Optimal Non-Uniform Sampling by Branch-and-Bound Approach for Speech Coding

SAKSHI PANDEY, AMIT BANERJEE Member, IEEE
Computer Science Department, South Asian University, New Delhi, India (e-mail: amit@cs.sau.ac.in)
Corresponding author: Amit Banerjee (e-mail: amit@cs.sau.ac.in).

ABSTRACT Speech coding plays a significant role in voice communication and improving network bandwidth efficiency for applications that require long-distance communication or storage space utilization. Non-uniform sampling (NUS) is a technique for the same, which performs data reduction by sampling at irregular intervals. In the literature, researchers use the structural property of the speech waveform for studying various NUS methods, such as LCSS, MMD, IPD, and zero-crossing point. However, in this paper, we consider the speech signal’s statistical properties to propose an optimal NUS approach. The proposed technique statistically analyzes the speech signal to sample the abrupt changes over a time frame and approximates the signal with minimal reconstruction error using cost and linear penalty functions for avoiding the over-fitting problem. The proposed technique further performs the optimization using the branch-and-bound. To evaluate the proposed NUS, we design a speech waveform encoder called Block Adaptive Amplitude Sampling (BAAS). A BAAS encoder can directly perform statistical analysis on the speech waveform to select data samples corresponding to the most significant changes in the signal. The decoder approximates the eliminated values using linear interpolation. We experimentally study the proposed technique using various matrices and measures such as POLQA and MUSHRA test. The evaluation shows that the proposed NUS technique retains only 25% of data samples to get an acceptable quality signal regeneration. In addition, comparative studies with MMD and IPD show that the proposed algorithm performs 1.6% better with 30% lower MSE scores.

INDEX TERMS Branch-and-Bound Approach, Linear Approximation, Non-Uniform Sampling, Speech Encoding, Statistical Analysis

I. INTRODUCTION

Speech coding is a process of obtaining a compact representation of a speech signal for efficient transmission over band-limited wired and wireless channels and/or for efficient utilization of the storage space [1]. Speech coding is required for many practical applications, including long-distance communication [2], high-quality speech storage [3], and message encryption [4], VoIP, voicemail, and voice response systems [5]. Recently, researchers have also used speech signals for medical [6] and marketing purposes [7]. The speech coders are classified into three broad categories, i.e., waveform coders, parametric coders, and hybrid coders, as illustrated in figure [1]. The parametric coders, synthesize the speech signal based on sound source properties while the waveform coders [8] rely on the shape of the waveform rather than the sound source. The hybrid coders combine the feature of both to produce a medium-rate, high-quality speech signal.

According to [20], waveform coding is a preferred technique for encoding speech signals, as it retains the naturalness of the signal by preserving its shape. It samples the speech signal by removing the redundancies and preserving the shape of the waveform. The waveform coders can perform both uniform and non-uniform sampling. The uniform sampling technique samples the signal at a fixed rate ignoring the variation in the signal, which at times results in processing trivial samples. Thus, uniform sampling often suffers from high redundancies [21], [22], which increases the complexity of the system in terms of bandwidth, storage, computational capabilities, and power consumption for processing the signal [23], [24]. Moreover, with advancements in technology like Internet-of-Things (IoT), the smart devices are increasingly exchanging voluminous data over...
modern wireless communication channels. A non-uniform sampling (NUS) can be a solution to the above problem. Thus, in the paper, we focus on a non-uniform waveform coder. NUS (variable sampling rate) relies on the idea that linear segments can approximate any signal in a short time frame to remove triviality in the sampled data. The various applications of NUS includes efficient computation of atmospheric turbulence spectra [25], mechanical scanning radar [26], non-invasive blood pressure (NIBP) oscillometry [27], a discrete-time sampling rate converter [28]. However, none of the above approaches use audio data.

NUS is a popular and widely studied technique in the field of speech coding also. There are many approaches for signal reconstruction using non-uniform samples like, zero crossings points [29], random points [30], level crossing points [31], local maxima and minima [32], or the inflection points [33]. The main limitation of random sampling is that it cannot skip more than two consecutive samples, which may result in redundancy. The Level crossing detection (LCD) is based on pre-determined thresholds, while sampling the speech signal at local maxima and minima (MMD) results in imprecision due to under sampling [32]. To overcome this problem extra samples are added based on a pre-determined distance value between the samples [34]. Further, in the inflection point detection algorithm (IPD) [33] the extra points added along with the local maxima and minima are points of abrupt changes in the slope of the signal. The idea is to exploit the geometrical structure of the speech signal. Researchers have also studied NUS in amplitude domain. The authors in [35], presents an iterative algorithm to recover a bandlimited signal from samples generated by amplitude sampling. In amplitude sampling the amplitude of the signal is quantized i.e. samples are only taken whenever the signal crosses a predefined set of amplitude levels while the time is represented with infinite precision [36], [37].

Recently, researchers also proposed adaptive level cross sampling (LCSS) for the next-generation data-driven applications [24] and mobile applications [23]. The LCSS technique only samples the amplitudes crossing the fixed threshold level, i.e. the samples above a pre-defined threshold are non-uniformly spaced along the time axis. In all the above approaches, the sample selection methodology relies on the structural property of the waveform, and to the best of our knowledge, few attempts are made to optimize the number of samples used for reconstructing the speech signal. In [34], the authors analyze the quantization properties of non-uniform sampling to optimize the speech samples using dynamic programming. The paper emphasizes quantization, i.e., on the optimal number of samples and not the perfect reconstruction of the signal.

However, in this paper we propose an optimal non-uniform sampling method based on the changes in the shape of a signal by studying its statistical properties in the amplitude domain to reduce the bit rates and retain the advantages of irregular sampling. Compared to the previous approaches, the proposed approach selects the time instance (or change points) of sudden changes in the amplitude of a signal that affects the marginal distribution of its statistical properties like Mean, Standard Deviation, and the linear statistical parameter (LSP, considering both mean and slope). The intuition is to capture time intervals where the signal differs from its statistical property within an upper bound. For this, we define a cost function(C) which minimizes the number of point selections based on the statistical property changes. In other words, the proposed approach analyzes the speech signal statistically to sample the abrupt changes over a time frame and reconstruct the speech signal via interpolation. The signal is approximated with minimal reconstruction error using the cost function(C) and a linear penalty function to avoid the over-fitting problem.

