Getting Information About The Room From Indoor Echo

Puqing Yang

Shanghaitech Universiy
yangpq@shanghaitech.edu.cn

January 21, 2018

A more precise acoustic response model for indoor sound is established. It has considered some not existed mirror point and the phase of acoustic wave. Through this model, you can get any order echo you want. This article explores the first echo and second order echo response when the input is sinusoidal signal. Also, we have done the experiments to test this model, and use this model to analysis predict experiment result, and hope to find the relationship between the room size and the echo response spectrum. From the result, it remains us that the echoes of indoor are very complex and the first order echo usually mixes with the second order echo. But we can use ML estimator to estimate the arrive-of-time of the maximum first order echo and second order echo.

1 Introdution

Begin Idea

From the signal process theory, we have known that the impulse response reveal information about the system. Similarly, the room impulse response also can be used to getting some useful information about the room characteristic.

Main Problem

Find a method that we could know the room size from the received acoustic.

Challenging

1. The multi-path effected
2. Noise
3. Distortion

2 Theory analysis

2.1 Physic model

In this artic, we need to research a method to estimate the information about a class room. So, a model of acoustic propagation in a room should be established to solve this problem. The room size is (L_x, L_y, L_z), The coordinates of speaker is $S = (s_1, s_2, s_3)$, the coordinates of receiver is R=(r_1, r_2, r_3). To simplify this model, we ignore the direction of sound source vibration. The speaker can produce acoustic wave, the acoustic wave can be described as

$$s(t) = u(t - t_0)ReA\sum_{\omega} e^{j[\omega(t-t_0)+\phi]} \tag{1}$$

The acoustic wave that receiver received in this room can be described as

$$\begin{aligned} r(t, r) &= \alpha_0 s(t) * \delta(t - t_0 - r_0/c) \\ &\quad + \sum_k \alpha_{1k} e^{-j\pi}(s(t) * \delta(t - t_0 - r_{1k}/c)) \\ &+ \sum_k \alpha_{2k} e^{-j2\pi}(s(t) * \delta(t - t_0 - r_{2k}/c)) + \cdots \end{aligned} \tag{2}$$

The first part of the equation is the source acoustic wave directly propagate to the receiver, and there is no reflection from any wall. The second part is the received wave with only one reflection from the walls. The third part is the received wave with two reflections from the walls. From the first part of $r(t, r)$, the $\alpha_0$is the decay factor of the acoustic propagation in the air. The propagation decay(D_p) and the absorption decay(D_a) from the air are considered here. For the second and third part, the reflected decay(Dr) should be included. Due to the reflected decay is very fussy, and it depends on material and contracture of surface, we assume the reflected decay is a constant like the reflection coefficient in optics.

$$\begin{aligned} log\alpha_0 =& D_p + D_a \\ =& -20log(r/r_s) - 7.4f^2 r/\zeta \times 10^{-8} \end{aligned} \quad (3)$$

$$\alpha_{1k} = \alpha_0 \bullet (1 - (1 - \alpha_{wk})u(t - t_0 - r_{wk}/c)) \quad (4)$$

$$\begin{aligned} \alpha_{2k} =& \alpha_0 \bullet (1 - (1 - \alpha_{wk1})u(t - t_0 - r_{wk2}/c)) \\ & \bullet (1 - (1 - \alpha_{wk2})u(t - t_0 - r_{wk2}/c)) \end{aligned} \quad (5)$$

For reflection model, we use Allen and Berkley's Image Method. Figure1 and figure2 respectively show one-reflection situation and two-reflection situation. But the author didn't considerate that some reflection

