Abstract

Generally characteristic of speech waveform is the continuous signal, which contains of voiced and unvoiced signal. Historically, speech waveform is coded by dividing it into frames; it is typically divided into 30 ms frame length, where each frame is coded separately. Speech is however created by a physical system and is substantially shaped by the vocal tract. As it is physically impossible for the vocal tract to move instantaneously from any state to any given state, trends should exist between successive vocal tract positions. In the coding techniques used in this paper, the vocal tract positions manifest themselves as Vector Quantised LSP coefficients.

Although speech coding is an entity in its own right, strong links exist between image compression and speech compression. In this work, the Address-VQ technique which used in the image compression arena, have been applied to the compression of speech coded parameters. Furthermore the technique, called Neural Address Prediction, which is a lossy technique, also applied to encourage further reduce the bit rate. This work exploits the repetitiveness of the attribute of a single speaker to further reduce the bit rate. Preliminary results indicate that approximately more than 33% additional compression is achievable using Neural Address Prediction with Address Vector Quantisation codebook. As Neural Address Prediction is a lossy compression scheme, the error of prediction directly affects to the quality of synthesis speech especially in the voice frames.

1. Introduction

Speech still exists the easiest way of communication between humans which a speaker must produce a speech signal in the form of a sound pressure wave that travels from the speaker’s mouth to a listener’s ears. However naturally the pressure wave generates from the mouth, sound also emanates from the nostrils, throat, and cheeks. Therefore speech signal is composed of a sequence of sounds that serve as a symbolic representation for a thought that the speaker wishes to relay to the listener. Hence speech signal is also one of the complex signals for engineering to study and apply the technologies to it [1]. Basically the speech signal is the continuous signal which contains of voiced and unvoiced properties signal. The major different of both signal is the energy and the form of the signal. Voiced signal is the high energy or amplitude speech signal and also has the form of repetitiveness in itself while the unvoiced signal is contradictory characteristics with voiced. It is low energy and also do not have the format on itself. The experiment in this paper is focused to use the characteristic of repetitiveness in voiced signal to encode and reduce the speech signal.

Encoding speech signal is the main idea and operation in speech coding system. Recently speech coding technologies has been developed over the past three decades and also made possible the achievement of many goals in speech applications in telecommunications bandwidth at various bit rates from 64 kbps down to 2.4 kbps which has become many applications for many international communication system standards.
such as telephone and mobile phone [2]. The appropriate of a particular speech coding algorithm in any applications is mainly selected by the requirement of speech quality and the bit rate. The ideally of the speech coding algorithm is high quality and low bit rate. However encoded speech signal does not used for transmission in communication only but also attempts to use them within propose of storage as well. This work is specifically directed toward investigating and developing the novel encoded speech signal technique to further reduce the bit rate. The model of this work had been experimented of the Code Excited Linear Predictive (CELP) coder [3] [4].

Speech coding algorithm strongly related to the speech compression and image compression. The technique namely Address Vector-Quantisation, which used in image compression area, has been applied in this work in the particularly to extract the repetitiveness of speech characteristic and create the codebook for contain those property. Furthermore the Neural Address Predictor scheme is also used to learn the characteristic of that repetitiveness in codebook and also performed as predictor scheme for propose to reduce the bit rate. In the following of this paper, the detail of Address Vector Quantisation and Neural Address Predictor are introduced. The performance of those techniques within characteristic of speech signal will be presents. Finally the results are shown the efficient of Neural Address Prediction applied to Address Vector Quantisation codebook in speech processing.

2. Address Vector Quantisation

Generally speech waveform is produced by a physical system of human and is substantially shaped by the vocal tract to generate a symbolic sound representation the communication information from an initiator to a recipient. As it is physically impossible for the vocal tract to move instantaneously from any state to any given state, trends should exist between successive vocal tract positions. To model the speech signal, it is required to divide the signal be a frame which in each frame still contains the spectral characteristic of speech waveform. The Linear Predictive (LP) Analysis algorithm [5] is widely used to model the frame of speech signal which the speech signal is represented in the form LP coefficients. The prediction coefficients attempt to model the spectral characteristics of each input speech signal frame where the different of spectral characteristics in speech signal is depended on the moving position of vocal tract.

For every position of vocal tract, a different consecutive set of coefficients will be produced. In this work the LP analysis is used to model the speech signal into the format of Linear Predictive coefficients (LPCs) and converted to Line Spectral Pairs coefficients (LSPs) [6]. However, speech signal consists of voiced and unvoiced waveform which voiced waveform is possible to be repetitiveness of moving vocal tract position. Using the advantage of repetitiveness especially in this work, it is hence necessary to measure and extract them.

2.1 Measuring the repetitiveness

The process of measurement and extract the repetitiveness of speech characteristic signal, namely LP-VQ [7], is shown in figure 1.

The prediction coefficients are sequentially presented to the LP-VQ encoder, after the speech signal has been through LP analysis. LP-VQ is a process of quantisation or classification which many similar sets of prediction coefficients are represented by a single set of coefficients. To achieve of this process, the LP-codebook is created which consists of the most popular set of coefficient within codebook.