More specifically, the proposed methodology samples the time instance where the amplitude distribution of the signal differs with respect to at least one of the statistical criteria such as mean, variance, or slope. The approach relies on approximating the signal with minimal reconstruction error, as shown in equation 1. The approach also defines a cost function C and a linear penalty function to avoid the over-fitting problem. To further remove redundancies in the sam-
We evaluated the performance of BAAS for various statistical parameters like mean, variance, and standard deviation. The results show that the number of samples retained by the proposed methodology depends upon the selected statistical parameter. That is, mean and standard deviation retains 90% and 50% sample points of a signal on an average, respectively, whereas a linear combination of both mean and slope retains only 25% sample points. Furthermore, we study the quality of BAAS using the five-point MOS (mean opinion score) scale using POLQA for various statistical parameters [41], and also evaluate the subjective quality MUSHRA score [42]. We also compare BAAS with other non-uniform time sampled waveform coders [32]–[40]. The evaluation shows that the number of change points selected by the proposed methodology is significantly less to the previous methods, and the reconstructed signal is a close approximation of the original sound signal.

The rest of the paper is organized as follows. Section II presents the related work. Section III presents the methodology of the proposed non-uniform sampling approach. In section IV, we discuss the experimental setup and simulation results. Finally, section V concludes the paper.

II. RELATED WORKS

Figure 1 shows a broad classification of speech coding techniques. Parametric coders analyze the speech signal to derive the acoustic parameters like pitch, fundamental tones, transient, harmonics, and uses them to re-synthesizing the speech signal. The output speech signal from these coders is an approximation of the input speech signal [43]. The main advantage of a parametric coder is that it re-synthesizes a reasonable quality signal at a very low bit-rate [44]. On the other hand, the waveform coders attempt to code the shape of the speech waveform with minimal error, ignoring the details of the speech production mechanism and speech perception. These are low complexity coders that can produce a high-quality speech signal with a high data rate (≥ 16 Kbps [44]). The hybrid coders combine the advantages of both parametric and waveform coders to produce a medium-rate high-quality speech signal. However, according to [45], the waveform coder is widely used as it can produce high-quality speech signals at low computational complexity.

| Works                  | Sampled point       | Description                                      |
|------------------------|---------------------|--------------------------------------------------|
| Fjallbrant T. [30] [1977] | Random Points       | Based on skipping uniform samples                |
| Mark J. et. al. [31] [1981] | Level Crossing     | Based on pre-defined threshold values            |
| Farokh A. et. al. [29] [1987] | Zero Crossing      | Based on points where waveform changes sign      |
| Budaes et. al. [32] [2005] | Local Maxima & Minima(MMD) | Based on peak and valley points in waveform |
| Greitans et. al. [38] [2011] | Combined peak and level-crossing sampling scheme | Based on peak and pre-defined threshold level in waveform |
| Byeong-Gwan [33], [39] [2014,2019] | Inflection Points Detection (IPD) | Based on the geometrical shape of the waveform |
| R. Viswanadham et. al. [23] [2018] | Level Crossing Sampling | Based on pre-defined threshold & number of bits M decide the quantization |
| V. Ravuri et. al. [24] [2021] | Adaptive Level Crossing | Based on pre-defined threshold |

This paper Abrupt Changes in the Statistical Property Optimization by branch-and-bound

TABLE 1: Summary of the existing works on non-uniform sampling techniques.
A. WAVEFORM CODERS WITH NON-UNIFORM SAMPLING

Since the proposed methodology is a waveform coder, in the following we focus our discussion on the same. The most prevalent works on the waveform coders are based on frequency [11], [12] and time-domain [13]–[15]. The waveform coders commonly use uniform time sampling to retrieve the exact shape of the speech signal. The disadvantage of time sampling is that the memory requirement increases with the increasing size of the signal [46].

As an alternative to uniform sampling, researchers proposed various non-uniform coding algorithms, like polynomial predictor [47] and interpolator using PAN-algorithm [48]. The non-uniform sampling in [47] is based on the observation that a signal may be approximated by a combination of linear segments in a small interval of time. Similarly, in [30], authors propose a non-uniform sampling technique by eliminating random sample points obtained from uniform sampling to perform data reduction. The algorithm establishes a linear relationship between known and estimated samples, to reduce it to sets of coefficients with which the known samples can be weighted to recover the skipped samples. Although the technique proposed in [30] can reduce the sampling rate (below the Nyquist rate) but the method works with a restriction of not skipping more than two consecutive samples. In [31], authors propose a signal reconstruction method by sampling the data at level crossing points (LCD) by defining threshold levels (or predetermined levels). This pre-determination in the LCD method results in roughly sampled data for a rapidly changing signal between the predefined levels.

Later, in [32], [49], authors approximate the speech signal reconstruction by using a non-uniform sampling technique that uses local extreme values (local maxima and minima (MMD)). The primary goal of [32] is to establish the importance of the local extrema values for the speech signal. The reconstruction is performed using a sequence of complex numbers called a skeleton, a pair (amplitude, distance). The study defines the limit of quantization for amplitudes and distances to preserve the intelligibility of the reconstructed speech signal. The method results in an imprecise sampling of the signal [33].

In order to reduce the imprecision and overcome the sparseness problem in [32], extra sample are inserted in [34]. These extra samples are selected using a predefined threshold, i.e., if the distance between the two consecutive sample locations is above a threshold, a mid-point is added. Recently in [33], [39], [40], the author proposes the concept of the inflection point in non-uniform sampling, i.e. instead of adding extra samples to overcome the sparseness problem of the MMD method, the algorithm uses the geometrical structure of the speech signal. For this author’s samples, the local maxima-minima and the points where the slope of the signal changes abruptly. The reconstruction of the speech signal is performed using interpolation. The non-uniform sampling techniques in [33], [39], [40] are the variants of the inflection point detection (IPD) method that mainly use the inflection points, i.e., the points where the slope of the signal changes abruptly.

In another approach, level crossing sampling is combined with the peaks of the signal [38]. The authors sample only the points closely located to the signal peaks. The reconstruction of the speech signal is done via linear interpolation. Recently researchers proposed the level cross NUS method for A/D converters to reduce the bandwidth and computational complexity (for various data processing applications by reducing the sampling rate of the signal). The technique’s sampling is directly proportional to the activity in the signal, viz. the sampling frequency is higher if the signal activity is more and vice-versa. V. Ravuri et. al. [24] and Senay et. al. [50] proposes an adaptive level cross sampling scheme (ALCSS), challenging the traditional uniform sampling analog-to-digital (A/D) converters as they sample at a constant rate which results in oversampling, making the data processing computationally intensive. Table [I] gives a brief overview of the current state-of-art non-uniform sampling algorithms.

B. SUMMARY

In comparison to the above approaches, the proposed methodology considers the statistical change points of the speech signal to perform non-uniform sampling. Figure[1] and Table[1] shows the distinction between the proposed and previous methodologies. Our intuition for using the statistical change points for NUS is to detect the lack of homogeneity (in terms of statistical distributions) in the sequence of a speech signal to find an optimized number of change points for reconstructing an acceptable quality speech signal. As, each selected change point is checked for a bound condition to reduce redundant samples, resulting in better compression. The approach further optimizes the selected sample points using the branch-and-bound solution to include minimum sample points based on the cost function (C). As suggested in [45], the statistical change points can model time-series data, which can be useful for application areas such as medical condition monitoring, climate change detection, speech and image analysis, and human activity analysis.

