Analysis and sound synthesis for human echolocation

Xiao Kai

Blekinge Institute of Technology
September 2008
Acknowledgements

With time flying, the thesis journey is now coming to the end. It might be the terminal of my master’s study here, but will never be the end of pursuing knowledge. People with countless support and care to my life and study here are the greatest treasure and irremovable memory I have gained. It is with great pleasure that I take this opportunity to express my heartfelt thankfulness to the support I have received.

First of all, I would like to express my sincere gratitude to my supervisor and examiner Dr Nedelko Grbic from my deep mind, for his enthusiastic supervision. The endless patience and encouragement from him are the confidence and motivation which support me from beginning to the end. Without his help, I could not go deep into the tough questions. His optimistic and diligent attitude will impact me in future career.

In the meanwhile, I would like to also deliver all my warmest thanks to Dr Bo Schenkman, my supervisor who gave all the guidance and support in perceptual analysis during my thesis time. His professionalism and sense of commitment really make me admiring. I shall never forget every discussions and nice chats that we shared with each other.

I would like to thank the research agency “Swedish council for working life and social research (FAS)” which provided necessary support on sounds which was used in related work.

I owe my parents Xiao Yu and Tangyu Hua who gave me birth, raised me up and provided chance to encourage me to get further education abroad. They are the first teachers in my life, who taught me the importance of honesty, perseverance and humility. Nothing could repay their kindness rather than my longing and blessing.
Abstract

Human echolocation is the way in which people use information in echoes to detect objects and locate themselves. Previous research were done in order to investigate this complex processing. One on-going research is based on empirical results of both blind and sighted people, and tries to set up the relations between behavioral data and physical analysis.

Repetition pitch and loudness are two of the mechanisms which human echolocation is based on. People’s perception of repetition pitch and loudness has been found to be strongly related to autocorrelation and the root mean square value of the sound signal respectively. This thesis project was conducted in order to analyze the relevant information in previously made sound recordings. By comparing theoretical results, some important sound properties were identified. Colorations in the original sound recordings were found to be able to influence participants’ judgments on echoes.

An acoustic model was set up in order to synthesize sound recordings with pre-defined properties. Physical results of autocorrelation and root mean square value of the synthesized sound signals were validated to follow the theoretical expectations. Discussions and proposals were made in order to further improve the sound syntheses’ correspondence to physical recordings.

The main achievements in this thesis project were to provide analysis from signal processing’s point of view in order to identify sound properties, to investigate relations between physical results and empirical results of human echolocation and to propose an acoustic model to synthesis sound.
Contents

1. Introduction ................................................................................ - 7 -
2. Related work .............................................................................. - 9 -
3. Requirement analysis and methodology ................................. - 11 -
4. Data extraction and analysis of the sound recordings .......... - 13 -
   4.1 Data extraction of ACF quotient ........................................ - 13 -
   4.2 Data extraction on RMS value .............................................. - 21 -
5. Sound synthesis and analysis ................................................... - 25 -
   5.1 Acoustic model identification .......................................... - 25 -
   5.2 Synthesis of sounds ............................................................ - 28 -
   5.3 Synthesis confirmation ..................................................... - 30 -
6. Confirmation on original sound recordings ............................. - 33 -
7. Conclusions and Discussions ................................................... - 36 -
   7.1 Main findings and implications ........................................ - 36 -
   7.2 Discussions and proposals for further study ..................... - 38 -
References ...................................................................................... - 41 -
Appendix .......................................................................................... - 45 -
1. Introduction

Human echolocation relates to the ability of people to detect objects in the environment by perceiving echoes reflected from those objects. It is different from passive acoustic localization which localizes the position of sound emitted by objects, whereas humans’ echolocation is an active acoustic localization involving the creation of sound in order to produce an echo which is then analyzed to determine the location of the object in question [1]. In the natural world the echoes are reflections of the original sound sources, so they may come from many different directions with different intensities compared to the original sound wave. In human echolocation, both the sound source and its reflections are used in the process [2]. The need of both the sound source and its reflections could be due to the complex sound environment in real world. Since there is a huge amount of different sound sources in reality, it is possible to have other sounds which have the same characteristics as the “echo” that was defined as the reflection of the original sound source. People might feel hard to distinguish the existence of a certain “echo” from all sound sources. But when sound source and its reflected wave are presented together, human beings have the ability to perceive “differences” between the direct-path and its reflections [3][4]. Such kind of “hearing mechanism” is how the hearing system gets information from echoes in order to detect the objects. It is hard to find a suitable model which could describe how human echolocation works, but previous research show that despite the differences among people, human echolocation involves many “hearing mechanisms” and combines them together so that people can get useful information from certain acoustic environment [2][4][5][6][7].

Among all of those “hearing mechanisms”, the repetition pitch and the loudness strongly contribute to human echolocation [1][2][7][8]. Repetition pitch is based on the perception of a frequency. When a sound and the repetition of that sound after a time delay $\tau$ are added and presented to a listener, monaurally, a subjective tone is perceived with a pitch corresponding to the reciprocal value of the time delay $\tau$ [8]
[9]. For the repetition pitch, people can also perceive the strength and it allegedly depends on the first peak in the autocorrelation function [10]. As for the loudness mechanism, it relates to the attenuation caused by the sound propagation. When a sound wave travels in space and gets reflected by an obstacle, it loses energy both from propagation loss and loss from the reflection because it is subject to energy loss arising from medium absorptions [11]. In other words, there is always a difference in loudness between the direct path and any of the reflected paths. For people’s hearing mechanism, by comparing this difference in loudness, useful information can be provided about an object’s position.

To further understand human echolocation, it is important to investigate how those hearing mechanisms function in different acoustic conditions and how they cooperate with each other. Previous research has been done [1][2][12][13][14], and both physical sound analysis and psychological analysis have been used (for more details on history review, see Ch 3, "Related Work").

