NON-INTRUSIVE NOISE REDUCTION IN GSM VOICE SIGNAL USING NON-PARAMETRIC MODELING TECHNIQUE

3 authors:

Safiu Abiodun Gbadamosi
University of Pretoria

A. M. Aibinu
Federal University of Technology Minna

Adeiza J. Onumanyi
Council for Scientific and Industrial Research, South Africa

www.seetconf.futminna.edu.ng

www.futminna.edu.ng

NON-INTRUSIVE NOISE REDUCTION IN GSM VOICE SIGNAL USING NON-PARAMETRIC MODELING TECHNIQUE

S.A Gbadamosi[1*], A. M. Aibinu[2], O.C.Ugweje[3], A. J Onumanyi[4], E. N Onwuka[5], & M. Aderinola[6]
[1,2,4,5]Federal University of Technology Minna, Niger State, Nigeria.
[3]Digital Bridge Institute, Abuja, Nigeria
[*]g_safiu.its@futminna.edu.ng, 08060511079.

ABSTRACT

Noise degrades the quality and intelligibility of speech. It impedes speech clarity, coding, recognition and speaker identification. To mitigate noise effect and improve speech quality, we propose Non-parametric modeling technique along with a Non-intrusive signal denoising system based on short time Fourier transforms. This paper aims to establish only the idea behind our proposed algorithm, however, we present argument to justify that our results will reduce an end to end acoustic background noise; improve quality of speech for both the speaker and the listener and eventually increase throughput. Ultimately, users' will be able to call and receive calls in a noisy environment while enjoying clarity of voice.

Keywords: *Non-parametric, Non-intrusive signal denoising system, Short time fourier transform, Speech quality.*

1. INTRODUCTION

Communication systems continue to be plagued by the limiting effects of noise. These varying noise levels degrade speech quality, especially in Global System for Mobile (GSM) communication networks, thereby creating difficulty in audio reception. An initial step towards addressing this noise challenge is being able to quantify its effect on voice signals. One performance metric for achieving this evaluation requirement is the use of speech quality[Barile et. al.2006]. Speech quality signifies the lucidity of any speaker's words as identified by the audience [Mahdi 2007].

Two main approaches used to evaluate speech quality are either the use of subjective and objective speech quality measures [Aicha 2012]. Each approach has its limitation especially in the area of reducing distortions such as noise reduction, echo cancellations, listening level, and loudness and so on [RIX, and HOLLIER 2000].

To reduce these effects, we propose noise reduction in GSM voice signals using non-parametric modelling techniques approach for evaluation of speech quality. This technique adopts the use of Non-Intrusive Signal Denoising algorithms which employs the use of noise suppression algorithms such as Short Time Fourier Transform (STFT) algorithms. This technique will extract noise time consumption. It is user-directed and hence offers precise understanding of quality features that lead to better service acceptance from the end users. It is designed to judge speech quality alongside signal distortion, noise distortion as well as overall quality. A correlation among this specific pair of distortions and the observed quality of speech will be developed.

The paper is organised as follows. Section II introduces the related work to our research alongside with the proposed method and noise reduction techniques adopted. In section III, we elucidated the processes adopted to achieve our speech quality. In section IV, we stated the results expected to be obtained. Finally, we concluded in section V.

2. RELATED WORK

Review of related work from [Mahdi & Picovici 2009, Quackenbush et. al.1988, Rix 2006, Hansen and Pellom 1998, Noll et. al. 1976, Noll 1974, Tribolet et. al. 1978] of early simple objective measures of quality, based on metrics of overall SNR and their extension were proposed.