III. PROPOSED METHODOLOGY

As mentioned above, non-uniform sampling is a technique of selecting the samples of time-series data at irregular intervals that further approximates the original signal by a sequence of polynomial functions [47]. In the following, we discuss the proposed non-uniform sampling methodology and its optimization using the branch-and-bound approach. The approach analyzes the speech signal statistically to sample the abrupt changes over a time frame. The selected data samples are optimized to reconstruct the speech signal via interpolation. The sampling of discrete-time instances is performed with an assumption that the statistical property remains constant between the consecutive data samples. Also, we discuss the extension of the proposed non-uniform
sampling methodology to design a waveform coder, called block adaptive amplitude sampling (BAAS).

A. STATISTICAL ANALYSIS TO PERFORM NUS

In the following, we start with a broad definition of the statistical change points, as instances where the statistical property of the signal changes significantly. That is, given a time series data \( x(n) = \{x_1, x_2, \ldots, x_N\} \), a change point occurs at instant \( \tau \), if the distributions of \( \{x_1, x_2, \ldots, x_{(\tau-1)}\} \) and \( \{x_{(\tau+1)}, x_{(\tau+2)}, \ldots, x_N\} \) differ with respect to at least one criterion such as Mean, Standard Deviation or Linear property (Mean and Slope). For sampling, we consider the fact that the distribution of \( x_i \) changes at unknown time instances \( t_1 < t_2 < \cdots < t_k \), where \( t_1 \) is the first instance of change, \( t_2 \) second and so on. Although the characteristics of a signal can change suddenly at unknown time instance \( t_j \), but within the interval \( (t_{j-1} : t_j, (1 \leq j \leq k)) \) it is bounded by a very small threshold \( (th) \). These time instances can be sampled to approximate the source signal using interpolation.

For example, considering Mean as the statistical parameter, we show binary segmentation of a signal in figure 2. In the figure the vertical line marks the value of \( \tau \) i.e. the change point for the signal and the statistical property remains within a threshold \( th \), for the intervals \( \{x_1, x_2, \ldots, x_{(\tau)}\} \) and \( \{x_{(\tau+1)}, x_{(\tau+2)}, \ldots, x_N\} \). The change point \( \tau \) is the instance that satisfies the following condition. In the following, \( (th) \) is a threshold used for selecting the change points of a signal.

\[
\frac{1}{\tau} \sum_{i=1}^{\tau} x_i - \frac{1}{N-\tau} \sum_{i=\tau+1}^{N} x_i \leq th \quad (2)
\]

Similarly, we can consider equation 3 as the Linear statistical property (LSP), by combining the Mean and Slope of the signal. The change point \( \tau \) can be calculated using the equation 4. The equation 3 calculates the LSP of segment \((t_1, t_2)\). A change-point is selected, if the difference between statistical estimates above a threshold.

\[
L(x_{t_1:t_2}) = (t_2 - t_1) var[(x_{t_1} \cdots x_{t_2})] - \sum_{i=t_1}^{t_2} (x_i - \mu(x_{t_1} \cdots x_{t_2}))^2 (i - \mu(t_1 \cdots t_2)) \\
= \frac{(t_2 - t_1) var[(x_{t_1} \cdots x_{t_2})]}{2} \geq 2\sigma \quad (3)
\]

Note that the threshold \( (th) \) in the above equations depends on the signal and the statistical parameter. It can be determined as the average of the minimum values in the segment. In other words, it depends on the rate of the variability of the signal with respect to the studied statistical parameter.

In the literature, similar definition of change points is used for ECG anomaly detection [51], machine anomaly detection [52], climate change detection [53] and several other applications [54]. As shown in equation 1 and 2, the change points are usually selected based on a pre-defined threshold value, which is application-dependent. In this paper, we use the concept of change points to perform non-uniform sampling of the speech signal and use the sampled data to design a waveform coder.

B. NAIVE STATISTICAL NUS

For minimizing the threshold \( (th) \), we associate a cost function \( (C) \) with every change point. This is important for reducing the redundancies in the sampled data. The cost function \( C(x_{t_1:t_2}) \) defined in the interval \( [t_1 : t_2] \), can be computed as the sum of deviation of each point in the interval from the selected statistical estimate \( (x_i^t_{t_1:t_2}) \) in the interval. That is, \( C(x_{t_1:t_2}) = \sum_{i=t_1}^{t_2} (x_i - (x_i^t_{t_1:t_2})) \). To find a change-point in an interval \([1 : n]\), we divide the signal in two parts at some instant \( \tau \) such that, sum of the cost of first and second segment \([1 : \tau]\) and \([\tau + 1 : n]\) is less than the total cost of the signal, as shown in equation 5.

\[
C(x_{1:\tau}) + C(x_{(\tau+1):n}) < C(x_{1:n}) \quad (5)
\]

Similarly, we can use equation 5 to find all \( K \) change points of a signal by using a bottom-up approach. As a base case, we consider the first data point as a change-point and iteratively add/remove the subsequent points using the cost function stated above. The \( K \) change points results in \( K + 1 \) segments of a signal, with an \( i^{th} \) segment containing all points in the interval \( x_{(i-1)+1:t_i} \).

Notice that equation 5 can sample all the data points in the worst case, particularly for a rapidly changing signal. To control the number of change points, we introduce a linear function \( \alpha g(K) = \alpha K \) in the equation above, where \( \alpha \) is a constant, which controls the rate of growth of the linear function and in turn controls the number of change points. The problem of finding \( K \) change points can be formulated using equation 6, where \( \alpha \) is a penalty added to the cost function \( (C) \) each time a new change point is detected.
Algorithm 1 Naive Statistical NUS

1: procedure MINSAMPLES(x)  \triangleright Input \((x_1, x_2, \cdots, x_n)\) time series audio sample
2:       \(k \leftarrow 0\)
3:     for \(i = 1, 2, \ldots, n\) do
4:          \(H(i) = \min_{t \leq \tau \leq \tau + 1}[H(t) + C(x_{t+1}; t; \tau)] + \alpha\)
5:         \(t' = \arg\min_{t \leq \tau \leq \tau + 1}[H(t) + C(x_{t+1}; t; \tau)]\)
6:         \(cp(k) \leftarrow t'\)  \triangleright Update change pts
7:       \(k \leftarrow k + 1\)  \triangleright Increment k
8:     end for
9:     return \((cp(n))\)  \triangleright return the change pts
10: end procedure