One on-going research project (more introductions in Ch 3, "Related Work") has provided experimental results from tests with people. In the earlier work, Schenkman and Nilsson recorded sounds in different acoustic conditions and used those sound recordings in their experiment to test people’s perception. This thesis project was set up with the purposes (1) To investigate those experimental results and to find out how they coincide with the empirical results by extracting the data information from the sound recordings; (2) To design an acoustic model and synthesize sound stimuli based on the properties of the original sound recordings for further study on the sound properties. The analysis in this thesis was aiming at investigating how the properties of original sound recordings could affect human perceptions in the earlier experiment.
2. Related work

Previous research in human echolocation dates back to the early 1900s. The 40s is usually considered as the beginning of scientific experimentation in this field. Before it was known to be based on localization of echoes, human echolocation was then sometimes described as “facial vision” [3]. Dallenbach and Cotzin showed that both necessary and sufficient condition for echolocation is a pitch change [15].

After that, although some studies were done on the discriminatory power of human echolocation [16][17][18], it’s rare to find a systematic research reported until the late 70s. Schenkman tested in a laboratory the ability of detecting objects by blind people in the 80s [2]. Yost did systematic research on repetition pitch and defined acoustic models for iterated ripple noise (IRN) [10][19][20] in the late 70s. During the same period, Bilsen also studied the repetition pitch and its relation to human echolocation [6][8][13]. More recently, many experiments and studies were taken to investigate human echolocation ability with focus on the difference between normal vision people and vision handicapped people [1][14][21]. Research is aimed at understanding how different hearing mechanisms work in human echolocation and how acoustic conditions affect a subject’s ability in audio perception.

One on-going research project by Schenkman and Nilsson [11] aims at investigating differences between sighted and blind people regarding human echolocation ability. In order to test an objects’ perception, a laboratory test was designed. They recorded sounds in two different room conditions (an anechoic chamber and a conference room), with three different reflection distances and six different sound durations. Those sound recordings were then presented to participants in a laboratory. Each participant was asked to judge if he/she could perceive the echoes based on his/her perception. Percentage of judgments for all participants is given in [11].

From those empirical results of tests, they studied the differences between different objects, to what extent the differences was a result from different acoustic conditions
and what was the relationship between the empirical results and the theoretical considerations. A further extension of their research was to find a suitable model which could describe how different hearing mechanisms work in human echolocation.

Among the empirical results from Schenkman and Nilsson’s research, the mean percentages of correct judgments show that blind people have a better ability to detect the echoes than the sighted people in all conditions. They also found that detections of all participants were better in the conference room than in the anechoic room. Compared to sighted people, blind people rely more on sound and echo in daily life. The more experience of using echo’s to locate objects might explain the difference in empirical results between blind people and sighted people [11]. The different percentage of correct judgments in different room conditions might indicate that human echolocation is influenced by how much information the listener can get, since multiple reflections are present in a conference room while there is only one in the anechoic chamber for the same sound stimulus [11].

Both sound duration and propagation distance were found to influence participants’ ability of perceiving echo. Empirical results showed that sound recordings with long durations were much more perceivable than those with short durations. When considering the propagation distance, the closer it is, the easier the sound echos can be perceived [11].
3. Requirement analysis and methodology

The related work by Schenkman and Nilsson introduced above provide empirical results from human participants. One aim was to develop a more formal model to account for the obtained results. Development needs to be done on the acoustically measured data and the perceptual judgments of the participants in different situations.

This thesis work is related to the work by Bo Schenkman and Mats Nilsson. Based on the results of their research, a data analysis was done in order to investigate how the relevant information from the sound recordings correlates with the empirical results provided by the test persons. Results from their research, like differences between blind participants and sighted ones, differences between two room conditions and differences between sound durations and reflection distances etc., need to be analyzed both from a psychoacoustic and a physical aspects.

The approach regarding the requirements could be put in the following processes:

First, the original sound recordings are investigated by extracting useful information. Different sound properties lead to differences in people’s perception of the sound recordings. An efficient way to figure out sound properties could be by estimating their relevant contents related to repetition pitch and loudness, which were discussed in the Introduction part. I.e. the investigation of the original sound recordings is to be based on the extractions of frequency and energy content.

Secondly, a suitable model is set up based on acoustic theory in order to identify the sound properties. The parameters in the acoustic model should be variable so that sound signals with pre-defined properties could be synthesized from it i.e. properties such as frequency and energy contents of the sound syntheses should strongly follow the theoretical expectation. The sound synthesis will then be taken as a comparison to study the original sound recordings.

Thirdly, with the frequency and energy extractions, a comparison should be done
between the sound recordings and the sound syntheses. The aim of the comparison is to further indentify the features of the original sound recordings. With sound syntheses, further studies could be proposed to understand how different sound properties affect human perception. Besides, whether the acoustic model could be improved to simulate the real acoustic room conditions should also be discussed. Using sound synthesis could be an alternative way in further study of human echolocation.
4. **Data extraction and analysis of the sound recordings**

Based on the analyses on the participants’ perception test, Schenkman and Nilsson formed hypotheses on how each factor impacted on the human echolocation ability [11]. To confirm the empirical results, physical analysis was done in this thesis on the sound recordings, especially for those differences between different acoustic conditions.

In order to investigate the relationship between the sound properties and the participants’ perception, data extractions were made from the original sound recordings which were presented to the participants in their experiments. The ACF (autocorrelation function) and the RMS (root mean square) value were found to provide important acoustic information, since they have strong relationship with the repetition pitch and loudness perceptions respectively [11]. By calculating strengths of ripples in the ACF, the power of the original sound and its reflections was identified. Information regarding the repetition pitch can also be found from the distance between the main peak (which refers to the original sound) and the side peak (which refers to the reflection) in the ACF. The RMS value is a related measure of a signal’s mean energy which relates to the loudness perception.

