Detection of Human Echo Locator Waveform Using Gammatone Filter Processing

RSA Raja Abdullah, NL Saleh, SMS Ahmad and Asem A. Salah
Wireless and Photonic Networks Research Centre of Excellence, Faculty of Engineering, University Putra Malaysia (UPM) 43400 Serdang, Selangor, Malaysia
r_syamsul@upm.edu.my

NE Abdul Rashid
Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor, Malaysia

***Abstract*—Human echolocation is a technique that is commonly used by blind persons to perceive their surroundings by analysing echo signals using an active signal (often tongue clicks). Over the years, studies into human echo location have explored vibrant disciplines, including the engineering perspective. The studies have been continuous and report on the human echo locator waveforms, which are individually unique, with the existence of multiple frequency components. However, possible explanations as to how blind people detect their own pair of transmitted and echo signals still remain vague. The detection process using the conventional matched filter has led to poor performance probably because the waveform consists of multiple frequency components. It was reported in a recent analysis that an ideal scheme for the detection of a human echo locator waveform click is possibly through the adoption of bio-inspired processing.Therefore, a similar detection mechanism based on a bio-inspired method incorporated with a gammatone filter was proposed in this paper for the transmitted-echo signal pair. The optimal detection output led to an ideal method for the detection of human echolocator signals. Furthermore, the need for alternative signal processing approaches for future man-made sensor systems has placed a demand on researchers to explore the perspectives in new fields of study. As such, the positive results explored in this paper can be beneficial for emerging concepts in new developments in the application of radar and sonar systems in the near future.**

Keywords— Human echolocator, gammatone filter, matched filter, basilar membrane, correlation.

I. INTRODUCTION

Human echolocation is a technique that is usually used by blind people to identify their surroundings by analysing the returning echoes which are actively generated from the person. Human echo location studies were initiated by Supa in 1944 [1]. The experiments revealed that blindfolded participants were able to sense the presence of obstacles by the noise created by scuffing their heels on the floor. In 1967, an experiment on human echo location was conducted by Rice, and it was discovered that hissing sounds and punctuated tongue clicks were the main sources of echo location among participants [2]. More researches in recent years have explored the cognitive and psychophysical factors involved in human echolocation [3][4][5]. This fascinating phenomenon has drawn the attention of scholars to analyse human echolocator waveform clicks, and it has been discovered that they are individually unique with the existence of two frequency component regions [6][7][8].

A recent analysis has demonstrated that the detection of human echo locator clicks is possible [9] by adopting bio-inspired processing. The proposed detection method, which incorporates a gammatone filter (GF), imitates the properties of the basilar membrane (BM) in the synthesis of sound signals. Studies have revealed that the waveforms are individually unique and different compared to bats and dolphins, which also use echolocation on a daily basis. A look at these mammals reveal that their signal properties follow the characteristics of linear frequency modulation.Hence, the conventional method of using amatched filter (MF) for detection purposes will work [10]. Therefore, these findings suggest that the strategy for echo location employed by humans is different from that of bats and dolphins.

Thus, this paper studied the detection performance of the human echolocator waveform using bio-inspired processing (proposed method) and a MF. The bio-inspired method incorporated a GF [11]. For the bio-inspired method, both the transmitted and echo signals were synthesized using GF processing in order to obtain the frequency response components. These frequency component coefficients were aligned into individual rows to construct a set of filterbanks containing synthesized information of the waveform. This concept is believed to imitate the theory of BM in synthesizing frequency components upon the entry of sound signalsinto the human ear. The process continued with the cross-correlation of each row from the set of filterbanks, followed by the obtaining of a single correlation output prior to the detection stage. Last, but not least, the information discussed in this paper can be disseminated and be of benefit to other researchers with regard to emerging concepts for new developments in radar and sonar system applications in the near future.

