Published: September 6, 2018

RESEARCH ARTICLE

Citation: Yao M. et al. (2018)
Detecting Frequency from
Randomly Sampled Data
Implementation of random
sampling in BRATUMASS. Science
Publishing Group Journal 1(2).

DOI:

Corresponding Author:
Meng Yao
E-mail: myao@ee.ecnu.edu.cn

Funding:
Shanghai Science and Technology
Development Foundation under
the project grant numbers:
03JC14026, 08JC1409200;
TI Co. Ltd through TI (China)
Innovation Foundation;
ZKGS (Beijing) Co. Ltd.

Copyright:
© 2018 Science Publishing Group
This open access article is
distributed under the terms of the
[Creative Commons Attribution
Non-Commercial License](https://creativecommons.org/licenses/by-nc/4.0/).

Detecting Frequency from Randomly Sampled Data Implementation of random sampling in BRATUMASS

Luxi Li¹, Yizhou Yao², Meng Yao^{1,3}, Erik. D Goodman³

Authors' affiliations:

1. School of Info Science and Tech, East China Normal University, Shanghai, China
2. College of Science and Engineering, Central Michigan University, Mt Pleasant, MI, U.S.
3. BEACON Center, Michigan State University, East Lansing, MI, U.S.

Abstract

In this article, a system which implements random sampling theory is presented—the system obtains each sample after a random time interval, and it is a part of the Breast Tumor Microwave Sensor System (BRATUMASS) which is designed to refine the data of BRATUMASS. The first part introduces the system in the following aspects—the components of the system, how the signal is obtained, and the algorithm we used to calculate the spectrum of non-uniformly sampled data. The second part introduces a set of experimental performances based on random sampling method to explore the features of random sampling, and the signals used in the experiment are single frequency sinusoidal waves, mixed sinusoidal waves, and a piece of bass music waves since the random sampling method is a prototype and not integrated with BRATUMASS yet. The data from BRATUMASS is a uniformly sampled data interpolated with the same mechanism—random time interval interpolation—random time interval interpolation—will be the next step of this study.

Introduction

BRATUMASS is developed to detect breast cancer at an earlier stage, more treatable stage. It is a safe, mobile and cost-effective method compared to the traditional diagnostic method such as soft X-ray mammography imaging and MRI (magnetic resonance imaging).

There was a significant difference in the electrical properties between the normal breast tissue and the mammary gland tissue in the microwave band¹. Based on this permittivity contrast between different tissue, backscatter would happen in edge interface of different tissue. By obtaining and analyzing the backscattered microwave signal, the interface of different tissue could be located. In BRATUMASS, the distance of the object interface is converted into frequency components, and the analysis of the data is significantly important because the target frequency components are merged in noise and are difficult to separate. Moreover, the frequency resolution is important in discriminating different breast tissue. Random sampling—the measuring method we used in frequency screening—is based on compressive sensing theory². It features a signal to information conversion. Due to the limit of information energy in the Nyquist-Shannon theorem, random sampling reduces the number of samples without much perceptual loss. This is ideal for BRATUMASS because, for a limited storage, random sampling fills the storage slower than Nyquist Sampling, and thus extends observation time window for the signal. And longer observation time means higher frequency resolution, which is important in imaging resolution in BRATUMASS. Besides, this quality is ideal in the developing of a portable cost-effective device. More importantly, the signal to information conversion, if successful, would increase the information to noise ratio while processing BRATUMASSs data and thus is meaningful in separating target signal from noise in BRATUMASS. This is a feature of compressive sensing that could improve the imaging resolution of BRATUMASS as well.

This work is to explore the features of random sampling and to prepare for its implementation in BRATUMASS. The Randomly Sampled Data system is based on previous simulation³. On this system, experiments can be performed to explore the features of random sampling using real signal.

Methods

In this section, the BRATUMASS breast tumor detecting system and the Randomly Sampled Data System are to be introduced. BRATUMASS is a system that mapping the distance of objects to frequency components. The introduction of BRATUMASS would be based on its modules and how the system converts the distance of an object into frequency components. The Randomly Sampled Data system is also a module of BRATUMASS that is not yet integrated with it. The purpose of this Randomly Sampled Data system is to implement the concept of compressive sensing and improve the image resolution of BRATUMASS. The implementation of this Random Sampling Data system will be discussed.