4.1 **Data extraction of ACF quotient**

According to previous research on pitch theory (Bilsen 1968; Yost, Patterson et al. 1996), the autocorrelation function is an essential component of repetition pitch. The autocorrelation sequence $r_x(k)$ for a WSS (wide-sense stationary) random process $x(n)$ is [22]:

$$r_x(k) = \sum_{n=-\infty}^{\infty} x(n)x(n+k)$$
\( r_x(k) = E \left\{ x(n)x^*(n - k) \right\}, \tag{1} \)

where the index \( k \) refers to the time-lag parameter and \( * \) refers to the complex conjugate.

It provides a time domain description of the second-order moment of the process. Since \( r_x(k) \) is an energy signal we can compute its discrete-time Fourier transform \( P_x(e^{j\omega}) \),

\[
P_x(e^{j\omega}) = \sum_{k = -\infty}^{\infty} r_x(k) \cdot e^{-jk \omega}, \tag{2}\]

which defines the power spectral density of the random process \( x(n) \) [23].

Given the power spectrum, the autocorrelation sequence may be determined by taking the inverse discrete-time Fourier transform of \( P_x(e^{j\omega}) \), i.e. \( r_x(k) \),

\[
r_x(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_x(e^{j\omega}) \cdot e^{jk \omega} \cdot d\omega, \tag{3}\]

where \( k \) is a time-lag parameter.

If \( k \) is set to 0 in (3), then

\[
r_x(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_x(e^{j\omega}) \cdot e^{j0\omega} \cdot d\omega
= \frac{1}{2\pi} \int_{-\pi}^{\pi} P_x(e^{j\omega}) \cdot d\omega, \tag{4}\]

where the right side of (4) is the integral of the power spectrum function \( P_x(e^{j\omega}) \) within the area \([-\pi, \pi]\), i.e. it is the total power of the random process.
In previous research, Yost has argued that the strength of the repetition pitch depends on the first peak in the autocorrelation function [10]. In the ACF of one IRN (iterated ripple noise [10][19]) sequences, several many other peaks besides the first one can be found. They are related to the iterations of the noise. In room acoustics, they are explained as the reflections of the original sound wave in certain acoustic environment.

According to the theory for repetition pitch, the pitch which is perceived corresponds to the reciprocal value of the time delay $\tau$ between direct path signal arrival of the sound wave and its reflection [8][10]. The quotient between the autocorrelation value of the reflection (ACF2) and the autocorrelation value at time 0 (ACF1) can be used to study the changes in power of the reflection when the sound characteristics are changed.

To get the average quotient, the autocorrelation for each sound recording was calculated, ACF2 was divided by ACF1 and then the average of each ten stimuli (sound recordings were made ten times for each acoustic condition in Schenkman and Nilsson’s experiment [11]) with the same sound properties was calculated. The sound speed was assumed as 342 m/s in all acoustic conditions and a 5% acceptable error was chosen to compensate measurement error in propagation distance. The result is shown in Figure 1.
Figure 1. Autocorrelation quotient for the signals, the value at the time-lag of the reflection divided by the value at time-lag 0, for the six different distances and the three time durations in the two rooms. Acceptable error of the propagation distance is ±5%.

Figure 1 shows the ACF quotients of sound recordings with six propagation distances and three time durations in two rooms. The solid curves with star-marks in both figures show the ACF quotients of 5ms sound recordings in two rooms, where the star-marks indicate the ACF quotient of each distance. Similar to it, the dash lines with diamond-marks illustrate ACF quotient of 50ms sound recordings and dash lines with round-marks illustrate ACF quotients of 500ms sound recordings respectively. Figure 1 a) shows the ACF quotients for the six different distances and the three time durations in the anechoic chamber. It can be observed that for each reflection distance, ACF quotient is mostly largest for 5ms sound recordings, second largest for 50ms ones and smallest for 500ms in general. This inverse proportion between the ACF quotients and duration of sound recordings in Figure 1 a) has one exception when the reflection distance is equal to 4m, where the ACF quotient of 50ms sound recordings is greater than the one of the 5ms sound recordings. Both Figure 1 a) and b) have the
same scale on both axes. Compared to the anechoic chamber in Figure 1 a), the relation between ACF quotient and sound duration is less consistent in the conference room shown in Figure 1 b). The ACF quotients versus reflection distance go up and down from one distance to another, which make them hard to interpret.

The ACF quotient is a function of reflection distance, sound duration and room condition. In each room condition, when sound properties are changed, there should be a consistent trend of the ACF quotient. However, the trend of the physical signals result shown in Figure 1 is inconsistent.

One should also note the starting point of each curve. The plots should have been correlated with each other because Schenkman and Nilsson used white noise as sound stimulus in all the experiments. Having a given bandwidth, the PSD (power spectral density) of a white noise has equal power in any band, at any centre frequency. As was mentioned in the introduction part, the repetition pitch corresponds to the reciprocal value of the time delay $\tau$ which is the time that the sound wave travels within the reflection distance. The ACF quotient is assumed to be an essential component of repetition pitch [8][10]. Therefore the ACF quotient and the reflection distance should have a potential relation, i.e. the ACF quotient should change with a certain tendency when reflection distance is increased. But it is hard to find such a relation in Figure 1. One possibility for those irregular behaviors could be the problems of the sound stimuli themselves, which means that the sound recordings used in previous experiments were not pure white noise, but comprised some additional colorations.

