Research Article

Rafed Sabbar Abbas*, Hayder Rafea Kareem Almosa and Yahya Jasim Harbi

Modulation and performance of synchronous demodulation for speech signal detection and dialect intelligibility

https://doi.org/10.1515/eng-2022-0410
received December 01, 2022; accepted January 24, 2023

Abstract: Speech processing is one of the fundamental operations in computer science and it is particularly difficult to process and distinguish speech in different Arabic dialects when background noise is present. In any nation, communication skills are crucial. Pushing a button is all it takes for the typical person to make phone calls and leave voicemails but for telecommunications experts, the process is very different. We understand how communication actually works. The terms detection and demodulation are commonly used when addressing the full demodulation process. The procedures and circuits are substantially the same under both designations. As the name implies, demodulation is the opposite of modulation, which is applying a signal, such as an audio signal, to a carrier. The demodulation process isolates the output signal from the audio or other signal that was transmitted using amplitude shifts on the carrier. In this study, a system for distinguishing speech signals was developed using modulation and demodulation to transmit speech by extracting it from a variety of factors, the most significant of which is background noise in addition to a wide variety of dialects, which poses a significant challenge in speech processing. The proposed system was applied to a dataset that was created for a group of voices in different dialects, and by using important techniques, the noise accompanying the voices was deleted and then the voices were processed with other techniques such as modulation and demodulation to distinguish the dialect. The system has proven effective by distinguishing dialects.

Keywords: speech signal detection, modulation, demodulation, noise, dialect intelligibility

1 Introduction

Frequency modulation (FM) is a nonlinear method of encoding information on a carrier wave. It may be used for many different applications, each with its own statistics that are mostly driven by the underlying producing process, including telemetry, seismic prospecting, and interferometry. Its most widespread application, though, is in radio transmission, which is frequently used to transmit audio signals that represent voices [1].

There are many different types of distortions, noise situations, and other impairments that might affect a communication channel. When a critical threshold is surpassed, the limitations significantly reduce FM demodulator performance. As a result, there is a dramatic decline in the intelligibility and quality of the detected speech; “threshold effect” is the name given to this occurrence [2].

The excitation (source) signal and the vocal tract transfer function (VTTF) are both represented in the acoustic speech signal. In some applications, it is critical to precisely predict the vocal tract transfer and to eliminate fluctuations brought on by shifting fundamental frequency or pitch. For instance, feature extraction for automatic speech recognition (ASR) frequently uses vocal tract transfer [3]. With an all-pole model, the vocal tract transfer is estimated through linear prediction coding (LPC) analysis. However, background noise might interfere with LPC-based features. Similar to this, a smoothed or fast Fourier transform spectrum will be susceptible to background noise. In this study, we present a noise-resistant method for estimating the speech spectrum envelope, which holds data on vocal tract transfer. In the frequency domain, the method resembles amplitude demodulation.

* **Corresponding author: Rafed Sabbar Abbas,** Department of electrical engineering, Faculty of Engineering, University of Kufa, Najaf, Iraq, e-mail: Rafids.abbas@uokufa.edu.iq
Hayder Rafea Kareem Almosa: Department of electronic and communications engineering, Faculty of Engineering, University of Kufa, Najaf, Iraq, e-mail: hayder.almusa@uokufa.edu.iq
Yahya Jasim Harbi: Department of electrical engineering, Faculty of Engineering, University of Kufa, Najaf, Iraq,
e-mail: yahyaj.harbi@uokufa.edu.iq

An audio signal modifies or varies the amplitude, frequency, and phase of a sinusoidal signal through the process of modulation. Modulation methods are often applied to extremely low-frequency sinusoids in the context of audio processing. An amplitude modulation (AM) or phase modulation (PM) of the audio signal might be considered, in particular, to occur when control settings for filters or delay lines are changed. Wah-wah, phaser, and tremolo are common AM and PM examples of audio signal [4]. We will first explain basic methods for AM, single-sideband modulation, and PM and highlight their use for audio effects in order to gain a greater knowledge of the potential of modulation techniques. These modulators may be used to create more complex digital audio effects, as will be shown through a number of examples. We shall discuss numerous demodulators in a later section that isolate the input signal or its properties for additional effect processing [5].