II. BIO-INSPIRED PROCESSING

A. Human Hearing Theory

The construction of the human hearing system is complex and can be divided into three main regions known as the i) outer ear, ii) middle ear, and iii) inner ear. The system is responsible for the processing and synthesis of sound signals upon entry into the human auditory system. Each region is assigned a different purpose in the processing and synthesis of sound signals. Multiple curved shapes of the pinna (outer cone shape of the human ear) serve their purpose in modifying the sound signals, especially the high frequency components, and are useful during the localization of sounds[12]. Impedance matching occurs in the middle ear, where the sound signals need to be amplified as they travel from one medium to another (air to liquid). Therefore, an adequate amount of energy needs to be relayed to the stapes in order to oscillate the BM in the inner ear at an optimum rate [12][13]. The unique construction of the width and stiffness along the BM varies from base to apex, being wider and softer at the apex and narrower and stiffer on approaching the base. As a result, only a specific region along the BM will oscillate upon receiving the sound signals. This illustrates that human hearing has non-linear properties, and serves as a filterbank, where each channel has a specific bandwidth range. This strategy allows the human auditory system to resolve in detail multiple frequency components from sound signals.

B. Synthesis of Waveform using Gammatone Filter (GF)

The human hearing theory has demonstrated that the unique construction of the BM is meant to serve its purpose in synthesizing the frequency components in detail and acts as a band-pass filter. As a result, only a distinct region along the BM corresponds to specific frequency components. These feedback responses along the BM create a non-linear characteristic upon the synthesis of sound signals. The Equivalent Rectangular Bandwidth (ERB) is a mathematical approximation that is used to represent the measured psychoacoustic width of the filter in the human auditory system, as given in (1),where cf_i is the estimation of the centre frequency index along the BM (assuming a human hearing range of between 20Hz to 20kHz) [11]. Hence, the artificial measurement of the filter width using a GF is specified based on the ERB theory. Thus, the strategy of the BM in synthesizing the waveform can be imitated by implementing the GF processing. The realization of the synthesis of the human echolocator waveform in this paper used two sub-processes, defined as the ERB space and the ERB filter bank. The details of the GF construction can be retrieved here [11].

$$ERB_i = 24.7\left(\frac{4.37 \cdot cf_i}{1000+1}\right)$$
$$i = 1,2,3, \ldots, N \tag{1}$$

The ERB space is responsible for the uniform spacing (logarithm factor) of the EE waveform based on (1). Then, the ERB filter bank is responsible for constructing the rows (channels), i of the band-pass filters (BPF). Each row, Gf_i is known as a frequency coefficient component, which consists of all the parameters given in (2). Now, the output filter bank from the proposed method contains the synthesized frequency response components extending from 20Hz to 20kHz (assuming the hearing range of humans). As a result, the GF processing should be able to resolve the frequency component from the sound signals based on the psychoacoustic theory.

$$Gf_i(t) = \alpha t^{n-1} e^{-2\pi B_i t} \cos(2\pi c f_i t + \theta)$$
$$B_i = 1.019 \times 2\pi \times ERB_i$$
$$i = 1, 2, 3, \ldots, N \tag{2}$$

where a is a constant value controlled by the gain, n is the filter order, which is set to be less than or equal to 4, B_i is the decay factor, which is related to cf_i, and θ is the phase carrier. The frequency responses of the individual channels from the GF are illustrated in Fig.1a, where the non-linear properties of the filter width are increasing in the logarithm factor, as shown in Fig. 1b.

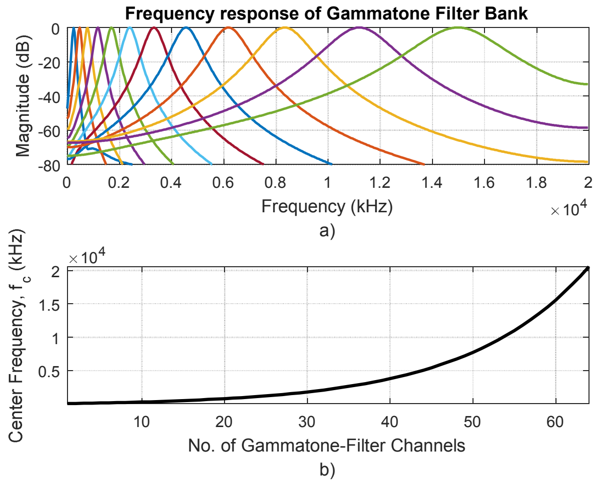

Fig.1. Characteristics of the Gammatone filter, a) Frequency response of Gammatone filter bank, b) Trend of individual channel widths in the Gammatone filter, which increases in the form of an logarithm factor