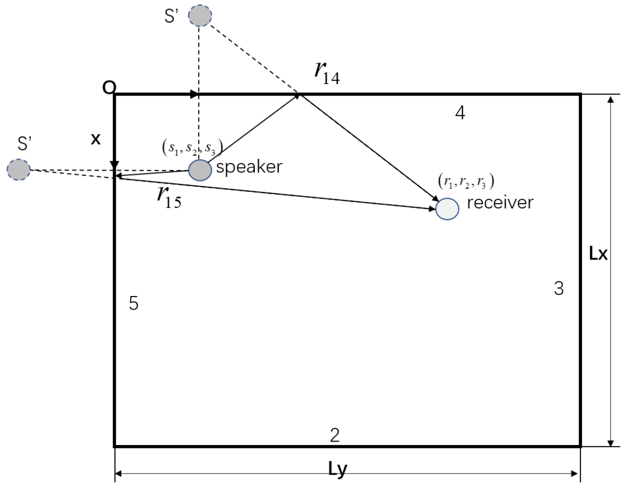

Figure 1: *First order echo path*

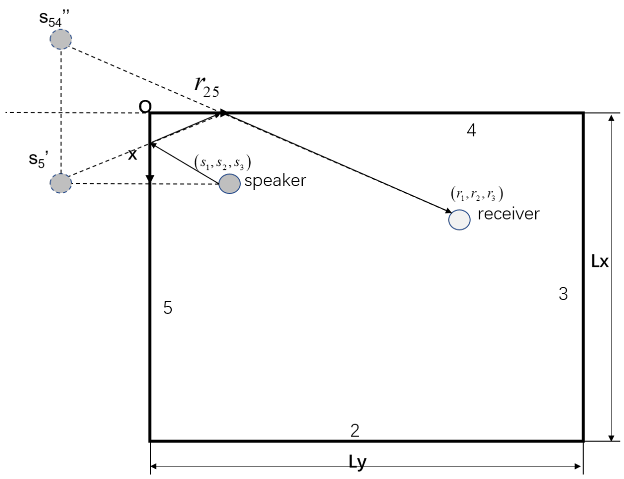

Figure 2: *Second order echo path*

not exit, picture3 gives an example This picture shows some mirror point couldn't transfer acoustic wave to the receiver. So, we need to exclude such mirror point. One of observation is that useful second-mirror point should satisfy that the connection between the second-mirror point and the receiving point is intersected with

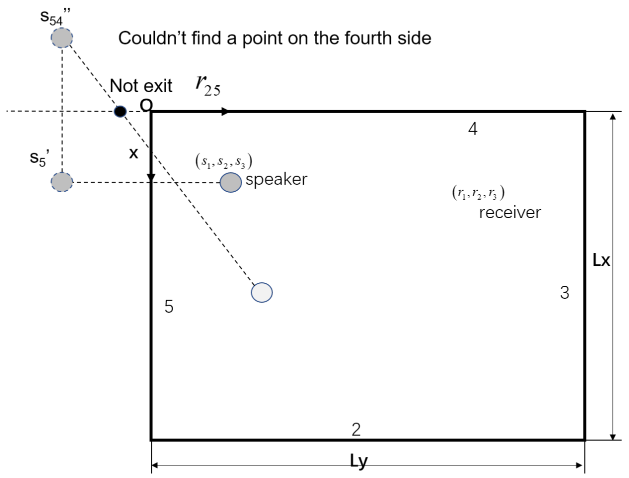

Figure 3: *The second order echo not exit*

Table 1: *the parameter of the experiment*

Object	Coordinates	Value
Room	(L_x, L_y, L_z)	$(13.40, 7.46, 2.81)$
Speaker	(s_1, s_2, s_3)	$(0.44, 0.66, 0.70)$
Receiver	(r_1, r_2, r_3)	$(3.18, 3.47, 0.75)$

the corresponding symmetric surface. It can also be extended to the third-mirror point. In three dimensions, a speaker and a receiver were put in a cuboid-like room, the key data show in the table 1

2.2 Model simulation

A simple sine-wave signal will applicate to this model, which is shown in figure5.

From physic model, we should know exactly which reflection is exit and the coordinates of reflection point on the wall. The flowing show this clearly. Each order refection has its own reflection feature, and we call it transfer matrix.