There might be two causes to the colorations. One is the nonlinearities in the electronic equipments and the electrical current transmitting medium used to produce the sound recordings. Since one cannot avoid the nonlinearities of the loudspeakers, cables etc., even when the signal from the generator is pure white noise, they might be colored before being played. Because the sequence from the signal generator was not recorded in Schenkman and Nilsson’s experiment, the hypothesis about the
nonlinearities cannot be validated.

The ACF quotients in Figure 1 might also depend on the assumptions that were made when they were calculated. As was mentioned before the sound velocity was assumed as 342m/s and that the acceptable error of the propagation distance was assumed as ±5%. Both were made in order to get the time delay $\tau$, which was calculated by dividing the propagation distance by the sound velocity. With the time delay $\tau$, one could calculate the time-lag $k$ in (3) so that the position of ACF2 could be located. The ACF quotient was calculated with ACF2 divided by ACF1. But as for acceptable error of the propagation distance, it was not certain that ±5% was a proper setting to compensate for the variability of the velocity by the influence of mediums’ properties (e.g. temperature, humidity and density) and to avoid the nonlinearities’ coloration mentioned above. If it was not a proper setting, all the derivations above might be influenced, which could lead to imprecise results in the ACF quotient.

In order to validate the influence from acceptable error of the propagation distance, 3 other settings than ±5% were also chosen. These were ±0.05%, ±0.5% and ±10%. The ACF quotients recalculated based on these 3 settings are shown in Figures 2 to 4, respectively.
Figure 2. Autocorrelation quotient for the signals, the value at the time-lag of the reflection divided by the value at time-lag 0, for the six different distances and the three time durations in the two rooms. Acceptable error of the propagation distance is ±0.05%.

Figure 3. Autocorrelation quotient for the signals, the value at the time-lag of the reflection divided by the value at time-lag 0, for the six different distances and the three time durations in the two rooms. Acceptable error of the propagation distance is ±0.5%.
Figure 4. Autocorrelation quotient for the signals, the value at the time-lag of the reflection divided by the value at time-lag 0, for the six different distances and the three time durations in the two rooms. Acceptable error of the propagation distance is ±10%.

Figure 2 and Figure 3 are based on the acceptable error of propagation distance being ±0.05% and ±0.5% which are less than ±5% used in Figure 1. Smaller acceptable error setting leads to narrower acceptable band (the acceptable band is equal to propagation distance multiplied with the acceptable error) where the ACF2 is located. It reduces the compensation of error from reflection distance, sound duration and sound speed assumption which will decrease the accuracy of ACF2. Because of the existence of coloration, the wider the acceptable band is, the higher the possibility will be for the coloration to be involved in. This could decrease the accuracy of ACF.

The effects from acceptable error of propagation distance can be observed by comparing Figures 2 and 3 to Figure 1. Among the comparisons, the ACF quotients in anechoic chamber at distance 4m are remarkable. In Figure 1, as was mentioned above, the only exception to the inverse proportion between the ACF quotient and duration of sound recording in anechoic chamber happened when the reflection distance was 4m. The ACF quotient of 50ms sound (dash line with diamond-marks in
the figure) at this distance decreases when the acceptable error of propagation
distance is decreased, until becoming smaller than the ACF quotient of 5ms sound
(solid line with star-marks in the figure) at the same distance. On the other hand,
when the acceptable error of propagation distance (e.g. Figure 4) is increased, the
ACF quotients do not change much, i.e. there is not a distinct difference between
Figure 1 and Figure 4.

Even though the acceptable bias of propagation distance affects the ACF quotients
slightly, the entire plots still have unusual shapes. The analysis above shows that the
sound recordings may have had some coloration which can not be compensated by
changing the acceptable error of propagation distance.

Another issue regarding the reflection needs to be clarified which might also have
influenced the ACF quotient of the sound recordings. Since the perceived repetition
pitch only depends on the first arrived reflection as was discussed before, and the
other reflections later on are not as important as the first one, they can be ignored [10].
But whether they can be disregarded in the ACF quotient analysis is not evident. If the
time interval between several reflections is too short to distinguish, it might result in
mis-selection when calculating the first reflection. This might be one explanation for
why the figure of the conference room is more irregular than the one of anechoic
chamber, since there are more reflections from different directions in the conference
room.

### 4.2 Data extraction on RMS value

As was mentioned above, RMS value measures the energy of the signal. Besides
repetition pitch, humans can also use information from the power of the sound by
perceiving the loudness [11].
RMS is known as the root mean square, which is a statistic measure of the magnitude of a varying quantity [23]. For the continuous time series \( x(t) \) defined over time interval \([t_1, t_2]\), is calculated based on the following formula [23]:

\[
x_{RMS} = \sqrt{\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} [x(t)]^2 \, dt}, \tag{5}
\]

The continuous time series \( x(t) \) is white noise, which has infinite bandwidth. By sampling the original continuous form, its discrete time form is as follows:

\[
x_{RMS} = \sqrt{\frac{1}{n} \sum_{i=0}^{n-1} x(i)^2} = \sqrt{\frac{x(0)^2 + x(1)^2 + x(2)^2 + \cdots + x(n-1)^2}{n}}, \tag{6}
\]

where \( n \) is the number of samples in the discrete time series and \( x(i) \) are the discrete time samples.

To calculate the RMS value for all the sound sequences, a sample frequency of 48000Hz was used, since it had been used to record those sounds. Then the RMS values were calculated with the formula (6) for each signal (based on the sound recordings from the study by Schenkman and Nilsson) and were plotted as Figure 5.
In Figure 5, RMS values of all the sound recordings are plotted in decibel scale. There are three conclusions which can be made from this figure:

First of all, in both rooms, the RMS values of sound recordings for all time durations decreased when the reflection distance was increased. For the sound recordings with the same reflection distance, the longer time duration they had, the larger their RMS value was. As was discussed before, when sound travels in space and gets reflected, the longer the propagation distance is, the more energy will be lost due to the propagation attenuation. On the other hand, the energy of white noise is proportional to the duration of the sound. Thus, for the same propagation distance, the longer the sound duration is, the larger the RMS value will be.