To date, the actual explanation on the recognition of sound signals is still vague. It has been suggested that a correlation process is involved prior to the sound being transported to the human brain [12].Thus, the correlation of each channel from the Gf_i output is needed in order to obtain a single correlation result by adding up the correlation output from each channel. The output that fulfils the threshold condition is identified as the true pair of transmitted-echo signals.

C. Data Used for Detection

The actual human echolocator tongue clicks were retrieved from [4].The clicks belonged to a middle-aged man (43 years old at the time of recording), who had been blind since the age of 13 months due to retinoblastoma. He was also known as an expert in the human echo location process.In this paper, he was denoted as anecho locator expert (EE). Prior to the experiments, a pair of in-ear microphoneswere placed at his pinna. Then, a straight pole covered with aluminium was placed at a distance of 150 cm facing him. The signal that was present during the active human echolocation process was recorded using binaural in-ear microphones inside the anechoic chamber. The binaural recording consisted of two signals, one at the left and one at the right ear, which were picked up by the microphones at the time of the data collection. More details of the experimental setup and procedure can be obtained from [4]. The transmitted and echo signals were extracted manually using (3) from the raw data, which was stored in Wave Audio File Format (WAV). The extracted transmitted and echo signals from the raw data were used as the input in the chapter, "Bio-Inspired Detection Mechanism", which will be discussed later. The shape of the extracted transmitted waveform is shown in Fig. 2a and that of theecho waveform is shown in Fig. 2b. The Power Spectrum Density (PSD) for the transmitted signal is shown in Fig. 3a, and the PSD for theecho signal is shown in Fig. 3b.

$$t = \frac{2d}{c} \cdot f_s \tag{3}$$

where $c = 342m/s$ (speed of sound in the air), d is the distance between the target and source signal, t is the time of the return trip of the signal from the source signal, and f_s is the sampling frequency that is equal to 44100 kHz.

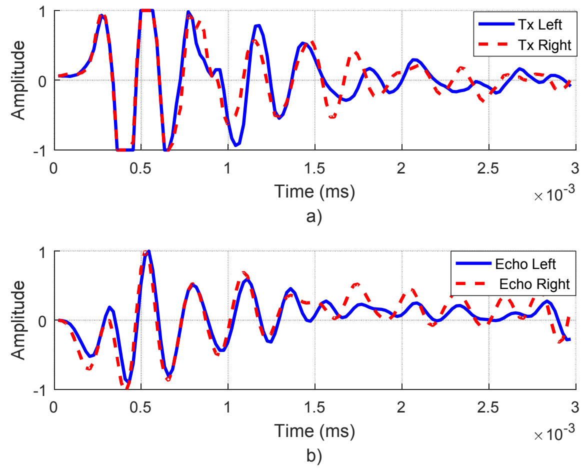

Fig. 2. Time domain of EE signal, a) Transmitted signal, b) Echo signal

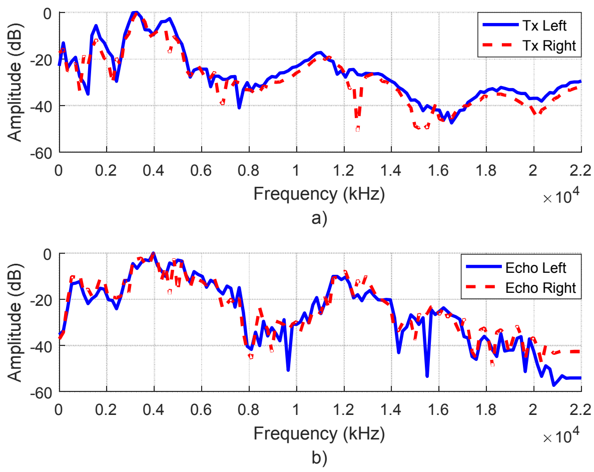

Fig. 3. Power Spectrum Density of EE signal, a) Transmitted signal, b) Echo signal