First-order-reflection

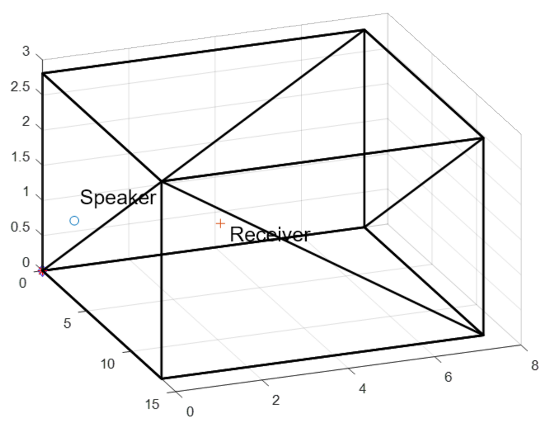

Figure 4: *The stucture of room*

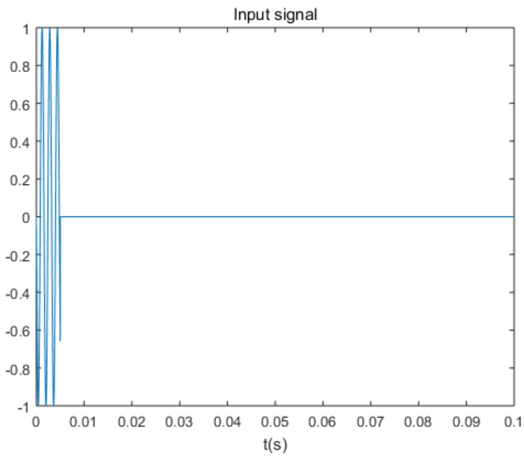

Figure 5: *Input signal*

Transfer matrix $T = [111111]$, it means that each six surfaces can reflection the acoustic wave to the receiver. All the mirror point satisfied first-order-reflection are shown in figure 6.

Second-order-reflection

Transfer matrix

$$T = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} \tag{6}$$

It means that two reflection surfaces satisfied the transfer condition is (1,2), (1,3), (1,6), (2,3), (2,4), (3,5) . . . All the mirror point satisfied the second-order-reflection are shown in figure 7.

Receive signal

We have used the physic model described in the begin and simulation it in MATLAB. After simulating, the receive signal contains directly receiving signal, first-order echo and second order echo.

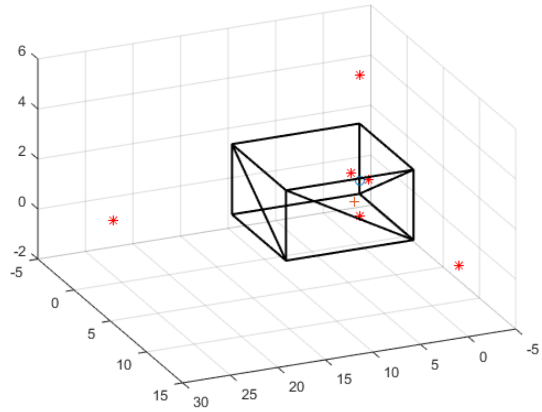

Figure 6: *First order mirror point*

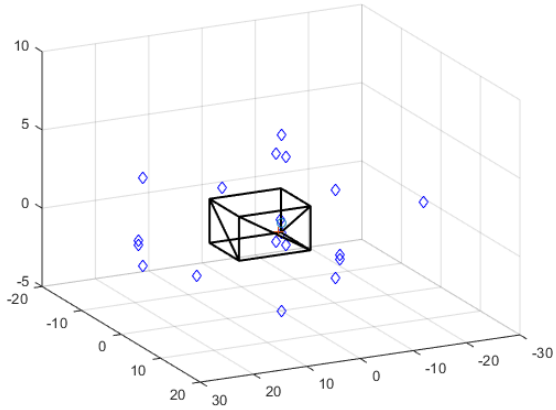

Figure 7: *Second order mirror point*

2.3 AoT estimation

Arrive-of-time estimation. Our target is get time information from the receiving signal. There need to estimate the AoT of first-order echo and second order echo. If the mixed-reflection is neglect, and if $s(t)$ is the transmitted signal, a simple model for the received continuous waveform is

$$x(t) = s(t - \tau_0) + \omega(t) \qquad 0 \le t \le T \tag{7}$$

In general, the model is a discrete time model. Letting $x[n]$ and $\omega[n]$ be the sampled sequences, the model is described as below.