In contrast to other uses, this study uses the word “modulation” differently. For instance, the “modulation spectrum” removes components with a high rate of change by applying low-pass filters on the time trajectory of the spectrum. In this article, a system was developed to separate speech signals using modulation and demodulation to transmit speech by extracting it from various factors, the most significant of which is background noise in addition to the variety of dialects, which represents a major challenge in speech processing. The article has been structured as follows: related work in Section 2, modulation and demodulation overview in Section 3, the proposed approach in Section 4, results and discussion in Section 5, and finally, the conclusion in Section 6.

2 Related work

Decoding radio transmissions and improving speech are typically seen as two distinct issues. However, the statistical characteristics of a measurement method and a previous statistical or deterministic model of the reconstructed signals are typically used to develop effective signal estimating algorithms. This is frequently a challenging topic that lacks an analytical answer and can only be somewhat resolved using oversimplified models of noise and signal. On the other hand, a nonlinear mapping from the input data to the intended output may be thought of as any signal estimate. We may learn such mapping utilizing a collection of study cases, pairs of input-modulated baseband signals, and the desired audio output signals when we have a method for approximating universal functions in our hands.

A technique that resembled amplitude demodulation in the frequency domain was proposed, and its use for automated speech recognition (ASR) was examined, in the study suggested by Zhu and Alwan [6]. The source (excitation) spectrum might be thought of as the carrier and the VTTF as the modulating signal in AM, which is thought to be the mechanism behind speech production. From this perspective, amplitude demodulation might be used to retrieve the VTTF. A nonlinear method that successfully conducts envelope detection by employing harmonic amplitudes and ignoring inter-harmonic dips was used to produce amplitude demodulation of the speech spectrum. The approach was noise resistant because low-energy frequency regions were ignored. The observed envelope was reshaped using the same theory. The method was then applied in order to build an ASR feature extraction module. It was established that this method outperforms Mel-frequency Cepstral coefficients in the presence of additive noise. Peak isolation was also conducted, which increased recognition accuracy even further [6].

A data masking strategy for AM broadcasting systems was suggested in the work given by Ngo et al. [7]. The data cloaking system in the AM domain is implemented using the cochlear delay (CD) digital acoustic watermarking approach they presented before. They looked at the viability of sending extra inaudible messages in AM signals using a CD-based inaudible watermark technique. The suggested method alters the carrier signal using the new duplex modulation, adding the original and watermark signals as lower and higher sidebands, and then sends the modified signal to the receivers. To extract the messages from the watermarked signal and the original signal using the CD-based watermark, special receivers in the proposed technique used dual demodulation to get both the original signals and the watermarks from the received signals. They were able to successfully extract messages from the observed AM signals thanks to the results of their computer simulations, which showed that the suggested approach may send messages as watermarks in AM signals. The outcomes also showed that the suggested technique and the traditional AM radio systems could maintain good sound quality for the demodulated transmissions. This indicates that the suggested method has the ability to function as a daemon transmitter and also has low-level AM radio system compatibility. Emergency warning systems and heavily trafficked AM radio services can both employ the suggested method [7].

A software-defined-radio receiver for FM demodulation that takes an end-to-end learning-based method that makes use of the prior information of the spoken message in the demodulation process was proposed by Elbaz and Zibulevsky [8]. The baseband version of the receiver’s in-phase and quadrature components was used to identify and improve speech. The new system was expected to outperform the existing one in terms of high-performance detection for both auditory disturbances and communication channel noise. There were recognized techniques for low signal-to-noise ratio (SNR) situations in mean square error and perceptual speech quality score evaluation [8].