III. BIO-INSPIREDDETECTION MECHANISM

A detection mechanism was proposed to mimic the closest possible strategy to employ with the human echo location process based on the author's knowledge of the human hearing theory. During an actual echolocation event, the human echo locator is likely to listen to both the transmitted leaked signal (*TxL*) and theecho signal (*Ec*) for them to be synthesized for further action. Therefore, both the *TxL* and *Ec* that had been extracted from the raw data earlier were required as the input for the purpose of detection in this paper. Basically, two phases of processing were involved in the proposed processing scheme,namely, i) the training phase and ii) the testing phase. The summary of the proposed detection mechanism is illustrated in Fig. 4.

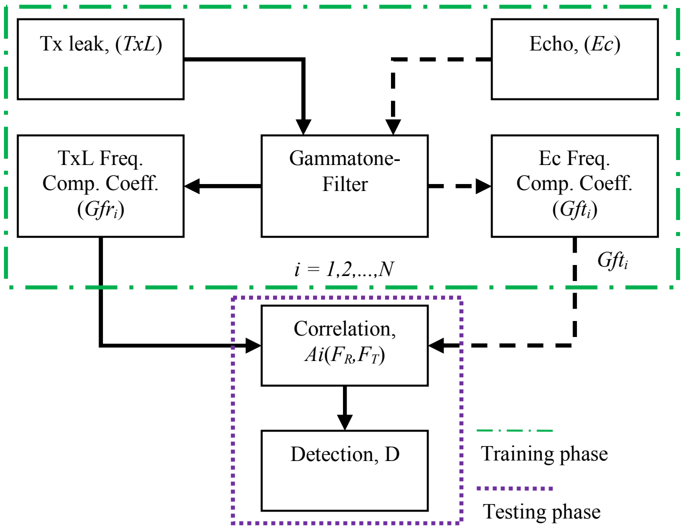

Fig. 4. Bio-inspired detection mechanism for human echo locator click signal

A. Training Phase

Both the *TxL* and *Ec* were passed through the GF in order to obtain the synthesized coefficients known as the frequency response coefficients, which were denoted as Gfr_i and Gft_i.These comprised all the parameters, as given in (2). The Gfr_i and Gft_iwere arranged in rows (channels), i of the band-pass filters (BPF), respectively. Next, these coefficients became the input prior to the testing phase, which involved the correlation of the individual channels from Gfr_i and Gft_i.

It was necessary to synthesize the waveform in such a way so as to allow the GF to resolve the multiple frequency components in greater detail rather than to treat the waveform itself as a wide-band signal. Consequently, the synthesis of the frequency coefficient from each channel only had a single pulse rather than a complex pulse. Therefore, upon the correlation of the waveform from each channel at a specific time, no multiple similarities were found, and this helped to eliminate the phenomenon of multiple local maxima prior to the detection stage. However, this strategy will not work if the MF method, which leads to poor output performance, is used.

B. Testing Phase

The detection process was continued with the use of the correlation technique. Each channel, i from the Gfr_i and Gft_i, was correlated separately, resulting in an array of correlations, A_i as given in (4). This correlation output was needed for the next phase of the detection processing. The correlations of each channel were then summed up to give a single correlation output, which provided information regarding the status of the detection, denoted as D, as described in (5). The output of D that fulfilled the threshold limit was known as the true pair correlation and led to accurate detection.