$$x[n] = s(n\Delta - \tau_0) + \omega[n] \qquad n = 0, 1, \cdots, N-1 \tag{8}$$

Assuming $\omega[n] \sim N(0, 1)$. Considering the signal is nonzero only over the interval $\tau_0 \le t \le \tau_0 + T_s$, so 8

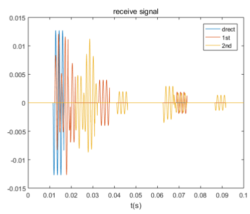

Figure 8: *The directly received signal, the first-order echo and the second-order echo*

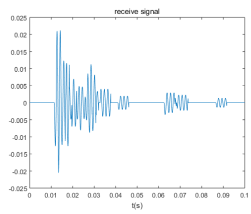

Figure 9: *total received signal*

reduces to

$$x[n] = \begin{cases} \omega[n] & \text{if } 0 \le n \le n_0 - 1, \\ s(n\Delta - \tau_0) + \omega[n] & \text{if } n_0 \le n \le n_0 + M - 1, \\ \omega[n] & \text{if } n_0 + M \le n \le N - 1. \end{cases} \quad (9)$$

Where M is the length of sampled signal and $n_0 = \tau_0/\Delta$ is the delay in samples. With this formulation we evaluate the CRLB. From the book have known

$$var(\hat{\tau_0}) \ge \frac{1}{\frac{\varepsilon}{N_0/2}\bar{F^2}} \quad (10)$$

Where

$$\varepsilon = \int_0^{T_s} s^2(t)dt$$

$$\bar{F^2} = \frac{\int_0^{T_s} (\frac{ds(t)}{dt})^2 dt}{\int_0^{T_s} s^2(t)dt}$$

$\bar{F^2}$ is a measure of the bandwidth of the signal, using the standard Fourier transform properties,

$$\bar{F^2} = \frac{\int_{-\infty}^{\infty} (2\pi F)^2 |S(F)|^2 dF}{\int_{-\infty}^{\infty} |S(F)|^2 dF} \quad (11)$$

Where F denotes continuous-time frequency, and $S(F)$ is the Fourier transform of $s(t)$. In this form it becomes clear that $\bar{F^2}$ is the mean square bandwidth of the signal. From 11 and 10, **the lager the mean square bandwidth, the lower the CRLB**.From 9 the PDF is seen to be

$$p(\mathbf{x}; n_0) = \prod_{n=0}^{n_0-1} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{1}{2\sigma^2}x^2[n]) \cdot \prod_{n=n_0}^{n_0+M-1} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{1}{2\sigma^2}(x[n] - s[n-n_0])^2) \cdot \prod_{n=n_0+M}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{1}{2\sigma^2}x^2[n]) \quad (12)$$

In this case the continuous parameter τ_0 has been discretized as $b_0 = \tau_0/\Delta$. The likelihood function simplifies to

$$p(\mathbf{x}; n_0) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} exp(-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1} x^2[n]) \cdot \prod_{n=n_0}^{n_0+M-1} exp(-\frac{1}{2\sigma^2}(-2x[n]s[n-n_0] + s^2[n-n_0])) \quad (13)$$

The MLE(Maximum likelihood estimator) of n_0 is found by minimizing

$$\sum_{n=n_0}^{n_0+M-1} (-2x[n]s[n-n_0] + s^2[n-n_0]).$$

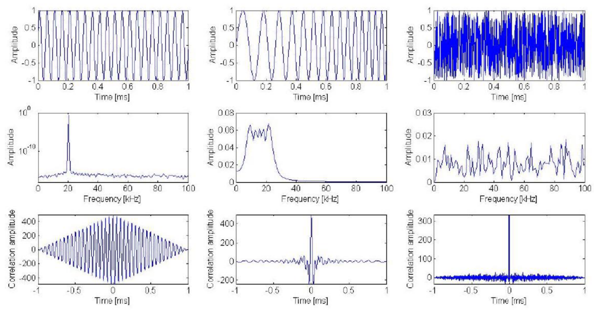

Figure 10: *the self-corralation feature of three different signal*

But $\sum_{n=n_0}^{n_0+M-1} s^2[n-n_0] = \sum_{n=0}^{N-1} s^2[n]$ and is not a function of n_0. Hence, the MLE of n_0 is found by maximizing