Algorithm 2 Optimization by Branch-and-Bound

1: procedure BBOpt(cp, \(t'\))  \triangleright Input \(cp\) is the list of change points, \(t'\) currently selected change point
2:       \(cp(0) \leftarrow NULL\)
3:     for all \(t \in cp(t - 1)\) do
4:         if \(H(t) + C(x_{t+1}; t; \tau) + \alpha < H(t')\) then
5:             Add \(t\) to the set \(cp'(t)\)
6:         end if
7:     end for
8:     \(cp(i) = cp(i - 1) \cap cp'(t)\)
9: end procedure

10: procedure MINSAMPLES_BOUND(x)  \triangleright Input \((x_1, x_2, \cdots, x_n)\) time series audio sample
11:     for \(i = 1, 2, \ldots, n\) do
12:         \(H(i) = \min_{t \leq \tau \leq \tau + 1}[H(t) + C(x_{t+1}; t; \tau)] + \alpha\)
13:         \(t' = \arg\min_{t \leq \tau \leq \tau + 1}[H(t) + C(x_{t+1}; t; \tau)]\)
14:         \(cp(i) = BBOpt(cp(i - 1), t')\)
15:     end for
16:     return \((cp(n))\)  \triangleright return the change pts
17: end procedure

Notice that if \(\alpha > 1\), the total cost function \(C\) increases significantly, resulting in redundant sampling and vice versa. Thus, \(\alpha\) is a linear function used to optimize the number of non-uniform data samples. In the experimental section, the cost function \(C\) is evaluated for various statistical parameters like Mean, Variance, Standard Deviation, and Slope in the signal.

\[
\sum_{i=1}^{K} C(x_{(t_i-1)+1}; t_i) + \alpha g(K)
\]  \(\text{(6)}\)

In equation \(7\), \(H(s)\) is the minimization of equation \(6\). The intuition is to choose a time instance \((\tau)\) such that the cost function \(C\) reduces with respect to the deviation from the statistical estimate of the signal, as given in equation \(5\). The idea is to reduce the cost function whenever a new change point is introduced to the sampled data set. Next, we discuss the elimination of certain sampled data points using the branch-and-bound approach to improve computational efficiency and remove the redundancy in the signal.

\[
H(s) = \min_{\tau \in T} \left\{ \sum_{i=1}^{K} C(x_{(t_i-1)+1}; t_i) + \alpha \right\}
\]  \(\text{(7)}\)

C. OPTIMIZATION BY BRANCH-AND-BOUND

To further optimize the non-uniform sampled data, we propose a branch-and-bound solution to remove the redundancies in the sample. The branch-and-bound algorithmic paradigm solves the combinatorial optimization problems. Each selected state of the solution is checked against the bound conditions defined in the algorithm to determine the next state of the optimal solution for the problem \([55]\). In the algorithm, we use the concept for optimizing the number of sample points. For this, we assume that the selected subset of the change points at any instance is the current state of the algorithm. Based on the optimality principle, we re-write equation \(H(s)\) as equation \(8\) in equation \(8\). In equation \(8\), \(H(t)\) represents the cost of a subset of the change points selected in the previous state, \(C\) is the cost of adding a new change point, and \(H(s)\) is the cost of the current state.

\[
H(s) = \min_{t} \{H(t) + C(x_{(t+1); n}) + \alpha\}
\]  \(\text{(8)}\)

Next, we define the bound condition for eliminating the overfitting points. For this, let us assume that for some time instance \(\tau\) the selected change points are \(T_\tau = \{t_1 < t_2 < \cdots < t_\tau < \tau\}\). That is, the current state of the algorithm is \(T_\tau\). To determine if the next data point \((\tau' = \tau + 1)\) can be included in the sample, or not, we use the bound condition as shown in equation \(9\). The point is included in the sample if the cost of the next state \((H(\tau'))\) is less than the cost of the previous state.

\[
H(\tau) + C(x_{(\tau+1); \tau'}) + \alpha < H(\tau')
\]  \(\text{(9)}\)

Algorithm \([1]\) shows the implementation of the proposed non-uniform sampling technique by selecting the change points based on the statistical deviation. It uses a time-series signal \(x\) as input to generate the set of change points \(cp(n)\).
The algorithm begins with initial values $cp(0) = NULL$ and $H(0) = -\alpha$, where $\alpha$ as a constant. In each $i^{th}$ iteration, the algorithm selects the change point $t'$ with minimum deviation from the statistical estimate as a non-uniform data sample (step 4 of algo[1]. Algorithm [2] optimizes the problem of sample overfitting of algo-[1]. The procedure MinSamples_Bound is similar to algo-[1] however it checks the previous state $(cp(i-1))$ and newly selected $t'$ against a bound condition (step 14 in algorithm [2]) to optimize the sample data. The procedure BBOOpt of algorithm [2] decides the inclusion of the newly selected change point by checking for the bound condition (step 4 in algorithm [2]), and is used for the optimizing of the selected change points.

D. BLOCK ADAPTIVE AMPLITUDE SAMPLING (BAAS)

To evaluate the proposed non-uniform sampling methodology, we design a waveform coder for the speech signals. The non-uniform sampling is performed in the amplitude domain based on the statistical changes in the signal’s amplitude property. One of the challenges in designing an encoder using the proposed methodology is to define the block size so that the sample points can reconstruct an acceptable quality speech signal. Another challenge is to determine the statistical parameter that can effectively reconstruct the signal.

Algorithm 3 BAAS Encoder

1: procedure BAAS_ENCODER(signal)
2: signal, $f_s$ ← read audio file
3: denoised ← Denoise the signal using DWT
4: $fsize$ ← size of the frame, $f_s/2$
5: $nof$ ← number of frames of frame size $fsize$
6: for $i$ ← 1 to $nof$
7: $frames(i)$ ← get the samples of length $fsize$
8: $cp(i)$ ← MinSamples_Bound ($frame(i)$)
9: $amp(i)$ ← amplitude corresponding to $cp(i)$
10: end for
11: for $i$ ← 1 to $cp$
12: $j$ ← $cp(i)$
13: $bitsream$ ← add 0 from current index to $j-1$
14: $bitsream(j)$ ← add 1 followed by 16-bit binary number of amplitude
15: end for
16: return (bitsream)
17: end procedure

Algorithm 4 BAAS Decoder

1: procedure BAAS_DECEDER(bitsream)
2: for $i$ ← 1 to $bitsream$ do
3: if $bitsream(i) == 1$ then
4: $encodedsignal$ ← Extract change point
5: end if
6: end for
7: $i$ ← 2
8: while $i <= length(encodedsignal)$ do
9: $slopes(i)$ ← slope of $cp(i - i)$
10: end while
11: for $i$ ← 1 to $length(slopes)$ do
12: $p$ ← $encodedsignal(i)$
13: $q$ ← $encodedsignal(i + 1)$
14: $k$ ← 1
15: for $p$ to $q$ do
16: $y$ ← $slope(i) \times k + p$
17: $k$ ← $k + 1$
18: end for
19: return ($y$) ≥ $y$ is the decoded speech signal
20: end for
21: end procedure

1) BAAS Encoder

Encoding is a technique of converting the source data into symbols which can further be used for communication or storage purposes. In the following, we discuss the proposed non-uniform sampling technique to design a waveform coder. Figure[5] gives a general flow of the proposed waveform coder by considering the transmission of the speech signal over the network between two parties, i.e., the sender and receiver.