$$A_i = \int_{-\infty}^{\infty} Gfr_i\,(\tau - t)\cdot Gft_i(t)dt, \quad -\infty < t < \infty$$

$$A_i(Gfr_i, Gft_i),\ i = 1,2,3, \ldots, N \tag{4}$$

$$D = \sum_i^N \frac{A_i}{N} \tag{5}$$

IV. RESULTS AND DISCUSSION

The possible explanation on the ideal detection from the GF processing was the unique approach of the GF processing itself.Each channel with a specific filter width was tuned to a different centre frequency by a given ERB space. This enabled only a specific frequency to pass through the channel and prevented other impulses (frequencies) from getting through. Now, each channel consisted of a single sinusoidal signal rather than a complex waveform. During the individual channel correlation, the output resembled the correlation of a sinusoidal waveform, consisting of multiple local maxima peaks, which were true of the left signal, as shown in Fig. 5, and the right signal, as shown in Fig. 6. Prior to the detection phase, the correlation products from individual channels, A_i were summed up to obtain a single correlation result. As a result, these multiple maxima peaks were gradually decreased as the local maxima was getting dominant and continued to rise to become a peak, while the weak side lobe got suppressed during the summation phase, as shown in Fig. 7.

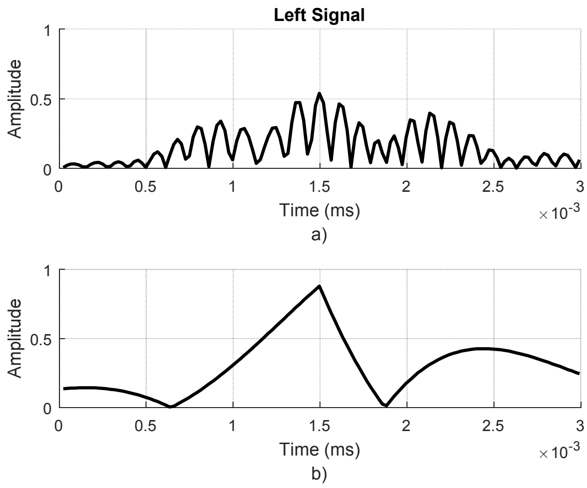

Fig. 5.Correlation of individual channels of Gammatone filter bank from the left signal, a) Output from highest channel order, i = 1 (containing highest frequency component from EE waveform), b) Output from lowest channel order, i = 10 (containing lowest frequency component from EE waveform)

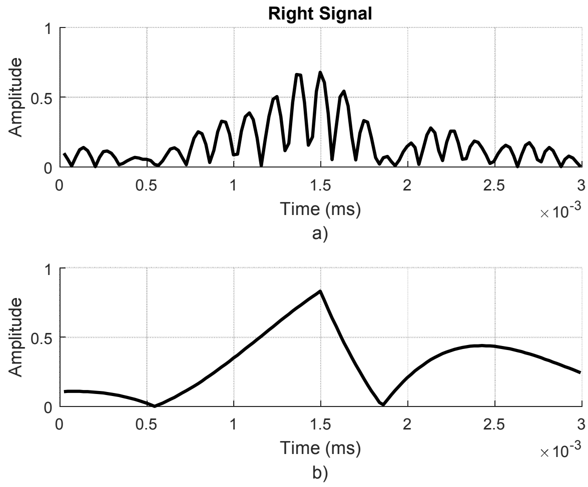

Fig. 6. Correlation of individual channels of Gammatone filter bank from the right signal, a) Output from highest channel order, i = 1 (containing highest frequency component from EE waveform), b) Output from lowest channel order, i = 10 (containing lowest frequency component from EE waveform)

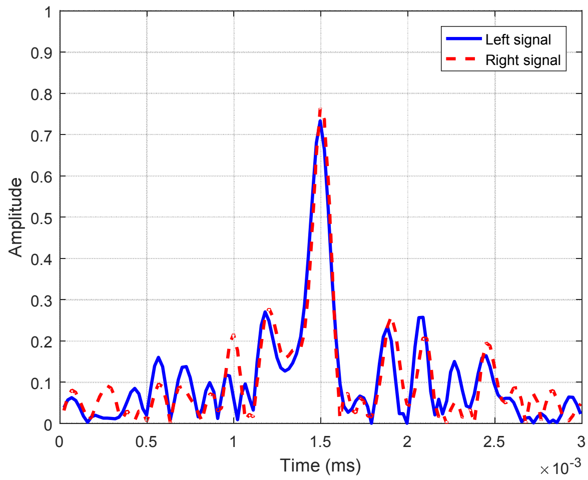

Fig. 7. Summation result of correlation products from individual channels for left and right signals