Orimoto et al. [9] offer a signal identification technique to de-noise genuine voice signals utilizing Bayesian estimates and bone-linked speech. More precisely, a new kind of algorithm for noise removal was theoretically created by adding Bayes' theory, which is based on monitoring speech conducted with air that has been contaminated by ambient background noise. In the suggested speech detection approach, the bone-made speech was employed to estimate speech signals accurately. The application of the suggested technique to air and bone speeches monitored in a real setting with background noise has empirically proven its efficacy.

A multilingual investigation was carried out by Tong et al. [10] (Chinese and English). All of the data gathered demonstrated that the graphite speech sensor had adequate sensitivity to extract acoustic wave parameters. The likelihood of the graphene layer randomly breaking was simultaneously decreased by the proposed cylindrical structure with microsurfaces. Additionally, a neural network was trained using voice data obtained from a microphone and a grapheme elastic sensor for speech identification. The dataset blended with the vocal cord speech signals has 75.9% recognition accuracy. The comparison demonstrated that there was sufficient unique information in the signals picked up by the sensor to carry out the speech recognition tasks .

3 Modulation and demodulation overview

In this section, modulation and demodulation will be discussed along with their uses for differentiating, changing, and extracting speech free of background noise.

3.1 Speech signal modulation

A sine wave is modulated with an information message $x_{\mathrm{m}}(t)$ by a method known as FM [11]:

$$y(t) = A_{\mathrm{c}}\cos\left(2\pi f_{\mathrm{c}} t + 2\pi f\Delta\int_0^t x_{\mathrm{m}}(\tau)\mathrm{d}\tau\right), \tag{1}$$

where $x_{\mathrm{c}}(t) = A_{\mathrm{c}}\cos(2\pi f_{\mathrm{c}} t)$ is the sinusoidal carrier, f_{c} is the carrier's base frequency, A_{c} is its amplitude, and $f\Delta$ is the frequency deviation, which denotes the greatest shift away from the carrier's base frequency. In this example, $x_{\mathrm{m}}(t)$ is the data signal, which is often a voice signal.

Model of noise: A number of limitations can affect the signal during transmission and reception. The modulator's job is to rebuild the original signal from the received signal as reliably as possible on the receiving side, overcoming any limitations imposed by the transmission and reception phases. The message signal experiences a number of distortions, as was previously noted [11]. There are two kinds of signal impairments caused by these distortions:

A. *Phase noise:* Impairments resulting from external factors like audio distortions and FM's operation are converted to phase noise from their original audio additive forms [12]:

$$r(t) = A_{\mathrm{c}}\cos(2\pi f_{\mathrm{c}} t + 2\pi f\Delta\int_0^t (x_{\mathrm{m}}(t) + n(t))\mathrm{d}\tau), \tag{2}$$

where $n(t)$ is the amplitude noise.

B. *Amplitude noise*: Deterioration brought on by distortions in the communication channel, such as convolution with the channel, multi-path, additive noise brought on by the propagation characteristics of the channel environment, etc., result in additive amplitude noise, $r(t) = y(t) + n(t)$ [11].

Each of the noise models mentioned above has a statistical model in communication systems that are often believed to be white Gaussian noise. Figure 1 shows a schematic of the communication system and its components for clarity.

3.2 Speech signal demodulation

The estimation of the VTTF and the elimination of pitch-related information are our objectives. This has to deal

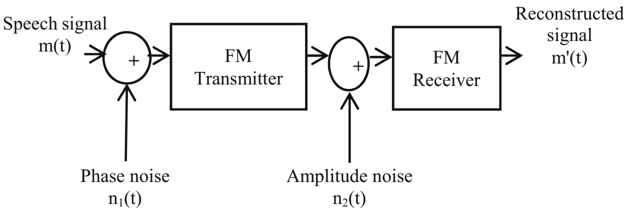

Figure 1: Communication system with sources of phase and amplitude noise.

with frequency domain demodulation of the voice spectrum [13].

Recovery of the carrier signal is necessary for coherent demodulation, which is used in FM radio, for instance. On the other hand, incoherent demodulation uses envelope detection with a rectifier and low-pass filter. Following full-wave rectifying the modulated signal, Figure 2 shows the incoherent demodulation process in the time domain. We will use a similar approach but carry it out in the frequency domain, where the resulting spectrum is "harmonically demodulated" [14].