$$\hat{n_0} = Arg \sum_{n=n_0}^{n_0+M-1} x[n]s[n-n_0] \quad (14)$$

Note that the MLE of the delay n_0 is found by correlating the data with all possible received signals and then choosing the maximum.

3 Experiment and Analysis

The total idea is researching and developing a method that we could know the room size from the received acoustic. From the conclusion of the analysis AoT estimation, a proper input signal can easier to be recognized, it is important for us to realize our idea.

3.1 Input signal in reality

Our purpose is to find input pulse. The input pulse should satisfy the following condition.

1. It can be played by music player.
2. It can be easy to access.
3. It sounds not terrible
4. It should be easy to identify.

From figure 10 we have known that the random signal is best input signal, but it would be very noisy and very easy to mix with environment noise. In combination with the above points, we choose piano tune as the input pulse. The top picture on figure 11 show one of the piano tune wave. As show from it, the single tune can't be used as the input pulse. Therefore, we need multiple tone to realize our idea. Considering the computation power and calculation time, we select five continuous tones as our input signal.

Also, different musical scale would affect the self-correlation variance, the figure 12 shows the variance change as the musical scale change. Intuitively, the larger the variance of the data, the easier it is to identify. However, the distortion is much worse at a low frequency scale.

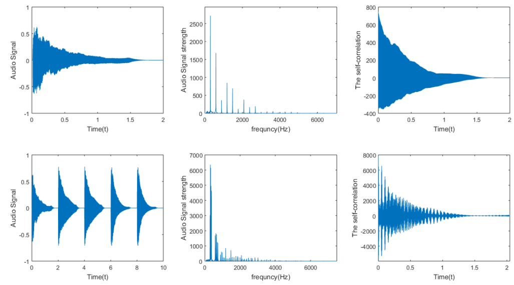

Figure 11: *the self-corralation feature of two different signal*

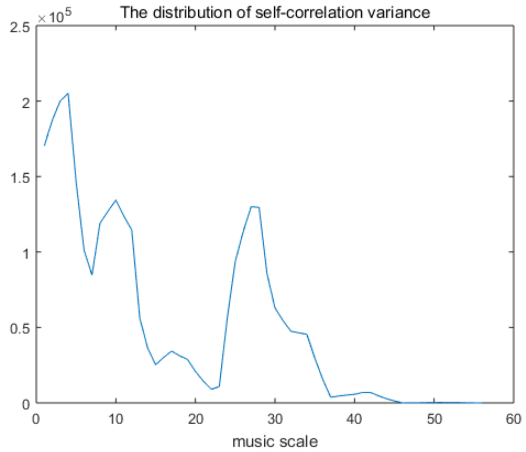

Figure 12: *The influence of differen music tone*

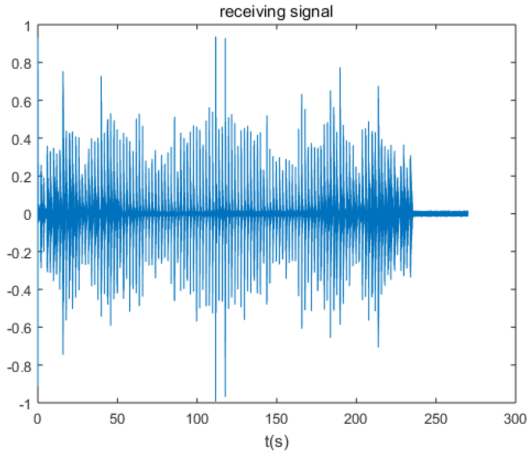

Figure 13: *One a receiveing sample*

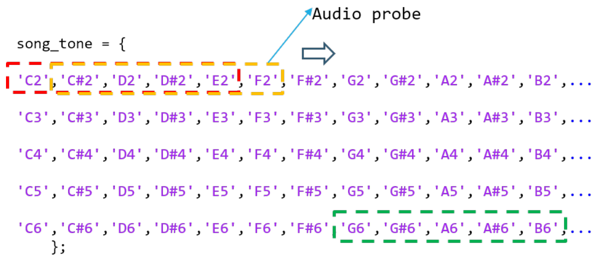

Figure 14: *The strategy of selecting probe*

3.2 Experiment situation and process