The encoder selects the data samples that correspond to the most significant change in the signal using the algorithm proposed above. The selected data samples are the encoded signal which sends over the network. At the other end, the receiver decodes the signal using linear stylization. Although we begin with denoising the signal, it is important to note that the analysis can be done directly on the raw speech signal, as well. 
Algorithm 3 shows the steps followed by the encoder to compress a speech signal. Firstly, we de-noise the input signal using the discrete wavelet transform (DWT). The coefficients of the transformed signal are changed appropriately into the wavelet signal according to their local features. Eliminating all the coefficients below the threshold value. Finally, reconstruct the signal from the remaining coefficients using the inverse wavelet transform. To retrieve the stationarity of the signal framing is performed in the subsequent step, a speech signal can be considered as stationary in a short period. The frame size selected is $f_s/2$, where $f_s$ is the sampling frequency of the speech signal. Finally, the encoder performs non-uniform sampling, discussed in section III-A, to reduce the size of the input signal.

Figure 4 shows the structure of a BAAS speech encoder. At the transmitter end, the encoder detects the change points using algorithm 2 and transmits the time instance of change points and the corresponding amplitude (16 bits). For the rest of the samples, only the time instances are transmitted. To distinguish between the two cases, we use the flag field. Figure 5 shows the structure of the bit-stream, where flag ‘0’ shows the interpolation point, and flag ‘1’ is the change point followed by a 16-bit binary number (the corresponding amplitude). For example, consider a bitstream generated by an encoder $< 1 \{a_1 \ 1 \ a_2 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ a_3 \}$, where 1/0 represents the flag and the $a_i$ (16 bits) are the amplitudes of the selected change points. The bitstream begins with flag 1, which represents that the next 16 bits is the amplitude of a change point. Similarly, the next flag bit (following $a_1$) is 1 and shows that the next 16 bits (of $a_2$) is a change point. Following this, the flag field in the bitstream contains a sequence of 6 zeros and the next change point is represented by flag 1 and amplitude $a_3$.

We can also modify the proposed BAAS encoder to design a fixed-rate waveform encoder. For this, we fix the number of samples extracted in the frame using the algorithm 1. The algorithm adapts the selected subset by pruning the excess samples using equation 9. The input to the algorithm is the raw speech signal. The algorithm iteratively selects the change points for each frame using the changes in the statistical parameters, as discussed in algorithm 3. The retrieved change points and the corresponding amplitude are stored in an array for transmission or storage for later usages.

2) BAAS Decoder

The encoded speech signal is approximated using linear interpolation between two consecutive non-redundant sample points [47]. In the literature, researchers have studied various ways of approximating a signal, including quadratic approximation [32], spline approximation [30], linear approximation [59]. However, in this paper, we use the linear approximation for decoding the original speech signal. The resulting waveform results in a saw-tooth shape in the voiced segment with a better approximation. The intervals of silence and unvoiced have randomness, represented by a straight line with zero slopes.

Algorithm 4 shows the steps followed by a decoder to reconstruct the speech signal. On receiving the bitstreams, the decoder reconstructs the speech signal. For flag 1, the amplitude of is obtained directly from the next 16 bits of the bitstream, to extract a change point ($a_{\text{amp}}$, $a_{\text{cp}}$), where $a_{\text{cp}}$ is the time instance and $a_{\text{amp}}$ is the amplitude at that instant). However, if the flag 0, the decoder performs interpolation to generate the intermediate amplitudes. The algorithm uses the straight-line equation to approximate the intermediate points between the change points.

![Figure 6](image-url)

**FIGURE 6:** Parameters used for block size selection by the proposed BAAS encoder.
IV. EXPERIMENT AND RESULTS

For the experimental evaluation, we use the speech samples from LibriSpeech ASR corpus [56] and EUSTACE speech corpus [57]. The LibriSpeech is a corpus of English speech containing approximately 1000 hours of audio samples of 16kHz, prepared by Vassil Panayotov with the assistance of Daniel Povey. The EUSTACE is the Edinburgh university speech timing archive and the corpus of English.

A. PARAMETER SELECTION

In this section, we study the effect of block size on the performance of the proposed methodology. The block size ensures that the number of samples selected is optimal, i.e., they do not inherit the waveform’s redundancy. Also, we investigate the performance of various statistical parameters and their effect on the reconstructed signal’s quality. Selecting the best statistical parameter for the proposed methodology such that the decoded speech signal is of acceptable quality or partial data approximation does not compromise the quality of the decoded signal.

1) Block Size Selection

In this experiment, we empirically try to determine the effect of the block size on BAAS by encoding and decoding the speech signal with various block sizes, i.e., ($f_s$, $f_s/2$, $f_s/4$, $f_s/8$, and $f_s/16$) and use the sample reduction rate (SRR) and mean square error (MSE) for selecting the appropriate parameter. Figure 6b shows the sample reduction rate (SRR) for the various block sizes. From the figure, we find that the SRR decreases linearly with block size. In other words, as the block size decreases, the number of data samples selected for the approximation increases. Besides, in figure 6a, we use mean square error (MSE) to test the accuracy of the decoded signal with decreasing block size. The graph shows that the MSE is low for $f_s/2$ and then nearly constant. We select the block size as $f_s/2$ for the following set of experiments, as it gives a better compression rate with acceptable loss in audio quality.

2) Statistical Parameters Selection

The statistical NUS is performed with an assumption to record the instance of significant changes in the statistical parameter within an interval. The various statistical parameters studied in this paper are Mean, Standard Deviation, and linear statistical parameters (LSP, considering both mean and slope). Figure 7 shows data sampled by the various statistical parameters. The figure shows that average samples retained by the proposed algorithm depend upon the chosen statistical parameter. The mean and standard deviation can retain about 90% and 50% of the signal on an average. However, considering the linear combination of mean and slope can retain about 25% of the samples on an average.