Different types of FM detectors and demodulators are available. In the past, when radios were composed of discrete devices, some types were more common. However, today, the phase-locked loop (PLL)-based detector and quadrature/coincidence detectors are the most extensively employed since they are the easiest to incorporate into integrated circuits and require few, if any, changes [15].

Normally, the intermediate-frequency (IF) stage may work so that the IF amplifier is forced into limiting in order to enhance the FM receiver's noise performance. By doing this, the noise-causing amplitude changes are eliminated, leaving just the frequency variations.

The following list includes common FM demodulators used in walkie-talkies, portable radios, radio communication systems, broadcast receivers, etc. [16,17]:

A. *Slope detection*: This is a very basic type of FM demodulation that gets its demodulation from the receiver's selectivity. Only when the receiver lacks FM functionality it is employed since it is not very effective. This method of FM detection has a great number of drawbacks, including the receiver's sensitivity to amplitude fluctuations and the radio's selectivity curves complete nonlinearity.

B. *Ratio detector*: When transistor radios employed discrete components, this form of the detector was one that was extensively used. Utilizing a transformer with a third winding was necessary for the ratio detector in order to provide an additional phase-shifted signal for the demodulation procedure. Two diodes, a few resistors, and capacitors were utilized in the ratio detector. The ratio FM detector was a costly type of detector due to the transformer it utilized, despite the fact that it worked effectively. These FM demodulators were expensive to produce since all wrapped components are more expensive than resistors and capacitors, and following the development of integrated circuit technology, when various circuits could be employed, the ratio detector was rarely used. Nevertheless, it did well in its day.

C. *Foster Seeley FM*: This was the other top candidate for the FM demodulator in radios back when they still utilized discrete components. The ratio detector and the Foster Seeley FM demodulator were quite similar in many ways. However, it employed a separate choke rather than the transformer's third winding. The Foster Seeley detector, like the ratio detector, lost popularity when integrated circuits were developed because other types of FM demodulators were much simpler to install with intermediate frequencies (IFs) and had higher performance.

D. *PLL demodulator:* A PLL may be used to demodulate FM. Excellent performance and few, if any, manufacturing changes are needed for the PLL FM detector. Another benefit of the PLL FM demodulator is that it can be readily added to an integrated circuit, which lowers the overall cost of the receiver chip and, ultimately, the radio receiver. Because the PLL was programmed to follow the instantaneous frequency of the incoming FM signal, the PLL FM demodulator worked. The voltage-regulated oscillator inside the loop has to monitor the incoming signal's frequency in order to keep the loop locked. The demodulated output of the audio or other modulation signal was therefore given by the voltage-controlled oscillator, whose tuning voltage fluctuated in accordance with the instantaneous frequency of the signal.

E. *Quadrature detector*: FM radio IFs now frequently employ the quadrature FM detector. It offers exceptional levels of performance and is simple to install. The quadrature coincidence version of an FM demodulator may be applied to an integrated circuit extremely quickly and for essentially no extra cost. Because of this, it is a particularly appealing solution for contemporary receiver designs. A quadrature detector/coincidence detector is

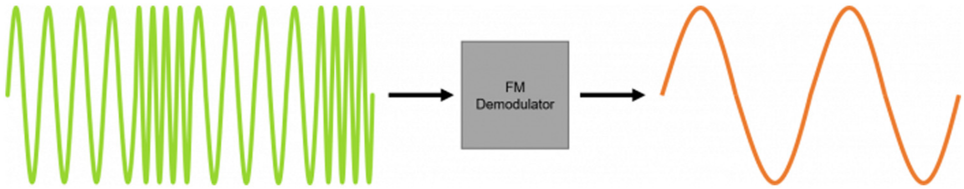

Figure 2: Envelope detector for FM demodulation.

often included in integrated circuits that are intended to perform the functions of a complete receiver or an IF strip. As a result, FM demodulation may be added to the final receiver for almost no additional cost.

