Sub - Band Coding and Speech Quality Testing

Karush Suri
Student, Department of Electronics and Communication Engineering, Amity University
D-56, sector -21, Noida, Uttar Pradesh, India

Abstract

Speech coding has proved to be a promising technique in modern telephonic and signal processing systems. The decoding process finds its applications in various telecommunication aspects such as video processing, verbal transmission, voice recognition devices, etc. It also has played an important part in spectrum analysis, radar systems, antenna systems and in enhancing voice privacy methods, globally. In coding the sub- band, the speech coding band is divided into a number of several bands which are coded separately. The audible frequency range (20 Hz – 20 KHz) is divided into frequency sub-bands using a bank of Finite Impulse Response (FIR) filter. The output of each signal is then sampled and encoded. At the receiving end, each signal is de-multiplexed, decoded and demodulated. These signals are then combined to reconstruct the main output signal. The basic building blocks of a digital signal processing (DSP) system[1] constitute the interpolators and decimators. These provide the up-sampling and down-sampling of the signal at a suitable bit rate. The method helps in obtaining a higher compression ratio of the coding signal. It also provides a flexible variable coding and helps in controlling the transmission of speech within a large range actively. Furthermore, measures of speech quality and speech quality testing have also been discussed.

Keywords : Linear Time Invariant, Speech Quality Index, Finite Impulse Response, Mean Opinion Score

I. INTRODUCTION

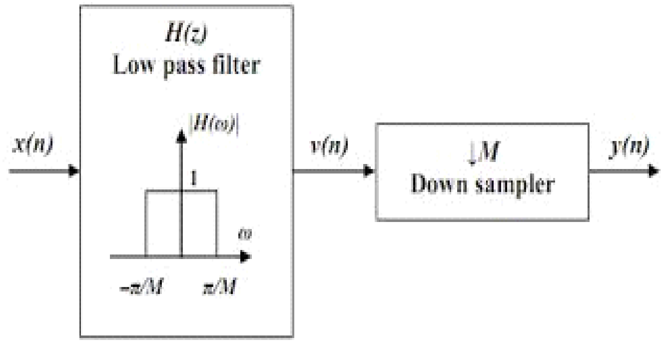
In practice we often encounter signals where most of the energy content, which is important to us, is present in a particular frequency band. In speech signals most of the energy is present in the lower frequency bands. Coding the complete signal that is by allocating same number of bits for the entire signal is not an efficient way of coding the signal for either transmission or storage. By taking advantage of the fact that most of the energy is present in a particular frequency band we can split the signal into various bands depending on the information content and then code the sub-band signals separately. The coding is carried out with the primary step being the splitting of the frequency spectrum. The first frequency subdivision splits into two equal width segments, a low pass signal and a high pass signal. The second frequency band splits the low pass signal into another set of low pass and high pass signals. And, the high pass signal splits into two signals of equal bandwidths. Thus, a total of 4 bands are obtained and decimation is performed. A reduction in the bit rate of the digitalized speech signal is achieved by allocating different number bits per sample to each signal.

II. MULTI RATE SIGNAL PROCESSING METHOD

The multi rate technique, in totality, deals with the alteration of sampling rates. Sample-rate conversion is the process of changing the sampling rate of a discrete signal to obtain a new discrete representation of the underlying continuous signal. Application areas include image scaling, and audio/visual systems, where different sampling-rates may be used for engineering, economic, or historical reasons. For example, Compact Disc Digital Audio and Digital Audio Tape systems use different sampling rates, and American television, European television, and movies all use different frame rates. Sample rate conversion prevents changes in speed and pitch that would otherwise occur when transferring recorded material between such systems. To achieve different sampling rates at different stages, multi rate digital signal processing systems employ the down-sampler and the up-sampler, the basic sampling rate alteration devices in addition to the conventional elements. Many multi rate systems employ a bank of filters with either a common input or a summed output.