The principle of BRATUMASS

For the breast tissues provide permittivity contrast between different tissue, backscatter would happen in edge interface of different tissue. The BRATUMASS obtain the microwave back signal via backscattered off different tissue and in every edge interface. A simplified detection process is showed in Fig. 1. The microwave travels to the permittivity contrast interface of tumor and

normal breast tissue and back to receiving antenna. For the signal is delayed after traveling, the distance of the interface can be obtained through analyzing the back wave signal.

Figure 1: Detecting model of BRATUMASS

BRATUMASS consists of signal obtaining device and data processing platform. The device consists of RF transceiver module, zero intermediate frequency (zero-IF) mixer, slot step frequency modulation signal generator, sampling module and data processing platform on PC. Fig. 2 shows the partial block diagram of BRATUMASS. Let $S_L(t)$ be the transmitted coupling signal, $S_r(t)$ be the received signal, $S_f(t)$ be the filtered zero IF output signal and $S_f(n)$ be the sampled signal.

Figure 2: System diagram of BRATUMASS

On a scan point, transmitting antenna transmits slot step frequency signal $S_L(t)$, Fig. 3 shows the frequency pattern of $S_L(t)$, the frequency is modulated by a triangular step wave. The transmitting signal goes through two paths. One directly goes through transmitting antenna coupling network and reaches one of the two inputs of the zero-IF mixer. Because this signal reaches the mixer input almost instantly, it can be seen as identical to $S_L(t)$. Another transmits into the detection space and backscatters off the tissue edge interface that comes in its way and then reaches the second input of zero-IF mixer through receiving antenna, becoming $S_r(t)$. Because the signal transmits through coupling network reaches the mixer input almost instantly, the delay of the received signal from receiving antenna can be seen as the time that the signal takes to travel from transmitting antenna to target tissues and then to receiving antenna. The output of zero-IF is $S_f(t)$ and its theoretical frequency pattern is shown in Fig. 3. It contains the frequency difference (Δf) of the emitting signal and back wave signal. And from $S_f(t)$, the Δt can be calculated, thus the distance from every edge interface of tissues to the antenna can be calculated. This is a simple model of the system. In the practical situation, the received signal is a

mix of backscattered wave and others. The distance of the target is now converted to frequency components. Hence the frequency resolution is crucial in discriminating targets at the different distance.

In application, 17 scan data from 16 orientations are collected to reconstruct the image of the interior tissue and, if exists, tumor. The reconstruction images were imaged via BRATUMASS imaging platform.

Figure 3: Frequency pattern of signal. (a) Transmitting and Receiving frequency pattern. (b) Output of zero-IF frequency signal.

Glance of Randomly Sampled Data system

This module is based on an Arduino⁴. Timer, ADC module and random number generator mentioned below are all integrated into this chip. The timer is set according to a sequence of random number. ADC module is triggered by the compare match of the timer and the value of compare match register is updated in each of the compare matches interrupt routine. Thus the ADC module is triggered after every random time interval and the time interval between two neighboring sample is controlled by the random number generated by the algorithm. Both ADC data and the random time interval is sent to PC through FIFO to USB module.

FIFO to USB Converter

The MCU of Arduino sends sampling data and its corresponding time interval to FIFO module. FIFO module stores the data from the sampling module before computer reads it. FIFO module also sends the message from computer to sampling module⁵. Through the instruction from the computer, the random number generation in Arduino is controlled by the computer. This module is also a conversion of FIFO data transmission protocol to USB that is common in computer hardware.

Data display and analysis platform on PC

This platform contains the driver for the USB module, real-time waveform display, and data processing. Before the data acquisition, the random number generation mode, as well as the observation time, is set on the computer. After setting a sampling mode and click the start button, sampling starts and will display the waveform within the observation window until save is clicked. Wave data and time interval is saved for later analysis.