Figure 8a shows the Average MSE (mean square error) for all three statistical parameters, showing the dif-
difference between the reconstructed and original signal. For this, we compute the average MSE of the signals of the same length. For example, considering 3 signals of 2 minutes duration with the individual MSE values as \( \{0.000290641, 0.000197009, 0.000272864\} \), the average of the three values is plotted corresponding to the value 2 on the x-axis. Similarly, in figure 8b, we study the perceptual objective listening quality assessment ITU P.863 (POLQA)\[41\]. This is an intrusive object quality test which is based on the comparison of the original signal to the transferred signal. Figure 8 shows that the average POLQA for Mean is 3.0 and, it retains approximately 90% of the data samples. The STD has a lower subjective quality score of around 2.3 for 50% of the change points. In the following experimental evaluations, we use LSP, as it retains about 25% of data samples and generates a good signal. Figure 9 shows the percentage of the NUS retained (using LSP as the statistical parameter) with the POLQA score of each signal, with green dotted lines showing the average POLQA for the Libre Speech Corpus\[56\]. Furthermore, figure 10 shows the variations in the SegSNR of each signal. Figure 11 shows a comparison of the original and reconstructed signal using the chosen parameter.

**B. FIXED RATE SAMPLING**

One of the primary disadvantages of non-uniform sampling is that it selects a variable number of sample points per frame, which may not be suitable for the communication...
Amit Banerjee et al.: Optimal Non-Uniform Sampling by Branch-and-Bound Approach for Speech Coding

FIGURE 10: Variation of SegSNR with the percentage of the retained samples for Libre Speech Corpus [56].

FIGURE 11: Original and reconstructed signal using statistical parameter as Mean and Slope.

applications [58]. To solve this, we modify BAAS to select a fixed number of samples per frame. Figure 12 shows the relationship between the retained number of samples of the encoded signal and MSE, POLQA of the decoded signal. Figure 12a shows an increase in the objective quality (POLQA) with increasing number of retained samples. On the other hand, the MSE remains almost constant at 20%, as shown in figure 12b. This shows that LSP can retain the structural property of the waveform with about 20% of samples.

C. COMPARISONS WITH OTHER SAMPLING TECHNIQUES

In this section, we begin with a preliminary investigation on the number of non-uniform sample points selected for interpolation by the MMD, IPD, and the proposed methodology. Further, we also compare the proposed sampling technique with ADPCM [59] and also with level cross algorithm.

1) Comparison with other NUS Techniques

The MMD method samples the signal by local maxima and minima, marked by solid blue circles in figure 13. However, the IPD method selects the abrupt changes in the signal’s slope along with the local maxima and minima to overcome the sparseness problem of the MMD method, as shown by diamonds. Finally, the proposed methodology selects the data samples using statistical analysis of the signal, marked by the squares in figure 13. We consider equation 3 to mark the change points using linear \((LSP)\) of the signal. The figure shows that the proposed methodology selects fewer change points compared to the previous methods. Also, figure 14 shows a close approximation of the original signal, where we use the retrieved change points to approximate the original signals using the linear approximation.

Figure 15a shows the average data samples generated by the three methods on the Libre Speech Corpus. It is evident from the figure that BAAS retains a comparatively less number of samples than the IPD method. Similarly comparing the three algorithms in terms of average POLQA [41] scores, as shown in 15b, we find that the performance of MMD and IPD methods are close to each other; whereas the performance of BAAS is better than IPD method with better compression.
Furthermore, we compare BAAS, MMD and IPD methods using the EUSTACE speech corpus [57]. Table 2 shows the average results of the three methods. As shown in the table, BAAS performs better with lower MSE and higher POLQA scores in comparison to MMD and IPD.

2) Comparison of BAAS with ADPCM

We compare the proposed BAAS with ADPCM Speech Codec (Adaptive Delta Pulse Code Modulation). ADPCM takes advantage of the high correlation between the consecutive speech samples by encoding the difference between a predicted sample and the speech sample. In contrast, the proposed approach samples the changes in the statistical property of the signal following BAAS (block adaptive approach).

We use the following parameters for evaluating the speech quality of the two algorithms: segmental SNR (segSNR) [60], weighted-slope spectral distance (WSS) [60], POLQA [41] and log-likelihood ratio (LLR) [60]. The results are shown in Table 3. The table shows that the proposed BAAS algorithm produces an acceptable quality signal with an average POLQA score of 3.0. However, we can achieve POLQA scores comparable to the standard wave-coder ADPCM by marginally increasing the sampling rate as the quality of the reconstructed signal by BAAS is directly proportional to the percentage of the retained sample (as seen figure 12a).
TABLE 2: Comparison of MMD, IPD, and BAAS using EUSTACE speech corpus [57].

| Input files | Sampling Technique | POLQA | LLR | SNRSeg | WSS |
|-------------|--------------------|-------|-----|--------|-----|
| S1.wav      | ADPCM              | 3.5   | 0.16| 27.9   | 1.4 |
|             | BAAS               | 2.9   | 0.33| 21.8   | 2.2 |
| S2.wav      | ADPCM              | 3.3   | 0.18| 26.9   | 1.3 |
|             | BAAS               | 2.8   | 0.39| 21.5   | 2.05|
| S3.wav      | ADPCM              | 2.6   | 0.58| 28.5   | 3.7 |
|             | BAAS               | 2.5   | 0.61| 25.3   | 2.3 |
| S4.wav      | ADPCM              | 3.2   | 0.30| 28.1   | 0.7 |
|             | BAAS               | 3.1   | 0.29| 24.2   | 1.2 |
| S5.wav      | ADPCM              | 3.2   | 0.24| 25.6   | 2.3 |
|             | BAAS               | 2.8   | 0.43| 23.5   | 2.1 |
| S6.wav      | ADPCM              | 3.3   | 0.19| 27.2   | 1.4 |
|             | BAAS               | 2.9   | 0.28| 27.7   | 1.2 |

TABLE 3: Comparisons of ADPCM (Time Domain Sampling) and BAAS (Amplitude Domain Sampling) using six random audio samples from Libre corpus [56].

| Sampling Technique | LLR   | SNRSeg | WSS  | POLQA | Cbak | Covl | Csig | MSE   |
|--------------------|-------|--------|------|-------|------|------|------|-------|
| BAAS               | 1.41  | 15.08  | 4.9  | 3.1   | 4.2  | 3.6  | 3.7  | 0.00008|
| Linear Quantization  |       |        |      |       |      |      |      |       |
| 16 levels (4-bit code) | 2.81  | -3.3  | 98.7 | 2.1   | 1.7  | 1.2  | 0.57 | 0.0002 |
| μ Law Quantization   |       |        |      |       |      |      |      |       |
| 256 Level (8-bit Code) | 0.69  | 2.3   | 24.9 | 2.2   | 2.6  | 2.8  | 3.4  | 0.001  |

TABLE 4: Comparison of BAAS, Linear Quantization & μ Law Quantization using EUSTACE speech corpus [57].