A. Decimation

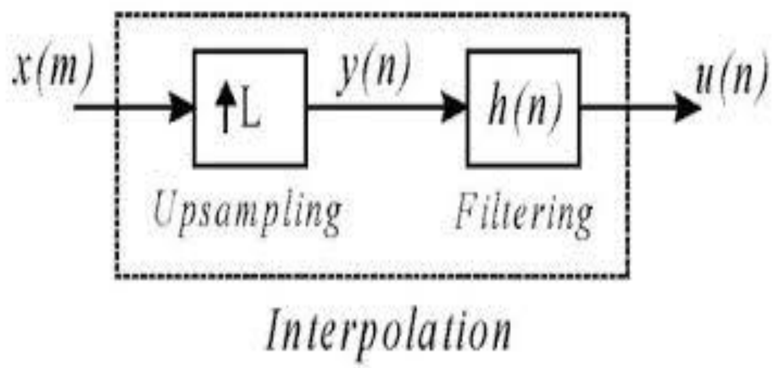
Multi rate signal processing applies down-sampling of the signal. In this manner, sampling rate of the signal is reduced by a definite factor. It can be regarded as the discrete-time counterpart of sampling. In decimation we start with a discrete-time signal x[n] and convert it into another discrete-time signal y[n], which consists of sub-samples of x[n].



$$X(k) = \sum_{n=0}^{N-1} x(n) W_N^{kn}$$

B. *Interpolation*

Interpolation[2] is the exact opposite of decimation. It is an information preserving operation, in that all sample of x[n] are present in the expanded signal y[n]. With reference to Figure, the expansion process is followed by a unique digital low-pass filter called an anti-imaging filter. Although the expansion process does not cause aliasing in the interpolated signal, it does however yield undesirable replicas in the signal's frequency spectrum.



C. *FIR Filters*

These are classified as a basic category of filters which alter the signal constituents, if required. A number of complex designs have been witnessed which range from the Linear Time-Invariant (LTI) filters to the constant coefficient design. They are composed of multipliers, adders, and delay elements. Multipliers are rather complex parts, so sometimes they may be replaced by simpler parts. For example, to multiply by 8,it is easier to shift the number 3 places to the left(since $2^3=8$). The delay elements can be thought of as registers. An FIR with constant coefficient is a LTI filter. The output of an FIR of order (or length) H, to an input time-series x[n], is given by a finite version of the convolution sum:

$$y(n) = \sum_{k=0}^{H-1} h(k) * x(n-k)$$

In a practical situation, the FIR[3] coefficients are obtained from a computer design tool and presented to the designer as floating point numbers. The performance of a fixed-point FIR, based on the floating-point coefficients, should be verified using simulation or algebraic analysis to ensure that design specifications remain satisfied. Furthermore, various designs of these filters are completely not suitable for all kinds of signals. Continuous efforts in developing the structure and design of the filter are being made in order to preserve the signal and process it without much disturbance. Following key steps are used in order to improve the design of the filter-

- Realize each filter coefficient with an optimal CSD code.
- Increase effective multiplier speed by pipelining.
- For FIR with symmetric coefficients, the number of multipliers can be reduced.

III. SPEECH QUALITY TESTING

Over the years, new and comprehensive methods have been revised to measure the quality of speech signals. Acceptable results have helped speech coders in designing better and authenticated mannerisms for processing the speech signal with less hindrance and at effective rates. Early military speech coders were judged according to only one criterion: intelligibility. With the advent of consumer grade speech coders, intelligibility is no longer a sufficient condition for speech coder acceptability. Consumers want speech that sounds ''natural.'' A large number of subjective and objective measures have been developed to quantify ''naturalness,'' but it must be stressed that any scalar measurement of ''naturalness'' is an oversimplification. A number of different parameters provoke the interest of developers for measuring the extent of reduced noise level and Speech Quality Index (SQI)[4].