Spectrum calculation

The anti-aliasing properties of random sampling make it possible to represent signal from a relatively small set of data. In this experiment, each sample is obtained after a random time interval. Simulation of this method has been done in previous work⁶. The advantage of this method is that aliasing can be avoided while sampling at an average rate below the Nyquist rate. Besides, lower sampling rate means more observation time at a limited data amount. In other words, it takes longer to full a limited storage at a lower sampling rate. In this case, lower sampling rate makes longer observation time and thus increases the resolution of the spectrum. In application, as the sampling time is limited, the signal we obtained is the view through the window. In other words, the signal we obtained is the product of the signal and a window function. The frequency resolution after sampling is $F = 1/T$ (where T is the length of window function). This means signals have frequency difference smaller than F is unable to be distinguished in its spectrum. Thus, for a fixed observation time window T , the frequency resolution is fixed, and the longer the observation time, the higher the frequency resolution. The frequency resolution is related to time window T regardless of the sampling rate. By random sampling, a signal can be represented while sampling at a slower rate (sub-Nyquist rate) thus can be represented by the relatively small set of data without perceptual loss. And for a fixed number of samples, random sampling enables longer observation time and thus improves frequency resolution.

The difference of spectrum calculation between randomly sampled data and uniformly sampled data is the difference of time should take into account while doing integration⁷.

Assuming that $x(t)$ is a band-limited signal, $X_c(f)$ is Fourier transform of $x(t)$, the sampling interval is T , the total number of sample is N , then NT is the total sampling time. Let $x(n)$ be the uniformly sampled data, $x(t_n)\{n = 1, 2, 3, \dots, n\}$ be the randomly sampled data, $X_D(f)$ be the Fourier transform of $x(t_n)$. We have (1) (2) (3):

$$X_c(f) = \int_0^{NT} x(t) e^{-j2\pi ft} dt \quad (1)$$

$$X_c(f) = \sum_{n=1}^N x(n) e^{-j2\pi fn} \quad (2)$$

$$X_D(f) = \sum_{n=1}^N x(n) e^{-j2\pi ft_n} (t_{n+1} - t_n) \quad (3)$$

Random number generation

The random number of three kinds of distribution is used in this experiment; they are Uniform, Normal, and Rayleigh distribution. These random numbers are interpreted as the time interval between neighboring samples. For comparison, Nyquist style samples (identical time interval between neighboring samples) is also taken and analyzed. Therefore we have 4 groups of data for one target signal. HIST of a typical series of the random number used in experiments is shown in Fig. 4. These numbers are generated on data acquisition board and then transmit to PC.

Figure 4: HIST of random time interval. (a)HIST of Uniformly distributed time interval. (b)HIST of Normally distributed time interval. (c)HIST of Rayleigh distributed time interval.

Results

Although perfect in simulation, practices in experiment show limits while distinguishing signal frequency at the low sampling frequency. During the experiment, while sampling frequency is lower than the signal frequency, the spectrum of the signal sometimes failed to distinguish signal frequency or has a higher level of noise frequency components. To ensure stable performance, the experiment results in this work are results when target signal frequency is below the sampling rate. As to the failure of some experiments, while signal frequency is above sampling rate, it may be because that the system itself has a limited timer. The timer is based on its clock and has a limited precision which is different from the random number generated on PC. Because the aim of this work is to optimize the frequency resolution of BRATUMASS, further study of this limitation and its improvements remain to be explored. However, distinguishing of the signal frequency lower than sampling frequency is successful and stable. Experiments and results will be discussed in this part.

Signal and spectrum

Five groups of the signal are selected as target signal in this experiment. They are 233Hz, 678Hz, (233.5+233.6) Hz, (678.5+678.6) Hz sine waves and a piece of bass music wave. As is shown in Fig. 5, first we tried to distinguish a single sine wave, then a mix of two sine waves with close frequency, finally a bass music wave.

Single sine wave

As is shown in Fig. 6, this sampling method is successful in distinguishing single sine wave. Aliasing is suppressed but frequencies with small amplitude are appearing where aliasing frequencies should be. This signal is obtained at an average sampling rate of 800Hz.

Mixed sine wave

In mixed sine wave experiment, a same amount of samples (16000 samples) is obtained from a mixed sine wave of 233.5Hz and 233.6Hz. The average rate of random sampling is 800Hz and the sampling rate of Nyquist sampling is 2000Hz. Thus the time window for random sampling is 20 seconds, and its frequency resolution should be $1/20=0.05\text{Hz}$. While the time window for Nyquist sampling is 8 seconds, so the frequency resolution should be $1/8=0.125\text{Hz}$. In other words, lower sampling rate means longer observation time window for signal when the data amount is limited. And the longer the observation time the higher the frequency resolution. The spectrum of this experiment shows that random sampling shows better frequency resolution at

the same number of samples, as it is in Fig.7 that randomly sampled data successfully distinguished mixed sine wave while Nyquist fails to distinguish two frequency components.