Secondly, the level of the RMS values between the anechoic chamber and the conference room, for the same sound duration, was higher in the conference room than in the anechoic room. The sound stimuli which were used in the two rooms were
the same. There is more reflections in the conference room than in the anechoic chamber, which would be added to the original sound wave. The RMS value of the sound recording with the same duration and reflection distance is higher in the conference room. In addition, because of the fixed room conditions (temperature, humidity, room layout etc.), the environmental effects were the same for different sound recordings. Thus, for the same sound recording (sound duration and reflection distance), the difference of its RMS values in two different room conditions should be a fixed value. This could explain the level-shift in dB scale between the two rooms’ RMS value, i.e. each curve has nearly the same shape as the one with the same marker in the other figure, but they have different levels.

Thirdly, comparing all figures with the ACF quotients, Figure 5 for the RMS value is more regular in shape. The colorations found on the white noise signals have little interference on the energy contents of the signals, but they have a big influence on the autocorrelation sequence of the sound which can influence the repetition pitch that participants perceived.

In conclusion, the plot of RMS value in dB scale follows expectation more than the plot of ACF quotients. Although there are one or two irregular points in the plot of the anechoic chamber, they still indicate the expected sound variation based on the different durations and propagation distances.
5. Sound synthesis and analysis

The analysis in the previous part illustrate the properties of the original sound recordings, which related to the human perceptions of repletion pitch and loudness. From the physical result of the ACF quotients and the RMS values, one might hypothesize that the original sound recordings have uncertain coloration. In order to further identify the properties of original sound recordings and relate psychological perception to the theoretical data analysis, one could reduce those interferences in the original sound recordings. One alternative way of study is to synthesize sound recordings with identified properties i.e. to set up a general acoustic model and synthesize different types of sound recordings with pre-defined features which can be taken as a comparison to the original sound recordings. With these sound syntheses, a new experiment can be set up, asking the participants to redo the same experiment.

5.1 Acoustic model identification

In an enclosed environment, any sound wave is received with multiple reflections. The combination of reflections can increase the strength of the pitch and the coloration of the rippled noise (Yost, Patterson & Sheft, 1996). There are two system models we can use to define the acoustic environment, which are illustrated in Figures 6 and 7:

![Diagram](image)

Figure 6. The network used to generate add-original iterated rippled noise, where d refers to the delay and g refers to the gain.
The models above describe two of the most common networks on how iterated noise can be generated. If the original sound is kept playing while getting reflected, then the source sound will be added to each reflection as illustrated in Figure 6. If the original sound has no continuation or a very short duration, then it will be only taken as an impulse to generate the iterated noise as shown in Figure 7.

The main difference between these two structures is the input in each section. In the first network, the original input $x(n)$ is long-duration stimuli to the network which last for the whole generation process. The original input $x(n)$ is added to every sections. In the second one, the input $x(n)$ has a very short duration which can only execute the first section in the network. For each section afterwards, the input will be the summation of the output of the previous section and its modified value from this section. Both of them can generate the sound stimulus which was suggested to be referred as IRN (Iterated Ripple Noise) by Yost et al. (1993). Therefore, they named the first model as IRNO which refers to Add-Original iterated rippled noise, and the second one as IRNS which refers to Add-Same iterated rippled noise correspondingly.

Furthermore, their system function (in spectral terms) can be generated from each model as follows:

IRNO:

$$H(\omega) = 1 + g \cdot e^{-j\omega d} + g^2 \cdot e^{-2j\omega d} + \cdots + g^n \cdot e^{-nj\omega d}.$$  \hspace{1cm} (9)
IRNS:

\[ H(\omega) = \left[ 1 + g \cdot e^{-j\omega d} \right]^n, \]  \hspace{1cm} (10)

where \( g \) is the gain which refers to the propagation attenuation, \( d \) is the time delay of the reflection and \( n \) is the number of the blocks.

In this thesis project, the most important part is the first reflection of the original sound stimulus, because it will mainly affect the analysis on the ACF quotient and the RMS value [8][10][24]. These two models can be simplified into a one-order term as follow:

Figure 8. The one-order network used to simulate sound stimuli, where \( d \) refers to the delay and \( g \) refers to the gain.

Figure 8 illustrates the acoustic model setup which was used to synthesize IRN sound with only a single reflection. The input \( x(n) \) is a white noise sequence with certain RMS value. In order to simulate the reflection of the input sound \( x(n) \), parameters \( d \) and \( g \) are chosen as the delay time and the attenuation respectively. The output \( y(n) \) is the summation of the direct path \( x(n) \) and its reflected path which are modified by \( d \) and \( g \).

The system function of the network illustrated in Figure 8 is given by:

\[ H(\omega) = 1 + g \cdot e^{-j\omega d}, \]  \hspace{1cm} (11)

where \( g \) and \( d \) are defined in the same way as they were in equation (10)
When it comes to the attenuation, the power loss from an acoustic propagation in space is proportional to $\frac{1}{R^2}$, where $R$ is the distance of propagation. In order to determine the attenuation factor $g$, a certain propagation distance was chosen in advance i.e. assuming a proper value $g_0$ which was between 0 and 1 as the initialization for $g$. And then adjust $g$ according to the ratio between the present distance and the initial distance.

$$g = g_0 \cdot \frac{R^2}{R_0^2}.$$  \hspace{1cm} (12)

### 5.2 Synthesis of sounds

Based on the system model which was identified above, one can generate sound stimuli with identified properties. As was described in Chapter 2, Schenkman and Nilsson tested objects’ perception in human echolocation by having sound recordings with different acoustic properties, i.e. sound recordings which were generated in varying conditions (room condition, propagation distance and sound duration). In the sound synthesis, it is hard to simulate the real environments as was used in their experiment. The way they used to classify sound recordings is not suitable here. However, by specifying the parameters in the acoustic model (Figure 8), one can modify properties of the sound synthesis. By setting up classification, sound syntheses with the same properties were grouped together. The purpose of the experiment is to categorize sound syntheses in different groups according to their properties instead of acoustic conditions, so that one can analyze them and investigate the test object’s perceptions on each category of them.