A. Mean Opinion Score

The Mean Opinion Score (MOS)[5] is perhaps the most commonly used speech parameter. MOS is computed by coding a set of spoken phrases using a variety of coders, presenting all of the coded speech together with ungraded speech in random order, asking listeners to rate the quality of each phrase on a numerical scale, and then averaging the numerical ratings of all phrases coded by a particular coder. The five-point numerical scale is associated with a standard set of descriptive terms: 5 = excellent, 4 = good, 3 = fair, 2 = poor and 1 = bad. Factors such as language and location of the testing laboratory may shift the scores of all coders up or down, but tend not to change the rank order of individual coders. For all of these reasons, a serious MOS test must evaluate several reference coders in parallel with the coder of interest, and under identical test conditions.

The diagnostic acceptability measure (DAM)[6] is an attempt to control some of the factors that lead to variability in published MOS scores. The DAM employs trained listeners, who rate the quality of standardized test phrases on 10 independent perceptual scales, including six scales that rate the speech itself (fluttering, thin, rasping, muffled, interrupted, nasal), and four scales that rate the background noise (hissing, buzzing, babbling, rumbling).

B. Algorithmic Measures

Psychophysical measures of speech fail to fully detect the quality. Due to the reliability of the score on the subject, a wide range of performances are observed. In order to avoid this inconveniece, algorithmic speech quality testers are implied, These wor kon the complete part of the speech using certain specific algorithmics. This simply minimises the error percentage and helps the coders to check th performance on much broader aspects.The development of algorithms that accurately predict the results of MOS or comparative testing is an area of active current research, and a number of improvements, alternatives, and/or extensions to the PSQM[7] measure have been proposed.

The PSQM measure allows automated, simulation based test methodologies to objectively rate speech clarity as well as transmission voice quality. This results in considerable savings of time and cost over the previous traditional practices. Moreover, it yields objective results that are reliable and reproducible. It uses a psychoacoustical mathematical modelling(both perceptual and cognitive) algorithm to analyze the pre and post transmitted voice signals, yielding a PSQM value which is a measure of signal quality degradation and ranges from 0 (low degradation) to 6.5 (high degradation). However, PSQM originally was not conceived for items such as packet loss, delay variation (jitter) or non-sequential packets. These conditions usually give inappropriate results under heavy network load simulations, failing to account for a very perceived loss of voice quality. Thus, PSQM is still considered a vast area for future research and alteration.

IV. CONCLUSION

Effective measures for coding speech signals such as sub- band coding are designed to alter the energy consistency in the various energy containing bands. These methods are applied in a number of fields such as radar systems, video processing and in the analysis for digital signal processing communication. Furthermore, in order to assess the quality of the speech signal obtained MOS comes out as a convenient measure which depends on the individual's perception of the speech. Algorithmic measures of speech testing have also been discussed which highlight a more elaborate manner of testing voice clarity. These also provide an insight into the upcoming studies related to the generation of new and efficient algorithms.

REFERENCES

[1] John G. Proakis and Dimitris G. Manolakis, "Digital Signal Processing: Principles, Algorithms and Applications", Third Edition

[2] Lalitha R Naik and Devaraja R Naik, "Sub band coding of speech signals using multirate signal processing and comparing the various parameter of different speech signal by corrupting the same speech signal"

[3] R. Cox et al., New directions in subband coding,IEEE JSAC 6(2): 391–409 (Feb. 1988).

[4] Computer Complexity of Adaptive Algorithms in Echo Cancellation, Mrs. A.P. Patil, Dr. Mrs. M.R. Patil, IJECE-Volume-2 Issue-7.

[5] ITU-T Recommendation, P.800, " Methods for subjective determination of transmission quality"

[6] Yi Hu and Loizou, P.C., "Evaluation of Objective quality measures for speech enhancement", 2008

[7] Antony W. Rix, John G. Beerends, Doh-Suk Kim, "Objective assessment of speech and audio quality-technology and applications"