Bass music wave

We also had a bass music wave experiment to explore how this mechanism fits sound wave with rich frequency components in it. This bass music wave is a wave (.wav) file⁸ with the sampling rate of 44100Hz. Original bass music wave and its FFT of is showed in the last line in Fig. 8. Most of its energy is in frequency components within 1000Hz. The entire sampling rate or the average of sampling rate in this bass music experiment is 800Hz.

Figure 5: Signal in time domain. (a) Sine wave in time domain from 0 to 4 seconds. (b) Mixed sine wave in time domain from 0 to 25 seconds. (c) Bass music wave in time domain from 0 to 30 seconds. (d) Sine wave (zoomed in). (e) Mixed sine wave (zoomed in). (f) Bass music wave (zoomed in).

Figure 6: Spectrum of single sine wave 233Hz (green line) and 678Hz (blue line) in one figure. All sampled at an average rate of 800Hz. (a) Sampled by uniformly distributed time intervals. (b) Sampled by normally distributed time intervals. (c) Sampled by normally distributed time intervals. (d) Sampled by identical time intervals (Nyquist style).

Figure 7: Spectrum of mixed sine wave. (a) Spectrum of mixed tone 233.5Hz and 233.6Hz sampled by Uniform (green line), Normal (pink line, and Rayleigh distributed time intervals. Average sampling rate is 800Hz. (b) Spectrum of mixed tone 678.5Hz and 678.6Hz sampled by uniform (green line), Normal (pink line), and Rayleigh (blue line) distributed time intervals. Average sampling rate is 800Hz. (c) Spectrum of 233.5Hz and 233.6Hz sampled by identical time interval. Sampling rate is 2000Hz. (d) Spectrum of 678.5Hz and 678.6Hz sampled by identical time interval. Sampling rate is 2000Hz.

Figure 8: Time and spectrum of bass wave. (a) Bass music wave from 0 to 30 seconds. (b) Spectrum of bass music wave from 0 to 1000Hz

The spectrum from randomly sampled data is no big difference from the signal's FFT, except for the aliasing around 700Hz. The anti-aliasing propriety seems to be weakened when the signal has rich main frequency components, as shown in the Fig. 8. The spectrum of Uniform, Normal, Rayleigh and Nyquist is roughly symmetric at an axis of 400Hz. All spectrums calculated from randomly sampled data is aliasing just like the spectrum calculated from Nyquist style sampled

data. However, the antialiasing propriety of random sampling is not totally gone. The amplitude of higher frequency, which is the aliasing frequency, is slightly lower than that of its symmetric frequency in Fig. 8.

Random number distribution and signal spectrum

In this section, a set of experiments are performed to see the effect of random number distribution on the signal spectrum. A set of mixed tones experiments will be discussed in this section. To explore the effects of random number distribution on the signal spectrum, this experiment has three variables; they are distribution of random numbers, average sampling rate, and sampling frequency range (the average sampling rate $\pm$ frequency range). The variable chosen is shown in Table 1. To evaluate the difference of the random sampled signal's spectrum between the original signal spectrums, the l_2 norm is chosen as a norm to evaluate the difference between two spectrum vectors.

The l_2 norm of a vector is the square root of the sum of the absolute values squared. To a vector $V(v_1, v_2, \dots, v_n)$, the l_2 norm of V , represented as $\|V\|_2$, is calculated by (4):

$$\|V\|_2 = \sqrt{\sum_{n=1}^{n=N} |v_n|^2} \quad (4)$$

The $\|V - M\|_2$, where $V_N(v_1, v_2, \dots, v_N)$ and $M_N(m_1, m_2, \dots, m_N)$, can be seen as distance between vector V and vector M . Thus it is chosen to evaluate the difference of two spectrums in this work. Besides, the l_2 norm is calculated from by sum of the absolute spectrum values squared, thus the l_2 norm is positive correlated to the energy of a signal. This is used to evaluate the energy of the aliasing frequency components.