Regarding the classification criteria, it is suggested that it could be an efficient way to start from two main elements which are repetition pitch and loudness. It means the sound synthesis need to be distinguished by its properties at least from aspects of
repetition pitch and loudness. According to these ideas, sound synthesis were classified on three categories as follow:

1) Sounds having both information in repetition pitch (frequency) and in loudness (i.e. power) from reflections. They are similar to the sound recordings in Schenkman and Nilsson’s study from real rooms. They are:
   - Sound with no reflections
   - Sound with reflections having both pitch and correct adjustment of RMS loudness.

2) Sounds having only information in repetition pitch. These sounds have the same RMS values (i.e. power).
   - Sound with no reflections
   - Sound with reflections

3) Sounds having only information in loudness. These sounds only differ in having different RMS values.
   - Sound with no reflections with a certain RMS
   - Sound with no reflections, with an increased RMS, related to the RMS contained in a reflection

Their difference should be equal to the difference between the sounds in category 1).

To generate those sound synthesis, Matlab was used based on the system model that were defined in Chapter 6. The input stimuli of the system was white noise. Since white noise is a random process, for each category 10 realisations of the sound synthesis were generated. All of the sound synthesis were normalized, and then named according to their properties and version, such as “soundstimuli_both_(propagation distance)_(sound duration)_(version number)”.
5.3 Synthesis confirmation.

Based on the sound synthesis, a data analysis was done in order to confirm their properties. As before, ACF quotients and RMS values is the focus of the analysis. Since synthesis in the second and the third categories are actually the components of the synthesis created in the first category, it is enough to check those sound synthesis in the first group instead of all of them.

![Figure 9. Autocorrelation quotient for the sound syntheses, the reflection divided by its value at time 0, for the six propagation distances and the three time durations](image)

As illustrated in Figure 9, the curves for different sound durations have nearly the same decreasing rate when the propagation distance is increasing. That follows theoretical expectations, because once the system model is determined, for a certain distance, the ratio between the attenuation and the original sound should be a fixed value, which means that the ACF quotient should be nearly the same.
Figure 10. RMS value for the simulations in six propagation distances and the three time durations

In Figure 10, three RMS curves array in certain order from the longest duration to the shortest one from up to down as was discussed before in chapter 6.2. Observe that for in long durations the RMS values only decrease slightly when distance increase. The almost constant parts in the figure might indicate that at long distance, the loudness of the reflection will not change too drastically.

In conclusion, the artificial sound synthesis have reasonable properties both in ACF quotients and RMS values. The comparisons between the sound synthesis and the original sound recordings (i.e. the comparisons between Figure 9 and Figure 1, Figure 10 and Figure 5) indicate that the colorations in the original sound recordings can influence the sound properties such as the ACF quotients and the RMS values. This indicates the colorations can also influence human perceptions of the repletion pitch and loudness.
In order to test those sound synthesis, a few sighted people were chosen to listen to them in pairs and see if they can perceive any difference in-between. All of them had difficulties on correct judgments especially when short duration sounds were presented. One person claimed that he even felt hard to perceive the difference in condition of 500ms duration and 1m distance, which is the case where differences could be clearly perceived in previous test. Regarding this result, further discussions will be drawn in Chapter 7.
6. Confirmation on original sound recordings

In the previous two parts, both ACF quotient and the RMS values of the sound recordings and sound synthesis were calculated. Comparisons illustrate that the sound recordings may not have been pure white noise, but might have been pre-filtered by an unknown system (this could be due to the recording equipment, storage medium etc.), which lead to the irregular ACF quotients and the RMS values. To confirm this hypothesis, autocorrelation of all the 18 sound recordings in the anechoic room were calculated to be compared with their theoretical results that were provided by the respective sound synthesis. No consideration was made of the 18 sound recordings in the conference room, because it is difficult to synthesize sounds in a conference room which is a complex acoustic environment. The investigation here was whether the sounds having been used were pure white noise or not. This has nothing to do with the room conditions, since the sounds were generated in advance. Therefore the stimuli in the anechoic chamber are representative. The comparisons between theoretical sound recordings and original sound recordings are illustrated pairwise in the Appendix.

In order to construct comparable results, all the autocorrelation functions in the Appendix were normalized. In each figure, the solid line designates the left channel and the dashed line designates the right channel, since both theoretical and original sound recordings were dual channel recordings. There are several noticeable points found for the 18 pairs of comparisons:

1) One can easily find the differences between the theoretical sound recordings and the original ones in the autocorrelation functions. Take the distance 1m and duration 5ms in Fig. A.2., as an example. According to the acoustic theory, assuming sound velocity is 342m/s, one can calculate the time position of its reflection based on the following process:

For an object at 1 m distance, the propagation time of the sound stimulus is:
\[ t = \frac{2d}{v} = \frac{2 \times 1}{342} = 0.0058s , \]  

(7)

The sample position of the equivalent time lag is:

\[ p = t \times F_s = 0.0058 \times 48000 \approx 280.7 , \]  

(8)

The theoretical position of the reflection should be around sample 281 in the autocorrelation function as in the left part of the figure. But in the right part which is the autocorrelation of the original sound recording, the most significant reflection is located somewhere around sample 20. As was mentioned in the section of Introduction, the repetition pitch that people perceive is related to the time lag of the reflection [8][9]. In other words, when we present this sound to the observers, the reflection which will influence their perception and judgment on repetition pitch is actually the one at sample 20 but not the one in sample 281 as was expected. The same thing holds for all the original sound recordings.