A 5W loudspeaker, TDA7297 digital amplifier with double channel and noiseless (type A), a mobile phone, a computer (ThinkPad T460) with a microphone, a normal class room.

Experiment operation process:

1. Record the room information (Lx, Ly, Lz), temperature and noise.
2. Build a coordinate system and record the speaker and receiver coordinates.
3. Open the receiver and start record sound.
4. Open the speaker and keep quiet.
5. Record about 4 minutes and 30 seconds, then rerecord another sample.
6. Change anther place and repeat 2 5 steps.

One of received sample has been illustrated in figure 13.

3.3 Directly receive time estimation

In the data processing, we need to know the time that speaker start producing a tune. With considering that, to make a probe to detect a special tune sequence is useful to estimation.

It is essential that a more accuracy start time should be estimate. Many probes have been used to estimate the start receive time t_s, which is shown in figure 14.

$$t_s = t_0 + r_0/c + 2(n(\text{probe}) - 1) \qquad (15)$$

Figure 15 shows that a probe contains different number tune can influence the estimation. A probe1 contains five tones on the left and a probe2 contains six tones on the right. The variance of arriving time estimation with probe2 is larger than probe1, so the probe2's data is more convincing.

When n=5

$$E(t_0 + r_0/c) = 5.6870s \qquad var(t_o + r_0/c) = 0.0060$$

when n=6

$$E(t_0 + r_0/c) = 5.6668s \qquad var(t_o + r_0/c) = 0.0027$$

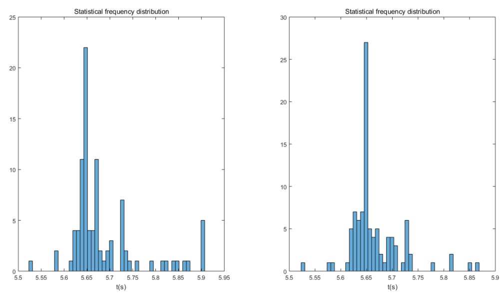

Figure 15: *Compare different number of probe*

From the arrive time estimation, the receive signal can be divided into many part. For each part, we can use same method to estimate the first-order-reflection arrive time and second-order-reflection arrive time.

3.4 Find First order echo

Ideal situation

We have discussed the receive signal in receiver when the input signal is part of sine wave. The following firstly analysis the receive signal and restore the first order echo, then apply the music signal to this physic model, and deduce some useful information.