Table 1: Variables in Sampling

Distribution	Average sampling rate(F_s)	Frequency Range
Uniform	500Hz	$F_s \pm 30\text{Hz}$
Normal	600Hz	$F_s \pm 40\text{Hz}$
Rayleigh	700Hz	$F_s \pm 50\text{Hz}$
Nyquist	800Hz	$F_s \pm 60\text{Hz}$

The target signal in this experiments is a mixed sinusoidal tones of (96 ,97.5 ,98 ,101 ,103 ,106 ,108, 111)Hz with amplitude (0.8 ,1 ,1 ,0.9 ,0.7 ,0.6 ,0.3, 0.2). Its spectrum and waveform is in Fig. 9.

To compare the frequency detecting accuracy and the anti-aliasing proprieties of random sampling, the spectrum of sampled data is compared to the signal spectrum. To evaluate the difference of spectrum after random sampling from the original signal spectrum, a range of the spectrum is selected from these two as two spectrum vectors. The l_2 norm of the subtraction of these two spectrum vector is taken as the difference for evaluation.

As in Fig. 10(b) and (d), the spectrum from 60Hz to 140Hz is compared with the original signal spectrum by subtraction, and l_2 norm of the subtraction is computed to evaluate the accuracy of the spectrum after sampling.

In addition, the l_2 norm of the frequency range from 400Hz to 600Hz, where the aliasing frequency components exist, is computed to evaluate the anti-aliasing properties of random sampling. The l_2 norm of all the experiments results is in Fig. 11.

Figure 9: (a) Spectrum of mixed tones. (b) Waveform of mixed tones.

Figure 10: The signal and spectrum of mixed tones experiments. This figure is mixed tone sampled at average rate of 600Hz, frequency range 100Hz ($600\text{Hz} \pm 50\text{Hz}$) (a)Waveform of mixed tones in 10 seconds. (b)Spectrum of waveform from 50Hz to 150Hz. (c) The difference between the spectrum of the random sampling signal and the spectrum of the uniformly sampled signal (from 50Hz to 150Hz). (d)Spectrum of waveform from 0Hz to 1200Hz. (e)The difference between

the spectrum of the random sampling signal and the spectrum of the uniformly sampled signal(from 400Hz to 600Hz).

Figure 11: Comparison chart of I_2 norm. I_2 norm under the experiments while the average sampling rate, the sampling rate range, the sampling time interval distribution as three variable. (a,c,e,g) Average sampling rate 500Hz(a), 600Hz(c), 700Hz(e), 800Hz(g), time interval distributed as Uniform(green line), Normal(pink line), Rayleigh(blue line) and Nyquist(purple line). X axis is frequency range from 60Hz(500±30Hz) to 120Hz(500±60Hz). And I_2 norm is calculated from the difference of spectrum(50Hz-150Hz). (b,d,f,h) Same as (a,c,e,g) but the I_2 norm is calculated from the difference of spectrum(400Hz-600Hz).

The left column of Fig. 11 is the spectrum comparison of spectrum where the signal frequency components located (from 50Hz to 150Hz). This comparison is to explore the trends of the I_2 norm at the different situation. From (a) to (c), (e) and (g), the average sampling rate is ascending. Though different in each experiment, they are all in range of 0.12 to 0.26. The I_2 norm is even steadier while removing the experiment results of the normal distribution at frequency range 120Hz in sampling rate 700 and 800Hz. But the I_2 norm is smaller when the average sampling rate is higher. This means the accuracy of the spectrum is increasing while increasing the average sampling rate. Four groups of experiments are not enough to tell some of the exact trends of I_2 norm changing with one variable. For some distribution of random sampling, such as uniform distribution, the I_2 norm seems larger while the frequency range is larger.

Refinement of BRATUMASS's Data

The randomly sampled data system is not integrated with BRATUMASS. A uniformly sampled data from BRATUMASS is interpolated with the random time interval to explore the effects of random interpolation upon the signal spectrum.

The signal is obtained at a sampling rate of 500Hz. It is interpolated at an average rate of 5000Hz with the uniformly distributed time interval, as is shown in Fig. 12. The samples interpolated between the original samples follow the linear relationship. A random time interval is generated as a time vector for interpolation location. The interpolated signal is calculated based on this random time vector.