There are several uses for these FM demodulators. Depending on the application, designers can choose from a variety of FM demodulator types, including broadcast, two-way radio communications, portable radios and walkie-talkies, high-end communication receivers, and so on [13].

Although PLL-based circuits and the PLL FM detector are the most often used detectors, so are quadrature detectors. Although occasionally still in use, the Foster Seeley and ratio FM detectors are mostly only found in older radios that employed discrete components.

4 The proposed approach

4.1 System dataset

FM stations require high amounts of frequency variation to enable high-quality audio broadcasts. The peak deviation value is set to 75 kHz, while the output signal's sampling frequency is set to 240 kHz, according to US FM broadcasting regulations. The modulating audio signal's default frequency is 48 kHz. Due to the aforementioned factors, the training set was created using Matlab FM modulation with the aforementioned minimum requirements. The amount of baseband samples the modulator generates (five in-phase and five quadrature) for each audio sample on its input is determined by the aforementioned system limits. We presume that conversion from IF to baseband will be handled by additional digital or analog gear in order to avoid directly influencing the FM pass-band signal.

Synchronous detection, also known as heterodyning the signal down to baseband, is the conversion process that is often carried out in the analog front end. The demodulator's computational needs are reduced when the high-frequency signal is converted to baseband, allowing for more convenient processing at a sampling rate lower than the original carrier frequency.

The continuous speech set was utilized to create the sound waveforms for our research. Each utterance is represented by a 16-bit, 16 kHz waveform file in the dataset. We employed male Arab speakers who spoke several dialects. We sampled the baseband signal's phase and quadrature components as two features for the system input. The waveforms were sampled at 48 kHz in accordance with the Arabic standard specification.

4.2 Speech signal modulation and demodulation

The ability to communicate is essential in any country. For the average person, speaking on the phone and leaving voicemails only need the push of a button but for telecommunication engineers, the situation is completely different. We are aware of how communication truly functions. Information is transmitted on a different carrier signal through the process of modulation.

When discussing the entire demodulation procedure, the detection and demodulation of words are frequently employed. The names essentially refer to the same procedure and circuit.

Demodulation, as the name suggests, is the reverse of modulation, which involves applying a signal, such an audio signal, to a carrier.

The demodulation procedure separates the audio or other signal conveyed by amplitude changes on the carrier from the resulting output signal.

AM is frequently used in audio applications; hence, the audio output is frequently used. It is frequently used for land communications for uses related to aviation, including broadcast entertainment, two-way radio communications, and frequently within walkie-talkies.

Although there are many other forms of modulation, we will be using AM for this work. In our current work, we will take the following steps:

- Consider the sound of the dataset, delete undesirable frequencies from the voice signal by filtering it, and use modulation and create a demodulation filter.
- Make signal to pass through the filter, demodulation procedure, and utilize a low pass filter.
- Use modulation and make a demodulation envelope detector and demodulation procedure.
- Utilize a low pass filter, plot time-domain graphs, and signals are Fourier transformed for frequency analysis.
- Shift the origin of the frequency waves and plot frequency domain graphs.

Figure 3 shows a flowchart of the proposed system.

5 Results and discussion

In this section, the most important results obtained will be explained. The dataset we obtained was utilized in Arabic and in several dialects, with varied speech speed and the presence or absence of background noise. The suggested system was created using the Matlab application and codes have been written for three different dialect sounds.

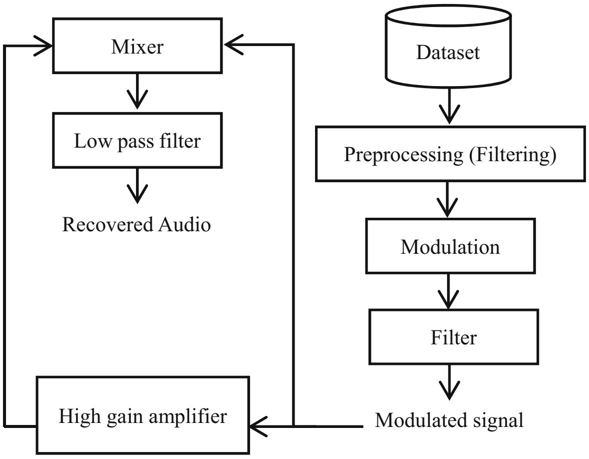

Figure 3: The proposed system flowchart.