3) Comparison with Level Crossing Sampling
In this section, we compare the performance of BAAS with the recently proposed level crossing sampling algorithms [23], [24]. For this, we use 16—level linear quantization of the amplitude and 256—level μ law quantization. The other parameters are similar to the ones discussed above. Table 4 shows the average results of BAAS and level crossing sampling. The comparative analysis suggests that the proposed technique has a lower MSE score, and thus the regenerated speech waveform is a better copy of the original signal.

D. PERFORMANCE OF BAAS
Next, in figure 16 we show the computational efficiency for encoding and decoding the speech signals with different time-durations for BAAS. The figure shows that the encoding time of BAAS increases linearly with the duration of the speech signal. However, the decoding time of BAAS is low and nearly independent of the time duration. It is because the complexity of the decoder is $O(n)$ (algorithm 4), where
$n$ is the number of retained samples. Besides, the algorithm computes intermediate points using basic arithmetical calculations.

We also perform the MUSHRA listening test (Multiple Stimuli with Hidden Reference and Anchor) following the ITU-R BS.1534-3 recommendation [42], to test the performance of BAAS. The experiment uses a web-based platform [61] to perform the MUSHRA test. In this, the participants compare and rate the audio quality of the processed and reference signals. The marking scale of $[0 \sim 100]$ is divided as Bad $[0 \sim 20]$, Poor $[20 \sim 40]$, Fair $[40 \sim 60]$, Good $[60 \sim 80]$, and Excellent $[80 \sim 100]$. It is a double-blind multi-stimulus test having a hidden reference signal. Similar to [62], we have added an anchor signal (low-pass filtered version of the original signal with a cut-off frequency of 3.5 kHz). In the experiment, we consider three test scenarios for three different audio samples. In each test scenario, evaluate five conditions. These include the reference signal (REF), low-anchor (Anchor35) signal, ADPCM, BAAS, and Level Cross Sampling methods speech codec. The audience consisted of 10 trained normal-hearing adults, with ages between 25 and 35 years.

Each speech sample is played twice before rating. Figure 17 shows the results of MUSHRA test. The first column shows the averaged value for all the three scenarios and the remaining three columns shows the individual ratings per test scenario. In the experimental result, we mark the average MUSHRA scores with confidence interval of 95%. The figure shows that the performance of BAAS in comparison to the other NUS sampling algorithms. From figure 17, BAAS has a subjective score higher than 92 on average, which indicates an acceptable quality of the regenerated signal.

Finally, we compare the performance of BAAS on two different corpora, i.e., Libre Speech Corpus [56], and EU-STACE speech corpus [57]. We perform the simulation with a block size set as half of the sampling frequency (fs). The encoder uses the proposed algorithm to retain the statistical changes in the signal. We use the following parameters for comparing the speech quality of the two datasets: segmental SNR (segSNR) [60], weighted-slope spectral distance (WSS) [60], POLQA [41] and log-likelihood ratio (LLR) [60]. We also verify the quality of the decoded signal using various quality scores like Csig Mean opinion score (MOS) to predict signal distortion [63], Cbak MOS predictor for background-noise intrusiveness [63] and Covl MOS predictor for signal quality on the whole [63]. Table 5 shows the average results, making it evident that the proposed NUS technique gives a
Amit Banerjee et al.: Optimal Non-Uniform Sampling by Branch-and-Bound Approach for Speech Coding

FIGURE 17: Comparison of ADPCM, BASS, and LCSS with MUSHRA Scores for three test scenarios with varying voice and silent regions. The colored markers represent the average score with 95% confidence intervals.

TABLE 5: Evaluating the performance of BAAS using EUSTACE and Libre Speech Corpus

| Dataset                  | LLR | SNRseg | WSS | POLQA | Cbak | Covl | Csig | MSE   |
|--------------------------|-----|--------|-----|-------|------|------|------|-------|
| EUSTACE speech corpus [57]| 1.41| 15.08  | 4.9 | 3.1   | 4.2  | 3.6  | 3.7  | 0.00008|
| Libre Speech Corpus [56] | 1.37| 15.34  | 4.68| 3.0   | 4.3  | 3.7  | 3.8  | 0.00008|

similar result for the two corpora.

V. CONCLUSION

In this paper, we propose an approach that statistically analyzes the speech signal to sample the abrupt changes over a time frame and approximate the signal with minimal reconstruction error using cost and linear penalty functions for avoiding the over-fitting problem. To further optimize the selected change points, we also propose a Branch-and-Bound algorithm. The methodology approximates the eliminated samples by linear interpolation. The proposed NUS technique retains only 25% of data samples to get an acceptable quality signal regeneration. The idea of using non-uniform sampling in the amplitude domain is to capture the time instance of sudden changes in the signal affecting the marginal distribution of the signal’s Mean, Variance, or Spectral Distribution. These time instances retrieve the structural property of the speech signal with minimal error. To evaluate the proposed non-uniform sampling methodology, we design a waveform coder viz. Block Adaptive Amplitude Sampling (BAAS). Compared to conventional non-uniform sampling methods that detect local maxima, minima, and inflection points, the proposed method shows a reduction in the number of detected samples and improvement in subjective quality. A primary disadvantage of non-uniform sampling is variable code rates that are not suitable for a fixed communication channel. We also propose a fixed-rate NUS based on BAAS and evaluate its performance on different standard databases.

REFERENCES

[1] M. Hasegawa-Johnson and A. Alwan, “Speech coding: Fundamentals and applications,” Wiley encyclopedia of telecommunications, 2003.
[2] M. Suman, H. Khat, M. M. Latha, and D. A. Kumari, “Speech enhancement and recognition of compressed speech signal in noisy reverberant conditions,” in Proceedings of the International Conference on Information Systems Design and Intelligent Applications 2012 (INDIA 2012) held in Visakhapatnam, India, January 2012, pp. 379–386, Springer, 2012.
[3] R. P. Sprague, “Compression of stored waveforms for artificial speech,” July 25 1989. US Patent 4,852,168.
[4] C.-P. Wu and C.-C. Kao, “Design of integrated multimedia compression and encryption systems,” IEEE Transactions on Multimedia, vol. 7, no. 5, pp. 828–839, 2005.
[5] L. Besacier, C. Bergamini, D. Vaufreydaz, and E. Castelli, “The effect of speech and audio compression on speech recognition performance,” in 2001 IEEE Fourth Workshop on Multimedia Signal Processing (Cat. No. 01TH8564), pp. 301–306, IEEE, 2001.
[6] M. N. Islam, “System and method for voice control of medical devices,” Aug. 21 2007. US Patent 7,259,906.
[7] S. Bassenyemukasa and R. J. Pilk, “Using voice signal analysis to identify authorized users of a telephone system,” Apr. 22 1997. US Patent 5,623,539.
[8] P. Sinha, “Waveform coders,” in Speech Processing in Embedded Systems, pp. 101–112, Springer, 2010.
[9] S. Tewksbury and R. Hallock, “Oversampled, linear predictive and noise-shaping coders of order n> 1,” IEEE Transactions on Circuits and Systems, vol. 25, no. 7, pp. 436–447, 1978.
[10] A. McCree, K. Truong, E. B. George, T. P. Barnwell, and V. Viswanathan, “A 2.4 kbit/s melp coder candidate for the new us federal standard,” in
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3138086, IEEE Access

Sakshi Pandey et al.: Optimal Non-Uniform Sampling by Branch-and-Bound Approach for Speech Coding

1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings, vol. 1, pp. 200–203, IEEE, 1996.