2) The intensity of a reflection should decrease when the reflection distance is increasing (see analysis on RMS values in the previous part for details) as was seen from the left part of each figure in the Appendix. But in the right part of each plot which shows autocorrelation of the original sound recordings, it’s very hard to see this trend. In some of the figures like Fig. A.15. it’s even difficult to identify which ripple is related to the reflection.

3) The white Gaussian noise that was used in the experiment was a WSS (wide-sense stationary) random process, which has zero-mean, i.e. it has no DC component. This makes some autocorrelations of the original sound recordings in the Appendix problematic. For example, when distance between sound source and obstacle is 200cm and the sound duration is 5ms, the autocorrelation of the sound recording was plotted as in Fig. A.3.. The left part is the normalized autocorrelation of the theoretical result, of which the ACF is close to zero for
values outside of the time-lag of the reflection. But the right one, which illustrates the normalized autocorrelation of original sound recording, has a DC component around level 0.3 in the normalized scale.

In conclusion, the original sound recordings which were used in the experiments by Schenkman and Nilsson [11] were probably not pure white noise. The unaccounted colorations may be the cause of the irregular plots of ACF. For participants in their previous experiment, those interferences from the colorations could have influenced the perceptions of the repetition pitch with the original sound recordings.
7. Conclusions and Discussions

7.1 Main findings and implications

The emphases of this thesis work were: a) understanding and analyzing the sound recordings and empirical results; b) finding a suitable model to generate sound syntheses in order to identify the properties of the original sound recordings. From the analysis, the interferences in the original sound recordings were found, which might affect participants’ perception of the repetition pitch. A model which can be used to synthesize IRN was proposed. The accomplishments in the thesis project are data analysis and sound synthesis.

First of all, from previous research and related work, a systematic background and a specific research aim were set. Human echolocation is the effect of many hearing mechanisms [2][12]. Each of those factors has its own working condition in which it might mainly function, but none of them can maintain the echolocation independently [4][5][12]. Each hearing mechanism is probably only regarding one type of information, e.g. repetition pitch is perceived as a sound frequency. The more information one can get the easier for him/her to perceive a sound. Human echolocation depends on how much information people can get from the environment. The higher detection in the conference room compared with the anechoic room could support this inference [11]. One main point in studying human echolocation is how to understand this multi-motivated bio-system.

Secondly, based on the original sound recordings, a set of data analysis were done to investigate how the information in them can coincide with the physical results. Data extraction on ACF quotients and RMS values showed that the physical sound recordings had uncertain properties which can influence people’s perception. A set of comparisons were done to confirm the hypothesis on whether the original sound recordings were pure white noise or not. From the confirmation, it is indicated that the sound recordings which were used in Schenkman and Nilsson’s experiment might
have spectral colorations due to the transmission and of interference besides the modeled reflection from the environment, which made them not pure white noise. From both empirical and physical analysis, the sound duration and propagation distance of the sound were shown to have effects on human echolocation. To investigate how different factors may influence human echolocation, it is important to know each of their working condition. If the working condition is chosen properly, one factor might be found as the dominate one of the human echolocation in this condition. For example, at far distances, the corresponding repetition pitch becomes too low to be perceived. One may more rely on the loudness of the sound. In this case, how the loudness mechanism may work as compensation in human echolocation could be investigated in further study. Further, the white noise which was used in Schenkman and Nilsson’s experiments is a common sound type which exists in reality and it covers the whole band of human being’s hearing spectra. Whether other kind of sounds could have different behavior in human echolocation could also be proposed in further study.

A model was used to synthesize IRN which was used as an alternative way to study human echolocation. The comparisons between original sound recordings and sound synthesis showed that colorations in the original sound recordings might mainly influence the ACF values which could affect human perception of the repetition pitch. As mentioned in Chapter 5, although the sound synthesis were supposed to be more distinguishable than the sound recordings used in Schenkman and Nilsson’s experiment, participants felt it more difficult to perceive differences between them. Regarding this question, some suggestions are raised:

1) The difference between theory and reality should be considered. Lots of interference always exist in real environment, and it is very difficult to simulate completely real conditions. One proposal is that some proper interferences could be added to the sound synthesis in order to simulate some of the complexities in reality. Since it will be nearly impossible to cancel all the differences between theory and
reality, one efficient way might be adding different types of interference into the synthesis and then try to investigate their effects on improving the sound properties. One alternative way is to adjust the reflection components artificially instead of using the network to generate sound synthesis.

2) Another proposal is to check whether there is any different factor between far distance propagation and close distance propagation. It is thought it would be helpful to choose another different system models for far distance cases, since differences in both repetition pitch and loudness are extremely hard to be perceived in the sound syntheses when the propagation distance is large.

3) The process to synthesize the sounds also needs to be improved. As was described before, there were many parameters which need to be set to design the model. But those values were chosen at the beginning and all used afterwards. The applicability for this method needs to be validated in a further study, i.e. those pre-settings have to be investigated whether they vary when the environment conditions change.

4) Since hearing is not a purely mechanical phenomenon of sound propagation, but also a sensory and perceptual event, another suggestion is that other aspects of psychoacoustics could also be considered in order to improve the model. Influence such as limits of human perception and masking effects [2][12] are also proposed for further study.