The modulated signal varied with time according to the dialect and volume of the signal, which has been proved (Figure 4). In terms of modulation, the interference between low-amplitude frequencies had the least impact on the final signals that were subtracted. High-amplitude frequencies had the majority of the sound characteristics and there was very little interference between them. As a consequence, the interference had no impact on our final results. That has better results compared with the results in Francis [18].

Three sounds with different dialects have been examined in this work. The magnitude in dB has been measured as shown in Figure 5, and the different amplitude values according to the frequency variation. The magnitude has a high value reaching 2.5 × 10^{-3} with 5 kHz and it decreased with high frequency. There was no phase shift during demodulation ($\varphi = 0$), the demodulated signals in sound (2) were identical to the original signal (Figure 5). The demodulated signals (voice signals) are shifted by 10°, 30°, or more degrees in sound (3), attenuating them and making them weaker than the original signals. The voice signals are severely weakened when the shift is 90°, and there appears to be no output speech. Interference between low-amplitude frequencies had the least effect on the final removed signals in our latest study in terms of modulation. The majority of sound characteristics are found in high-amplitude frequencies, and there was little overlap between them. Therefore, the intervention had no effect on our final results.

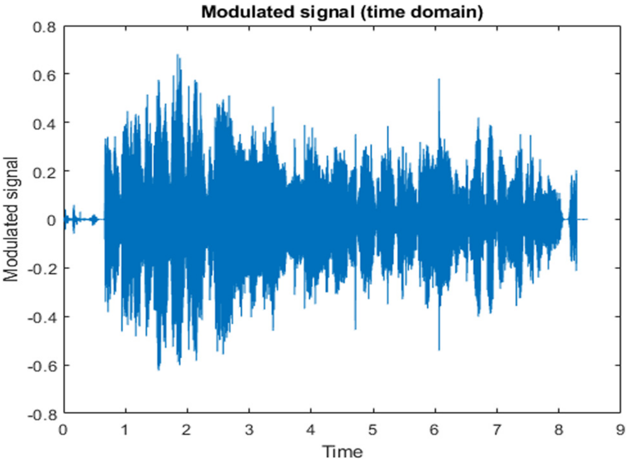

Figure 4: Speech signal modulation with time domain.

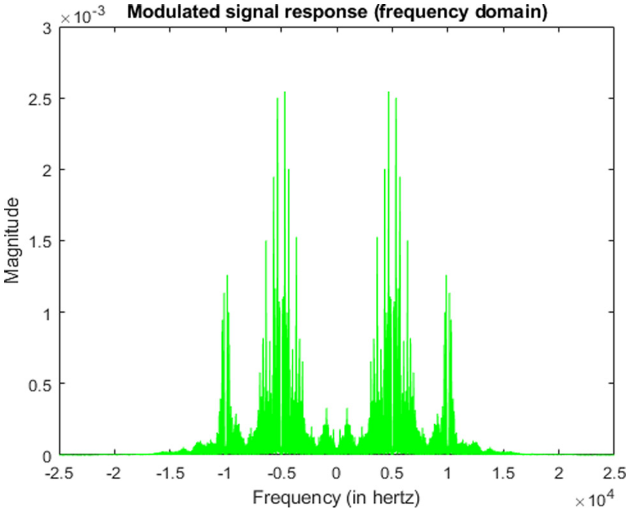

Figure 5: Speech signal modulation with frequency domain.