R. G. Van Der Waal and R. N. Veldhuis, “Subband coding of stereophonic digital audio signals,” in [Proceedings] ICASSP 91: 1991 International Conference on Acoustics, Speech, and Signal Processing, pp. 3601–3604, IEEE, 1991.

R. Zeilinski and P. Noll, “Adaptive transform coding of speech signals,” IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 25, no. 4, pp. 299–309, 1977.

N. S. Jayant, “Digital coding of speech waveforms: Pcm, dpcm, and dm quantizers,” Proceedings of the IEEE, vol. 62, no. 5, pp. 611–632, 1974.

J. D. Johnston, “Transform coding of audio signals using perceptual noise criteria,” IEEE Journal on selected areas in communications, vol. 6, no. 2, pp. 314–323, 1988.

R. Richey, “Adaptive differential pulse code modulation using pic® microcontrollers,” Microchip, Ed, 2007.

J. Foster and T.-K. Wang, “Speech coding using time code modulation,” in IEEE Proceedings of the SOUTHEASTCON'91, pp. 861–863, IEEE, 1991.

J.-H. Chen, R. V. Cox, Y.-C. Lin, N. Jayant, and M. J. Melchner, “A low-delay celp coder for the ccit 16 kb/s speech coding standard,” IEEE Journal on Selected Areas in Communications, vol. 10, no. 5, pp. 830–849, 1992.

J.-M. Valin, K. Vos, and T. Terriberry, “Definition of the opus audio codec,” IETF, September, 2012.

B. Bessette, R. Lefebvre, and R. Salami, “Universal speech/audio coding using hybrid acelp/rcx techniques,” in Proceedings,ICASSP'05, IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005., vol. 3, pp. iii–301, IEEE, 2005.

W. C. Chu, Speech coding algorithms: foundation and evolution of standardized coders. John Wiley & Sons, 2004.

S. Bae, M. Kim, and M. Bae, “A new hybrid non-uniform coding with low bit rates for sound signal in near field communication,” in 2013 International Conference on ICT Convergence (ICTC), pp. 1039–1040, IEEE, 2013.

Z. Perić and J. Nikolić, “An adaptive waveform coding algorithm and its application in speech coding,” Digital Signal Processing, vol. 22, no. 1, pp. 199–209, 2012.

R. Viswanadhnam, T. S. Kumar, and M. V. Subbarao, “A level cross-based nonuniform sampling for mobile applications,” in Microelectronics, Electromagnetics and Telecommunications, pp. 767–777, Springer, 2018.

V. Ravuri, S. K. Terlapu, and S. Nayak, “Adaptive level cross sampling for next-generation data-driven applications,” in Microelectronics, Electromagnetics and Telecommunications, pp. 633–640, Springer, 2021.

J. Krimmer, C. Füllner, and S. Randel, “A new approach to nonuniform sampling of clipped audio signals for lte smart phones,” in 2013 IEEE Third International Conference on Consumer Electronics, Berlin (ICCE-Berlin), pp. 129–130, IEEE, 2013.

L. Davissson, “Data compression using straight line interpolation,” IEEE Transactions on Information Theory, vol. 14, no. 3, pp. 390–394, 1968.

L. Ehrman, “Analysis of some redundancy removal bandwidth compression techniques,” Proceedings of the IEEE, vol. 55, no. 3, pp. 278–287, 1967.

M. Bae, W. Lee, and D. Kim, “On a new vocoder technique by the nonuniform sampling,” in Proceedings of MILCOM'96 IEEE Military Communications Conference, vol. 2, pp. 649–652, IEEE, 1996.

S. Senay, L. F. Chaparro, M. Sun, and R. J. Schlapass, “Adaptive level-crossing sampling and reconstruction,” in in 18th European Signal Processing Conference, pp. 1296–1300, IEEE, 2010.

S. Pal and M. Mitra, “Detection of eg characteristic points using multiresolution wavelet analysis based selective coefficient method,” Measurement, vol. 43, no. 2, pp. 255–261, 2010.

M. Hajar, M. El Badaoui, A. Raad, and F. Bonnardot, “Discrete random sampling: Theory and practice in machine monitoring,” Mechanical Systems and Signal Processing, vol. 123, pp. 386–402, 2019.

J. F. Ducrée-Robitaille, A. Vincent, and G. Boulet, “Comparison of techniques for detection of discontinuities in temperature series,” International Journal of Climatology: A Journal of the Royal Meteorological Society, vol. 23, no. 9, pp. 1087–1101, 2003.

S. Aminikhanghahi and D. J. Cook, “A survey of methods for time series change point detection,” Knowledge and information systems, vol. 51, no. 2, pp. 339–367, 2017.

T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, Introduction to algorithms. MIT press, 2009.

P. Daniel, “Librispeech asr corpus.”

L. White and S. King, “The eustace speech corpus,” 2003.

J. Walrand and P. P. Varaiya, High-performance communication networks. Morgan Kaufmann, 2000.

V. Ramamoorthy and N. S. Jayant, “Enhancement of adpcm speech by adaptive postfiltering,” AT&T Bell Laboratories technical journal, vol. 63, no. 6, pp. 1465–1475, 1984.

J. H. Hansen and B. L. Pellom, “An effective quality evaluation protocol for speech enhancement algorithms,” in Fifth international conference on spoken language processing, 1998.
[61] M. Schoeffler, S. Bartoschek, F.-R. Stöter, M. Roess, S. Westphal, B. Edler, and J. Herre, “webmushra—a comprehensive framework for web-based listening tests,” Journal of Open Research Software, vol. 6, no. 1, 2018.

[62] P. Govalkar, J. Fischer, F. Zalkow, and C. Dittmar, “A comparison of recent neural vocoders for speech signal reconstruction,” in Proc. 10th ISCA Speech Synthesis Workshop, pp. 7–12, 2019.

[63] Y. Hu and P. C. Loizou, “Evaluation of objective quality measures for speech enhancement,” IEEE Transactions on audio, speech, and language processing, vol. 16, no. 1, pp. 229–238, 2007.