The codebook search is instigated, where the computed coefficients are compared with all the codevectors in the codebook to find which codevector is the most similar to the coefficients. The aim is to find which codevectors offers the minimal distortion to the coefficients and then substitute the computed coefficients with that codebook entry. The method that is used to measure the distortion in this paper is Euclidean distance measure [8] which can be used to determine the closest codevectors.

Once the process complete, the coefficients will be replaced by the codevectors in LP codebook. At this stage it can be seen that the repetitiveness of speech signal which is formed in the coefficients, had been measure and extract to be
class within LP-codebook. The similar characteristics of speech signal will be located in the same class and also represent those coefficients with only a set of codevector. The output of LP-VQ is the consecutive of index LP-VQ codebook. Regarding to characteristic of speech signal, thus the consecutive of indexes exist the correlation in each another. This work studies and exploits the correlation of speech characteristics for further coding the speech, thus the Address Vector Quantisation technique is investigated to use.

2.2 Address Vector Quantisation codebook

Initially designed to take advantage of repetitions that exist within an image compression, Address-VQ is a powerful technique in reducing the bit rate of an image, without any loss in quality [9] [10]. In this work Address-VQ has been exploited both the relationships between these consecutive codebook indexes and the probability distribution of a group of indexes occurring together. The structure of Address-VQ codebook is shown in figure 2.

However, the total number of unique address makes this impractical and typically more than 100000 unique combinations are stored when the LP-VQ codebook size is 128. Obviously, the size of the Address-VQ codebook makes an exhaustive search impractical and also the number of bits required to represent each codebook address would be too great. This is the matter to make the system too complicate. Thus solution to solve this problem is to separate the codebook into two regions, one being addressable (active region) and the other being non-addressable (non-active region), and use block transition probability matrices to re-order the Address-VQ codebook contents. When the system operates, the re-order of probability combination addresses are inserted into the Address-VQ codebook in the section of active-region and continued sequentially into non-active region followed order of most popular till popular less. The non-active codebook can be thought of as a large database, which contains popular less combination. This part of the codebook is completely invisible to the actual coding system and hence has no influence on the overall bit rate. In other hand, the ordered most popular combinations are in the active region which is smaller region and also is addressable to the system. Therefore there are three main factors to consider when design the Address-VQ codebook; the size of LP codebook, the size of the active region and the number of using the same unique address in active region.

As it can be seen from figure 2, the first codebook is designed using the LP-VQ algorithms, while the second codebook, named the Address-codebook, consists of the unique addresses. The unique address is generated by combining of consecutive address of LP-codebook which there is four consecutive addresses in each unique address. The output of LP-VQ scheme is analysed and each consecutive four addresses are removed and stored in the Address-VQ codebook as an individual codevectors. The Address-VQ codebook is assumed to contain all possible combinations of addresses that can be occurred.

However the Address-VQ codebook requires the time computation and the storage of repetitiveness codevectors. The process of Address-VQ commonly extracts the repetitiveness and creates the codebook, which consists of most popular used of consecutive combination address LP codebook. To use more advantage of relation in consecutive unique address in active region, the Neural Address Predictor is applied to work with Address-VQ codebook.

3. Neural Networks

Neural Networks [11] is the artificial intelligent system which has been developed and applied to perform complex computational tasks such as in linear and nonlinear system. [12]. A neural network operates as parallel distributed information processing system. They are composed of a large number of simple processing units, namely neuron, which each neuron is connected and created to be a network. In this paper, the novel technique of neuron networks has
been introduced and applied to work as predictor with Address-VQ codebook for speech signal.

3.1 Neural Address Predictor

The Address-VQ technique has been associated the most popular unique address codebook and inserted into the active region. Neural Address Predictor has exploited the associated of repetition speech frames to further reduce the bit rate. Actually the propose of the Neural Address Predictor is to reduce the number of actual address codebook (index codebook) transmissions by exploiting from previous and present address codebook to predict the forthcoming address as show in figure 3.

This can be achieved if the common positions of consecutive address codebook can be learnt or trained.

Figure 3. The general structure of prediction address

At the first step of this technique, Neural Address Predictor needs to learn the input pattern, and then neural will be able to predict the target. In order to each index codebook consists of codevectors inside which each individual codevector also has a correlation to all other codevectors in codebook. This correlation is called as a distance, \( \text{Dist}_A(x) \), which can be measured by Euclidean Distance function. The novel technique in this paper uses the relation of the distance for generate the learning process for Neural Predictor.

The training of the predictor is in condition of supervise learning, which the target matrix is produced to let the networks to be learn and predict as less as error.

Figure 4. Neural Address Predictor structure

The structure of the training phase is shown in figure 4, the training data is defined as \( a_{i,j}, a_{i,j+1}, a_{i,j+2}, a_{i,j+3} \) until \( a_{i+n,j}, a_{i+n,j+1}, a_{i+n,j+2}, a_{i+n,j+3} \) and the distance between each addresses codebook is present as the different of the present codevectors and previous codevectors as \( D_{n-(n+1)} \). Each training data element consists of four popular addresses used from the Address-VQ codebook.