Because the target signal in BRATUMASS is in 0-50Hz, the spectrum between 0 and 50Hz is compared in Fig. 12 (c). The difference of the spectrum is visible by eyes that the difference vector fluctuates. And l_2 norm is used as a standard to evaluate the difference after interpolation. The spectrum of the signal after interpolation deviates from the spectrum before interpolation as the frequency range grows. As is shown in Fig. 12 (c) that the difference of the two spectrum vector fluctuates as the frequency range grows. And the l_2 norm of the difference of the spectrum is also getting larger when the frequency range is larger. The result might be different while using larger frequency range or different average interpolation rate. The difference of the spectrum after interpolation suggests the interpolation has the influence on the spectrum and the possibility of optimizing the spectrum for target locating by interpolation. Yet how different distribution of random numbers affects the signal spectrum is still to be explored.

Figure 12: Signal and its spectrum before and after interpolation. (a)Signal waveform from 0.67 to 0.7 seconds. Black line is signal before interpolation and others are after interpolation. (b)Spectrum of signal before and after interpolation from 0 to 500Hz. (c)Spectrum of signal before and after interpolation from 0 to 50Hz. (d)Difference of each signal spectrum with the spectrum of original signal from 0Hz to 50Hz.

Conclusions

This work implements previous simulation and this implements Randomly Sampled Data system successfully distinguishes the spectrum of the sine wave, the mixed sine wave and the signal composed of rich frequency components such as bass music wave. It shows the advantage of improving the resolution of spectrum, and in the suppressing of aliasing. In addition, it is efficient in the usage of data. This work also explores the effects of different time interval on the spectrum of the signal. Three qualities of the time interval are chosen as variables in experiments; they are distribution of time interval, frequency range, and average sampling rate. The l_2 norm is used to evaluate the difference of the spectrum between randomly sampled data's spectrum and uniformly sampled data's spectrum. After comparison of the l_2 norm under different experiments, the trends of the l_2 norm is not obvious but steady in a range of values. And though the aliasing is suppressed under random sampling, the l_2 norm of the spectrum vector where aliasing exists is also steady, which indicates the aliasing is suppressed but the energy of the aliasing frequencies remains the same. However, how the distribution of random numbers affects the signal spectrum and how to relate the distribution of random interval and the information energy is still to be explored.

Acknowledgements

This work has been performed while Prof. Meng Yao was a visiting Professor in Beacon Center, Michigan State University, thanks to a visiting research program from Prof. Erik D. Goodman. M. Yao would also like to acknowledge the support of Shanghai Science and Technology Development Foundation under the project grant numbers 03JC14026 and 08JC1409200, as well as the support of TI Co. Ltd through TI (China) Innovation Foundation and ZKGS (Beijing) Co. Ltd.

References

1. Zhonglin Han, Zhifu Tao, Meng Yao, B. Fleet, E. D. Goodman, and J. R. Deller. Detection transmission loss model of bratumass. In Image and Signal Processing (CISP), 2011 4th International Congress on, vol. 5. IEEE, 2011:2646–2649.
2. E. J. Candes, M. B. Wakin. An introduction to compressive sampling. IEEE signal processing magazine, 2008 vol. 25, no. 2:21–30.
3. Lucy. Li, Weiming Ji, Yizhou, Yao, et al. Refinement bratumass'data of breast phantom processing based on compressive sensing. In proceedings of the International Conference on Bioinformatics & Computational Biology (BIOCOMP2015):105.
4. A.LLC. (2016) Arduino products. Available. <https://www.arduino.cc/en/Main/Products>
5. F.T.D.I.Ltd.(2015) Ft2232-hi-speed dualusbuart/fifoic.Available <http://www.ftdichip.com/Products/ICs/FT2232H.htm>
6. Zhen Cai. Non uniform sampling used in breast cancer detection system. Master's thesis: East China Normal University Shanghai China; April 2015.
7. A. Wang. Methods of signal frequency measurement and implementation based on nonuniform sampling. Ph.D. dissertation: Huazhong University of Science and Technology Wuhan China; April 2004.
8. freesound.org.(2012) double bass test1.Available. <https://www.freesound.org/people/wescwave/sounds/169885/>