The limitation of the above metrics is that they lack predictive power. This lead to the discovery of linear prediction coefficients, Itakura-Saito (IS) distortion measure, loglikelihood ratio (LLR) measure, cepstral distance measure which uses cepstral coefficient and Bark spectral distortion (BSD) which measures mean squared Euclidean distance within spectral vectors of the coded utterances and the original [Rix 2006, Noll et. al 1974, Klatt et. al.1982]. The performances of the above measures predict speech quality moderately. The mathematical representation of all the above metrics can be denoted as follows [Kondo 2012];

$$SNR = 10\log_{10}\frac{\sum_n x^2(n)}{\sum_n (x(n)-d(n))^2} \qquad (1)$$

$$SNRseg = \frac{10}{M}\sum_{m=0}^{M-1}\log_{10}\left(\frac{\sum_{n=Nm}^{Nm+N-1} x^2(n)}{\sum_{n=Nm}^{Nm+N-1}[(d(n)-x(n))]^2}\right) \qquad (2)$$

$$fwSNRseg = \frac{10}{M}\times\sum_{m=0}^{M-1}\frac{\sum_{j=1}^{K} W(j,m)\log_{10}\frac{|x(j,m)|^2}{(|x(j,m)|-|X(j,m)|)^2}}{\sum_{j=1}^{K} W(j,m)} \qquad (3)$$

$$Where \quad W(j,m) = |X(j,m)|^{\gamma} \qquad (4)$$

$$d_{WSS} = \frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{j=1}^{K} W(j,m)(S_C(j,m)-S_p(j,m))^2}{\sum_{j=1}^{K} W(j,m)} \qquad (5)$$

For loglikelihood ratio (LLR) measure:

$$d_{LLR}(\vec{a}_p,\vec{a}_c) = \log\left(\frac{\vec{a}_p R_c \vec{a}_p^{\,T}}{\vec{a}_c R_c \vec{a}_c^{\,T}}\right) \qquad (6)$$

For Itakura-Saito (IS) distortion measure[Kondo 2012]:

$$d_{IS}(\vec{a}_p,\vec{a}_c) = \frac{\sigma_c^2}{\sigma_p^2}\log\left(\frac{\vec{a}_p R_c \vec{a}_p^{\,T}}{\vec{a}_c R_c \vec{a}_c^{\,T}}\right) + \log\left(\frac{\sigma_c^2}{\sigma_p^2}\right) - 1 \qquad (7)$$

For cepstral coefficient calculation:

$$d_{CEP}(\vec{c}_c,\vec{c}_p) = \frac{10}{\log 10}\sqrt{2\sum_{k=1}^{p}[C_c(k)-C_p(k)]^2} \qquad (8)$$

Where $\vec{a}_p, \vec{a}_c$ are the Linear Predictive Code (LPC) vector from the reference (original) voice signal frame and Linear Predictive Code vector from the enhanced voice frame, respectively. While R_c is the autocorrelation matrix of the reference (original) voice signal. σ_c and σ_p are the LPC (increase) gains from the reference (clean) and improve (enhanced) signals, respectively. Also $\vec{C}_c$ and $\vec{C}_p$ are the cepstrum coefficient vector from the reference (clean) and improved (enhanced) signals, respectively.

For Bark Spectral Distortion (BSD):

$$BSD = \frac{\frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{O}[L_x^{(m)}(i)-L_d^{(m)}(i)]^2}{\frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{O}[L_x^{(m)}(i)]^2} \qquad (9)$$

Where M will be the amount of frames (speech segments) processed, O is amount of critical bands, $L_x^{(m)}(i)$ is the bark spectrum from the mth critical frame of reference (original) speech, and $L_d^{(m)}(i)$ is the mth critical frame of coded speech of bark spectrum.

