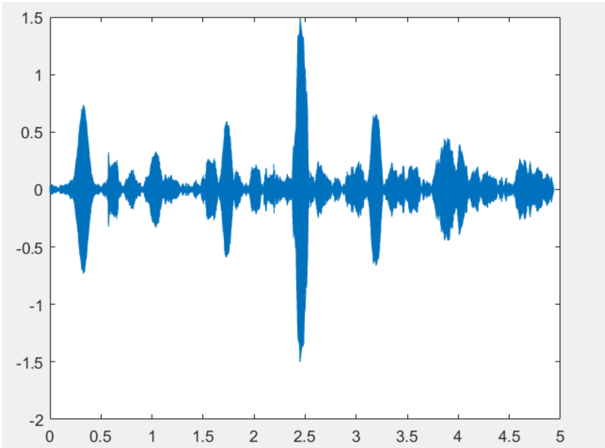

Figure 13: Cross-correlation plot of received signal by mobile phone (tone)

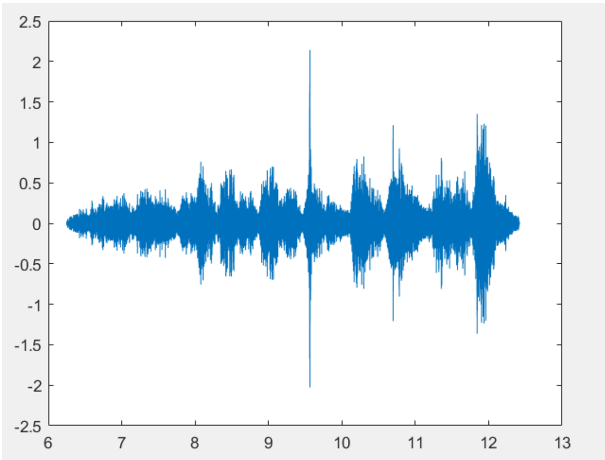

Figure 14: Cross-correlation plot of received signal by mobile phone (chirp)

Acknowledgements

I wish to thank Yang, Zhijian for enlightening comments on this report.

Appendix

Table .1 and Figure .1 show the ranges of sound card volume that satisfy the requirements. To be specific, the requirements are:

1. the signal power should be strong enough to be detected in the pocket, which determines the lower boundary of volume range.
2. The sound should not be perceptible by human, which determines the upper bound of the volume range.

The result is obtained when the distance between speaker and microphone is 80cm, the speaker and microphone are confronted with no angular bias. Also the microphone is in the pocket during the measurement.

The volume range presented here can only be used as a reference, the condition may change with different hardware and environment. If the environment is relatively noisy, the upper bound should be promoted.

Start Frequency (kHz)	Bandwidth (kHz)	End Frequency (kHz)	Duration (s)	Feasible Volume Range
33	0	33	0.2	[35, 44]
34	0	34	0.2	[35, 44]
35	0	35	0.2	[36, 43]
36	0	36	0.2	[36, 38]
37	0	37	0.2	{38}
38	0	38	0.2	[30, 37]
39	0	39	0.2	[31, 39]
33	1	34	0.2	[38, 41]
34	1	35	0.2	[37, 42]
35	1	36	0.2	[39, 41]
36	1	37	0.2	[40, 42]
37	1	38	0.2	[37, 39]
38	1	39	0.2	[34, 39]
33	2	35	0.2	[37, 41]
34	2	36	0.2	[38, 39]
35	2	37	0.2	{39}
36	2	38	0.2	[40, 43]
37	2	39	0.2	[37, 38]
33	3	36	0.2	[41, 43]
34	3	37	0.2	{43}
35	3	38	0.2	[40, 44]
36	3	39	0.2	[41, 43]
33	4	37	0.2	$\emptyset$
34	4	38	0.2	$\emptyset$
35	4	39	0.2	[38, 39]
33	5	38	0.2	$\emptyset$
34	5	39	0.2	$\emptyset$
33	6	39	0.2	$\emptyset$

Table .1: List of recommended signals with feasible volume range

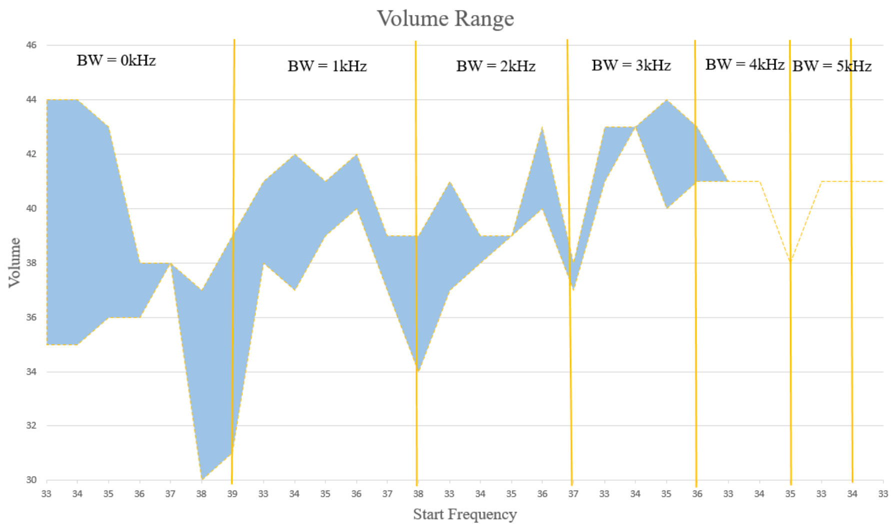

Figure .1: Recommended Volume Range for Different Start Frequency and Bandwidth