Next, the detection performance of the transmitted-echo signal pair from the individual processing schemes was observed. The detection results showed that the performance of the proposed method, which incorporated the GF,surpassed the performance of the MF, and this was true at both the left and right signals, as shown in Fig. 8. In daily life, binaural listening helps humans during the localization of sound by providing information regarding the status of the sound signal in a 3D perspective with respect to the human head. Therefore, in this paper, it was important to test both the left and right signals in order to validate the detection performance. Perhaps, the analysis could be extended in future work for a binaural detection strategy.

The proposed method demonstrated an ideal detection performance, where: i) there was a low cross-correlation, which exceeded 10dB, thereby indicating that the features of the main lobes were well-preserved, and ii) the multiple local maxima no longer existed compared to the MF output.

Meanwhile, the MF detection output revealed poor features, where: i) the threshold level between the main lobes and side lobes, which was less than -3dB, masked the actual features from the main lobes, and ii) the existence of multiple local maxima demonstrated multiple signal similarities found during the correlation, which led to false detection and poor performance.

Based on the detection performance using the proposed method, some conclusions can be made: i) the GF processing is likely an ideal strategy to synthesize the human echo locator waveform by estimating the width of the filter in the human auditory system according to the psychoacoustic theory, ii) the logarithm properties of the GF processing have the potential to better resolve the multiple frequency component from the human echo locator waveform, and iii) the correlation of the individual channels in between the transmitted-echo signal prior to the detection stage greatly suppresses the multiple local maxima phenomenon during detection,thereby leading to an ideal detection output.

The trend for alternative signal processing approaches for future man-made sensor systems has placed demands on researchers to explore new study perspectives. Thus, the knowledge that has been discussed so far could be beneficial for new developments in radar and sonar system applications. Next, the advantage of having multiple frequency components is that it allows more details regarding targeted features to be exploited from the waveform rather than by performing a demanding signal compression method in order to achieve such a high resolution. Finally, the preliminary studies that have been discussed in this paper could result in the emergence of a new bio-inspired concept which can be useful for developments in radar and sonar system applications in the near future.

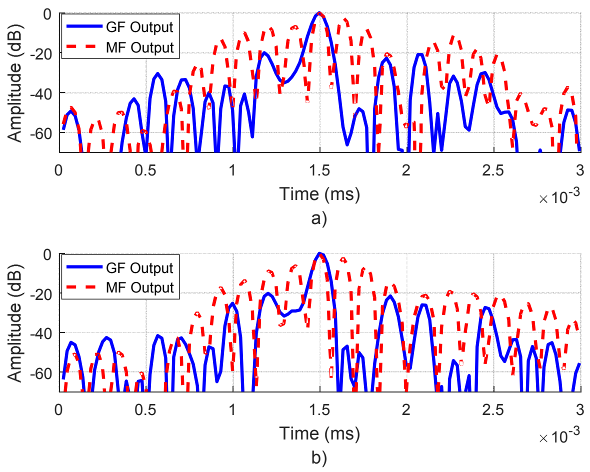

Fig. 8. Detection output between propose method (incorporated GF) and MF processing, a) Left signal, b) Right signal. The result is converted into a normalized representation in order to observe the cross-correlation level, which reflects the overall detection results

V. CONCLUSION

This paper investigated and analysed the detection performance of human echo locator clicks by GF processing and MF processing. The detection technique proposed in this paper involved training and testing phases prior to the final decision stage using both transmitted and echo signals. In the training phase, these signals had to be synthesized by passing themthrough a GF in order to obtain a set of frequency component coefficients, which were arranged into individual channels.