As the training of neural predictor follows the supervised learning structure, the accuracy of predictor is depended on the characteristic of training data and amount of neural. To prepare the data for neural networks training, the both distance of codevectors between \( a_{i,j} \) and \( a_{i,j+1} \) defined as \( D_{a_{i,j}-a_{i,j+1}} \) and between \( a_{i,j+1} \) and \( a_{i,j+2} \) denoted as \( D_{a_{i,j+1}-a_{i,j+2}} \) are computed. The target distance \( D_{a_{i,j+2}-a_{i,j+3}} \) is also computed.

When the training data is firstly presented to the training process, the neural predictor will be trained or learnt to get information of training data. The neural predictor will then compute as estimation target (\( P_i \)), which used the knowledge of the actual target. Generally the output of training process is the predictor target. At this stage, the objective of predictor is to provide the most accuracy of predictor target when compare to the actual target. In order to this work using the distance of actual address codebook to be input of neural predictor, which the interval of distance amount of codevectors in each address codebook is so small. By this reason this technique will then naturally offer minimise error of prediction for speech characteristic.
4. Implementation and results

The speech data input is generated from four speakers: two males and two females. Each speaker contributed a total of 90 minutes of speech, of which 60 minutes was used to be initial learning system data and the final 30 minutes was used to be the testing system data. The 60 minutes speech was fed through an LP analysis to convert the speech characteristics into the form of LP coefficients. In this work, the LP coefficients were formatted in two types: LPCs and LSPs. Then it was presented to the LP-VQ scheme to measure and classify the similar characteristic speech into the LP-VQ codebook. The results of this technique are shown in figure 5.

![Figure 5 Statistic of LP-VQ Codebook journeys of Male1, Male2, Female1, and Female 2.](image)

The size of LP-codebook was design as 128 in this work. The results show that the same address LP-codebook was used by the group of similar characteristic speech signal. Therefore the frames of speech signal, which failed in the same address, were presented by the codevectors of that address. The output of the LP-VQ scheme was a large string of consecutive addresses LP-codebook which represent all the speech frames. The consecutive address was then analysed to extract all the unique combinations of address, which each unique address was contained of four LP-VQ addresses. At this stage, two information of unique address were recoded in a database: the address combinations and the number of times used of each address combinations. The reordered process was operated to arrange the sequence of all unique address under the condition of most popular used. The re-ordered unique addresses were then contained in the Address-VQ codebook in the section of active region only. In this work, it was experimented in the various size of Address-VQ codebook. The results of this stage are shown in figure 6.

![Figure 6. Performance of difference size of Address-VQ codebook](image)

Regarding to the results, it shows that the large size of Address-VQ codebook were contained the unique addresses more than the small one. However, the larger size also uses more time to search and uses more memory to contain the database. Moreover, the number of time used of each unique address was also considered as well. By all reason, the size of 128 was chosen to be address size in Address-VQ codebook.

To take more advantage of repetitiveness LP Address in unique addresses, the neural address predictor was applied to use it. Generally the neural address prediction used the actual address to learn and predict but in this work it is different.
The new technique of neural address predictive used the relation of distance in each address codebook in unique address to be training data for neural. The process of this technique was introduced in section 3. After the training process, the neural was ready to operate as predictor for the testing speech data.

The 30 minutes testing speech data was provided to the system for examination the performance of the system. Some of results of testing are shown in figure 7-10.

![Figure 7. CELP VS NN predictor in Male 1 voiced](image1)

![Figure 8. CELP VS NN predictor in Male 2 voiced](image2)

![Figure 9. CELP VS NN predictor in Female 1 voiced](image3)

![Figure 10. CELP VS NN predictor in Female 2 voiced](image4)

The results were compared to the original speech waveform which can be analysed that the predictor speech waveform is lose the information in amplitude but in the frequency domain the predictor speech still remain the properties similar to the original. Furthermore, the quality of predictive speech was tested under the Mean Opinion Score (MOS) which still existed in the level of fair.

**5. Conclusion**

This paper is introduced the new technique and illustrated the efficient of Neural Address Predictor applied to Address-VQ codebook which used the repetitiveness characteristic of speech signal for further reduce the bit rate. The system can reduce the bit rate because the LP-codebook address is transmitted to Address-VQ instead of the actual coefficients. Also the Neural Address Predictor can be predict the forth index in each unique address; this means the unique address is transmitted only three index instead of four. This system was generated base on the Fs 1016 CELP model which used 34 bits per frame to transmit the LSP coefficients. Using these techniques in this paper, it can be reduce bit rate of Fs 1016 CELP more than 33% when the size of LP-VQ codebook is 128, the size of unique address is 4 indexes. However there are some further work that needs to be study such as the various size of unique address, the difference size of LP-VQ Codebook and also the number of neuron that create the networks.

**6. Reference**