6 Conclusion

Speech processing is one of the main processors in computer science, and processing and distinguishing speech in different Arabic dialects is a particular challenge because of the difficulty of doing so when there is background noise, and because of the abundance and diversity of Arabic dialects. The Arabic dialects were classified by building a system that relied on modulation and demodulation in two stages. The first is extracting sound characteristics and the second is distinguishing those characteristics according to the proposed system. Speech processing and accent recognition can be improved by introducing deep learning and artificial intelligence (AI) techniques such as neural networks, fuzzy logic, and other AI technologies. Future aspects will examine more dialects with different frequencies and find the SNR.

Conflict of interest: The authors state no conflict of interest.

Data availability statement: Most datasets generated and analyzed in this study are within this manuscript. The other datasets are available on reasonable request from the corresponding author with the attached information.

References

[1] Dutilleux P, Zölzer U. Modulators and demodulators. DAFX: Digital Audio Effects; 31 March 2002.

[2] Hohkawa K. Modulators and demodulators, electrical: Addendum. Digital Encycl Appl Phys. Vol. 33; 2002. p. 3021.

[3] Amini M, Balarastaghi E. Universal neural network demodulator for software defined radio. J Mach Learn Comput. 2011;1(3):305–10.

[4] Fan M, Wu L. Demodulator based on deep belief networks in communication system. International Conference on Communication, Control, Computing and Electronics Engineering; 2017.

[5] Fan Y, Qian Y, Xie F, Soong FK. TTS synthesis with bidirectional LSTM based recurrent neural networks. Interspeech; 2014. p. 19641968.

[6] Zhu Q, Alwan A. AM-demodulation of speech spectra and its application io noise robust speech recognition. Proceedings 6th International Conference on Spoken Language Processing (ICSLP 2000). Vol. 1; 2000. p. 341–4.

[7] Ngo MN, Unoki M, Miyauchi R, Suzuki Y. Data hiding scheme for amplitude modulation radio broadcasting systems. J Inf Hiding Multimed. 2014;5:324–41.

[8] Elbaz D, Zibulevsky M. End to end deep neural network frequency demodulation of speech signals. In: Arai K, Kapoor S, Bhatia R, editors. Advances in information and communication networks. FICC 2018. Advances in intelligent systems and computing. Vol. 886. Cham: Springer; 2019.

[9] Orimoto H, Ikuta A, Hasegawa K. Speech signal detection based on Bayesian estimation by observing air-conducted speech under existence of surrounding noise with the aid of bone-conducted speech. Intell Inf Manag. 2021;13:199–213.

[10] Tong K, Zhang Q, Chen J, Wang H, Wang T. Research on throat speech signal detection based on a flexible graphene piezoresistive sensor. ACS Appl Electron Mater. 2022;4(7):3549–59.

[11] Goehringa T, Bolnerb F, Monaghana JJM, van Dijkc B, Zarowskid A, Bleecka S. Speech enhancement based on neural networks improves speech intelligibility in noise for cochlear implant users. Hear Res. 2017;344(183):194.

[12] Graves A, Mohamed AR, Hinton G. Speech recognition with deep recurrent neural networks. IEEE International Conference on Acoustics. Acoustics, Speech and Signal Processing (ICASSP); 2013.

[13] Hatai I, Chakrabarti I. A new high-performance digital FM modulator and demodulator for software-defined radio and its FPGA implementation. Int J Reconfig Comp. 2011;342532:10.

[14] Kumar A, Florencio D. Speech enhancement in multiple-noise conditions using deep neural networks. arXiv. 2016.

[15] Wornell GW. Efficient symbol-spreading strategies for wireless communication. Cambridge, Massachusetts: Research Laboratory of Electronics, Massachusetts Institute of Technology; 1994.

[16] Xu Y, Du J, Dai L-R, Lee C-H. A regression approach to speech enhancement based on deep neural networks. IEEE/ACM Trans Audio Speech Lang Process. 2015;23(1):7.

[17] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. J Mach Learn Res. 2014;15:1929.

[18] Francis CD. Noise pollution filters bird communities based on vocal frequency. PLoS One. 2011;6:e27052.