Next, in the testing phase, each channel from the transmitted-echo signal pair needed to be correlated prior to the detection stage. To complete the processing, the correlation products from the individual channels were summed up to obtain a single correlation product for decision making. The detection performance using the GF technique demonstrated that an ideal output was achieved by fulfilling the basic requirement for detection. Meanwhile, the output from the MF method revealed a poor performance by showing a high cross-correlation level, which can cause the system to deliver a false detection.

This evidence indicated that the proposed method is probably an ideal method for the detection of human echo locator signals. By understanding the strategies on how the detection process is conducted by human echo locators, perhaps a similar technique can be adopted and realized in the development of a man-made sensor system. This could give rise to the need for a concept in bio-inspired processing which can be beneficial for new developments in radar and sonar systems in the near future.

REFERENCES

[1] M. Supa, M. Cotzin, and K. M. Dallenbach, "'Facial Vision': The Perception of Obstacles by the Blind," *Am. J. Psychol.*, vol. 57, no. 2, pp. 133–183, 1944.

[2] C. E. Rice, "Human Echo Perception," *Science (80-.).*, vol. 155, no. 3763, pp. 656–664, Feb. 1967.

[3] B. N. Schenkman and M. E. Nilsson, "Human Echolocation: Blind and Sighted Persons' Ability to Detect Sounds Recorded in the Presence of a Reflecting Object," *Perception*, vol. 39, no. 4, pp. 483–501, Apr. 2010.

[4] L. Thaler, S. R. Arnott, and M. A. Goodale, "Neural correlates of natural human echolocation in early and late blind echolocation experts," *PLoS One*, vol. 6, no. 5, 2011.

[5] R. S. A. R. Abdullah, N. L. Saleh, S. M. S. Ahmad, N. E. A. Rashid, G. Reich, M. Cherniakov, M. Antoniou, and L. Thaler, "Bio-inspired radar: Recognition of human echolocator tongue clicks signals," in *2017 IEEE Asia Pacific Microwave Conference (APMC)*, 2017, vol. 1, pp. 861–864.

[6] G. E. Smith and C. J. Baker, "Human echolocation waveform analysis," *IET Int. Conf. Radar Syst. (Radar 2012)*, no. c, pp. 103–103, 2012.

[7] C. J. B. and D. K. Xinyu Zhang, Graeme E.Smith, Lore Thaler , Galen.M.Reich, Michael Antoniou , Mikhail Cherniakov, “Echolocation in humans: waveform analysis of tongue clicks,” *IEEE IET Lett.*, vol. 53, no. 9, pp. 580–582, 2017.

[8] L. Thaler, G. M. Reich, X. Zhang, D. Wang, G. E. Smith, Z. Tao, R. S. A. B. R. Abdullah, M. Cherniakov, C. J. Baker, D. Kish, and M. Antoniou, “Mouth-clicks used by blind expert human echolocators – signal description and model based signal synthesis,” *PLOS Comput. Biol.*, vol. 13, no. 8, p. e1005670, 2017.

[9] R. S. A. Raja Abdullah, N. L. Saleh, N. E. Abdul Rashid, and S. M. S. Ahmad, “Bio-inspired signal detection mechanism for tongue click waveform used in human echolocation,” *Electron. Lett.*, vol. 53, no. 22, pp. 1456–1458, 2017.

[10] M. Vespe, G. Jones, and C. J. Baker, “Lessons for Radar,” *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 65–75, Jan. 2009.

[11] M. Slaney, “An Efficient Implementation of the Auditory Filter Bank,” *Apple Comput. Tech. Rep.*, vol. 1, no. 35, p. 40, 1993.

[12] B. C. J. Moore, “An Introduction to the Psychology of Hearing,” *Brill*, no. Sixth Edition, 2013.

[13] M. Haggard, *Hearing: An Introduction to Psychological and Physiological Acoustics*, vol. 45, no. 12. 1982.