In order to get better performance, many objective measures modelled approach-driven were developed, given rise to models such as Perceptual Audio Quality Measure (PAQM), Perceptual Speech Quality Measure (PSQM) [Côté 2011], Measuring Normalizing Block [BEERENDS and STEMERDINK 1994], Perceptual Evaluation of Audio Quality Measure (PEAQ) [Hekstra 2001], Perceptual Analysis Measurement system (PAMS) to Perceptual Evaluation of Speech Quality (PESQ) [Pocta 2010a]. PSQM and PESQ were adopted as ITU-T recommendation P.861and P.862 respectively [Pocta 2010b]. In spite of the high success recorded on PESQ. The search for more better, intuitive and predictive measure continues as all the above models required both the degraded signal and its corresponding clean version. The limitation of the above models is in the obtainment of the reference (clean) signal. This brings us to the counterpart method called Non-intrusive measures. The measure uses only the degraded speech signal to predict quality as against the intrusive measures. These models range from Auditory Non-Intrusive QUality Estimation (ANIQUE) [Cote 2011], single ended assessment measure (SEAM) and so on. The basic principle adopted by Non-intrusive measure was that it depends on the relationship of three ideas released by the ITU-T as the ITU–T Rec. P.563 (2004) [Cote 2011]: (i) any derivation in the degraded signal, of various parameters relevant to the speech production system, (ii) following reconstruction of a referral signal from a degraded signal, both signals are evaluated by an intrusive model, and (iii) discovery of specific distortions in the degraded signal. Subsequently, the extracted parameters are linearly merged to anticipate voice transmission quality. Above the aggregation stage, the perceptual influence of every parameter is quantified via a distortion-dependent weighting operation [Cote 2011]. The problem of quality of speech above is that, it measures the impact of radio frequency (RF)–related impairments on hearing or Listening quality, which are not able to capture and calculate how other important voice quality impairments present in live calls e.g. background noise, acoustic echo and mismatched speech levels affect users experience. In trying to adopt this basic principle, we proposed Non-Intrusive Signal Denoising System (NISDS) using Short Time Fourier Transform (STFT) to achieve real-time execution of the algorithms as well as to obtain a trade-off in between high quality reduction in noise along with minimal computational heap. The NISDS was used to derive our reference signal before using linear predictive code (LPC) and correlation of the subjective MOS value between the Noisy speech and denoised speech (clean) to predict our quality of speech.

3. NON-PARAMETRIC MODELLING TECHNIQUE

In contrast, non-parametric models discover the parameters for a statistical model describing a signal, system, or process. These techniques use known information about the system to determine the model. Applications for non-parametric modelling include speech and music synthesis, data compression, high-resolution spectral estimation, communications, manufacturing, and simulation [Mathwork 1988-2015]. It usually have no sound signal to process (and thus help to make restricted utilization of perceptual techniques), but rather estimate MOS from measured properties of the underlying transport and/or terminal, for instance, echo, delay, speech levels and noise [Aicha 2012], VoIP network characteristics [Rix et. al 2000], [Mahdi and Picovici 2009], or cellular radio reception measures [Quackenbush et. al 1988]. Parametric models are traditionally used for network planning, to construct MOS estimates based on tabulated values such as the codec type, bit-rate, delay, packet loss statistics, etc. [Rix 2006]. This process demand total characterization of the system under test and consequently could be considered as Glass box approach in which no expertise in the system underneath test is required. A Non-intrusive parametric assessment such as these, would be use to address two critical issues. Firstly, estimate the occurring distortions (noise) which is a challenge since the original signal is unknown. Secondly, predict the subjective impacts of the estimated distortions. The non-intrusive parametric approach determines first, the well characterized distortions such as impulsive noise and deduced a mathematical relation among the limited set and the subjective opinions [Mahdi 2009].

4. NOISE REDUCTION PROCESSES

In speech communication, the speech signal is always accompanied by some noise. In most cases, background noise of the environment, where the source of speech lies. The obvious effect of this noise addition is to make the listening task difficult for a direct listener. The purpose of P.835 listening tests is to assess the trade off between background Noise attenuation and foreground spe ech degradation that arises in noise reduction processing.

It should be noted that the improvement of overall listening quality is not the only purpose of noise reduction processing; the removal of noise during speech pauses also helps reduce the amount of data needed for transmission. Hence, it is another form of data compression but this time eliminates distortion caused by noise. The noise reduction algorithm used in this proposal is Short time Fourier Transform (STFT). The goal of noise reduction in speech is to improve the quality of degraded speech. Two general classes of problems have been assumed depending on the point where the corrupting noise is included to the voice signal. The assumed scenarios are;

- CORRUPTING NOISE INCLUDED AT THE TALKER ENVIRONMENT: It is presumed that the person speaking is in a noisy environment while the listener is in quiet.

- CORRUPTING NOISE INCLUDED AT THE LISTENER ENVIRONMENT: It is presumed that the listener is in a noisy environment while the speaker is in quiet. i.e. a supervisor delivering guidelines to employers on a noisy factory floor [Wuhan University 2004].