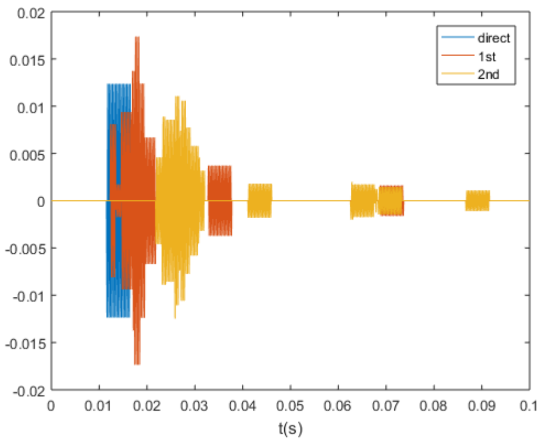

Figure 16: *The receiving signal in ideal situation*

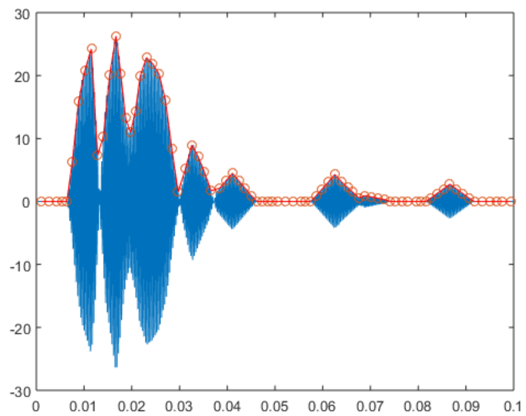

Figure 17: *The correlation between ideal signal and receiving signal*

From figure 17, due to many reflections mix with each other, we couldn't get some information about the room size. One of method is that we can put it as a footprint. If a room's size like this one, it would have same footprint. Otherwise, analysis the room size, receiver and speaker position. Noting the coordinates of speaker is very closed to the corner of wall. So, some of first-order echo could be neglected. In this condition, 1^{st}, 5^{th} and 4^{th} wall's first-order echo can be neglected. With considering this condition, it should be clear that 2^{nd}, 3^{rd} and 6^{th} wall's first-order echo is less difficult to find, but from the figure 16, it lets us know the second order echo would be received in the front of 2^{nd} and 3^{rd} wall's first order echo, so there left 6^{th} wall's first-order echo is able to be recognized. From the difference of time arrived between directly receiving signal and 6th wall's first order echo, it will give us the information about the L_z.

$$\sqrt{(2(L_z - s_3))^2 + (s_1 - r_1)^2 + (s_2 - r_2)^2 + (s_3 - r_3)^2} = (t_{1-6^{th}} - t_{\text{direct}})c \quad (16)$$

If we want to know more about the size, the third order echo information should be analysis.

Half-ideal situation

If we put a music probe as input signal, from the caculation of physic model that we have proposed in the begin, the receiving signal could be predicted. That has been shown in figure 18.

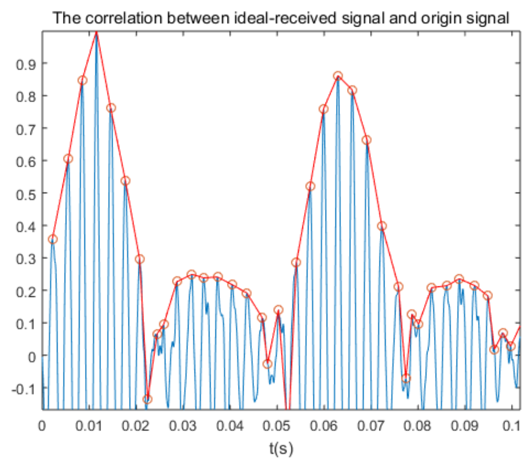

Figure 18: *Ideal receiving signal in a room*

References

[1] Fundamentals of Statistical Signal Processing: Estimation Theory/Detection Theory, Steven M. Kay, Prentice Hall, 1993

[2] Principles of Signal Detection and Parameter Estimation, Benard C. Levy, Springer, 2008

[3] Detection, Estimation, and Modulation Theory, Part I, Harry L. Van Trees, John Wiley Sons, Inc., 2001

[4] Room Impulse Response Generator, dr.ir. Emanuel A.P. Habets, September 20, 2010

[5] Method for Eciently Simulating Small Room Acoustics, J. Allen and D. Berkley,Journal of the Acoustical Society of America, 1979

[6] Acoustic signal detection through the cross-correlation method in experiments with different signal to noise ratio and reverberation conditions, S.Adrián-Martínez, M.Ardid*, M.Bou-Cabo, I.Felis, C.Llorens, J.A.Martínez-Mora, M.Saldaña, 2009.

[7] Localization of Sound Sources in a Room with One Microphone, Helena Peic Tukuljac, Herve Lissek and Pierre Vandergheynst, 2017.