7.2 Discussions and proposals for further study

The physical analysis in this thesis work focused on the ACF quotient and the RMS value which related to empirical result of human perception. Echo features are widely referred to in echolocation and room acoustic investigation [27][28][29]. Besides the repetition pitch and loudness which were of main scope in this thesis work, others
(like echo duration and timbre) were also found useful in human echolocation [27][29]. Both human beings and dolphins are reported to use different combinations of echo features that permit object discrimination [27]. In this thesis project, the same object was used in different acoustic conditions. Whether the combinations of echo features which people rely on are also different due to the different acoustic conditions could be investigated in further study. If this hypothesis could be proved true, it will be helpful to explain why people had higher correct percentage of detecting obstacle in the conference room than that in the anechoic chamber [11], since much more acoustic information are available in a conference room.

As a measurement of sound magnitude, RMS value was investigated in this thesis. Loudness is a subjective measurement of a sound’s quantity that is a primary psychological correlate of physical strength [12]. Even if loudness is strongly related to the RMS value of a sound, it still varies for many other reasons. Recent research on psycho-acoustics found that loudness was increased by perceiving the previous sound in the contralateral ear [30]. Research on induced loudness reduction (ILR) found that it can also affect loudness judgments [31]. Investigating the relation between the physical strength of a sound and its psychological quantity could also be useful to explain why people made different judgments on the sound recordings with the same RMS value [11].

Research on spatial hearing made comparisons between monaural hearing and binaural hearing. Both of them were established to contribute to forming the position of the auditory event [32][33][34]. But binaural unmasking was found to play a major role to detect signals using the interaural differences [32][35]. As was seen in the Appendix, the solid line is correlated to the dash line in each autocorrelation figure of the original sound recording. The correlation between them could offer information to human echolocation [36][37]. The acoustic model which was designed to synthesize IRN in this thesis project did not generate interaural differences. This could affect binaural hearing and make the sound synthesis less distinguishable. Improvement
could be done to this acoustic model so that binaural sound synthesis are taken into account.
References


echolocation. American Foundation for the blind. Research Bulletin, 4,
1-13.

perceptibility of pitch. Journal of the Acoustical Society of America, 47,
469-475.


loudness in the perception of obstacles by the blind. The American Journal
of Psychology, 63, 485-515.


ability of the blind: size and distance factors. Journal of Experimental
Psychology, 70(3), 246-251.

the pitch strength of iterated rippled noise. Journal of the Acoustical Society
of America, 99, 1066-1078.

discrimination of broadband signals with rippled power spectra. Journal of
the Acoustical Society of America, 63(4), 1166-1173.

Echolocation: a study of auditory functioning in blind and sighted subjects.

Analysis and sound synthesis for human echolocation


http://www.hearingresearch.org/Dr.Ross/Binaural-Better.htm

- 43 -


Appendix

Comparison between sound synthesis and original sound recordings

Figure A.1. Autocorrelation comparison between synthetic data (left) and original sound recording (right) when distance between sound source and obstacle is 50cm and the sound duration is 5ms
Figure A.2. Autocorrelation comparison between synthetic data (left) and original sound recording (right) when distance between sound source and obstacle is 100cm and the sound duration is 5ms.

Figure A.3. Autocorrelation comparison between synthetic data (left) and original sound recording (right) when distance between sound source and obstacle is 200cm and the sound duration is 5ms.
Figure A.4. Autocorrelation comparison between synthetic data (left) and original sound recording (right) when distance between sound source and obstacle is 300cm and the sound duration is 5ms

Figure A.5. Autocorrelation comparison between synthetic data (left) and original sound recording (right) when distance between sound source and obstacle is 400cm and the sound duration is 5ms
Analysis and sound synthesis for human echolocation

Figure A.6. Autocorrelation comparison between synthetic data (left) and original sound recording (right) when distance between sound source and obstacle is 500cm and the sound duration is 5ms

Figure A.7. Autocorrelation comparison between synthetic data (left) and original sound recording (right) when distance between sound source and obstacle is 50cm and the sound duration is 50ms
Figure A.8. Autocorrelation comparison between synthetic data (left) and original sound recording (right) when distance between sound source and obstacle is 100cm and the sound duration is 50ms.

Figure A.9. Autocorrelation comparison between synthetic data (left) and original sound recording (right) when distance between sound source and obstacle is 200cm and the sound duration is 50ms.
Analysis and sound synthesis for human echolocation

Figure A.10. Autocorrelation comparison between synthetic data (left) and original sound recording (right) when distance between sound source and obstacle is 300cm and the sound duration is 50ms

Figure A.11. Autocorrelation comparison between synthetic data (left) and original sound recording (right) when distance between sound source and obstacle is 400cm and the sound duration is 50ms
Figure A.12. Autocorrelation comparison between synthetic data (left) and original sound recording (right) when distance between sound source and obstacle is 500cm and the sound duration is 50ms.

Figure A.13. Autocorrelation comparison between synthetic data (left) and original sound recording (right) when distance between sound source and obstacle is 50cm and the sound duration is 500ms.
Analysis and sound synthesis for human echolocation

Figure A.14. Autocorrelation comparison between synthetic data (left) and original sound recording (right) when distance between sound source and obstacle is 100cm and the sound duration is 500ms

Figure A.15. Autocorrelation comparison between synthetic data (left) and original sound recording (right) when distance between sound source and obstacle is 200cm and the sound duration is 500ms
Figure A.16. Autocorrelation comparison between synthetic data (left) and original sound recording (right) when distance between sound source and obstacle is 300cm and the sound duration is 500ms.

Figure A.17. Autocorrelation comparison between synthetic data (left) and original sound recording (right) when distance between sound source and obstacle is 400cm and the sound duration is 500ms.
Figure A.18. Autocorrelation comparison between synthetic data (left) and original sound recording (right) when distance between sound source and obstacle is 500cm and the sound duration is 500ms