5. PROPOSED ALGORITHM

This research work adopts the approach of Non-Parametric Modeling method of estimating the quality of speech using Non-Intrusive Signaling Denoising System to extract Noise parameters from Speech. A speech signal was recorded using MATLAB command and a quantified amount of synthesized Gaussian noise is mixed with it using MATLAB code to get a Noisy Speech signal. The technical conditions were in accordance with the guidelines given in ITU-T recommendation P.800. The Noisy speech has been obtained using Synthetic approach. The Noisy signal was passed into channelization chamber where it convolved with hamming window in order to obtain a windowed segmented signal. The output of the windowed segmented signal was passed into the STFT process where the signal is analyzed, manipulated and the noise parameter in the signal is removed and any abnormal variation in the speech signal was considered as degradations. Secondly, a reference signal (denoised signal) is reconstructed from the Noisy signal with the aid of Non-Intrusive Signal Denoising system as depicted in Fig 1. In a signal comparison approach, the overall quality is finally determined from the degradation measures weighted, according to the dominant distortion identified in the speech signal and a correlation between the subjective MOS value of the degraded speech signal and the denoised (clean) speech signal.

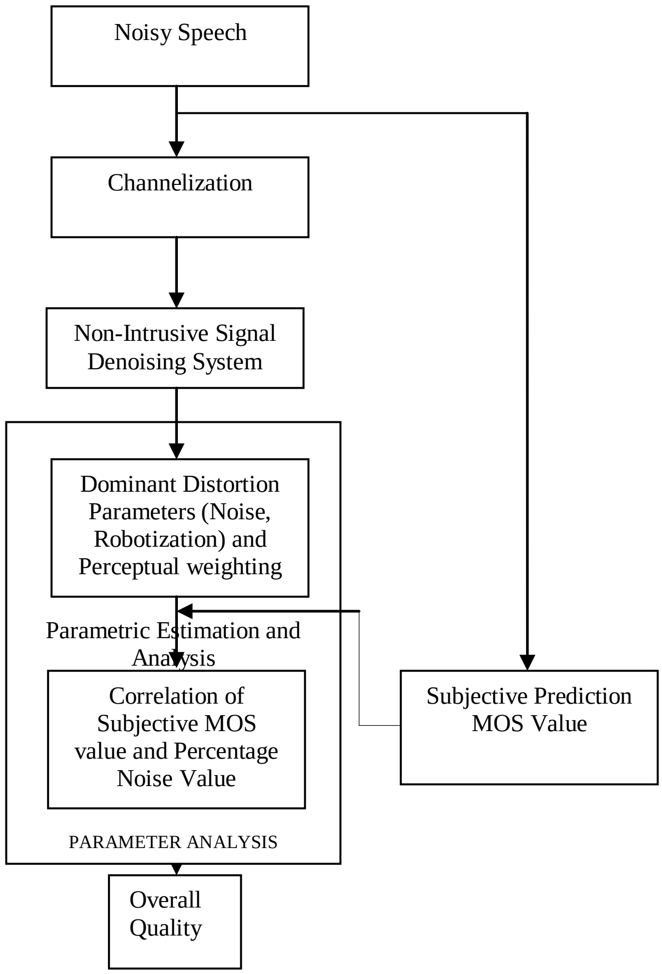

Fig 1: Block Diagram of Non-Parametric Modelling Technique

6. EXPECTED RESULT

The expected outcome of this research work will achieve an efficient model for noise reduction in GSM voice signal which will improve speech quality. Secondly, the removal of noise parameter from speech is a form of compression which increases throughput and help reduce the amount of data needed for transmission.

7. CONCLUSION

We have described how to reduce background acoustic noise in GSM voice signal using Non-Parametric modeling technique which adopts the use of Non-intrusive signal denoising system (NISDS) to reduce noise. The NISDS was based on the method of Short time Fourier transform to manipulate and analyze the noisy signal. This model when fully developed will be useful in speech analysis, recognition and of course noise reduction, which will improve speech clarity and intelligent.

ACKNOWLEDGEMENTS

My praise and adoration goes to the Almighty GOD for His guidance and protection. I appreciate my supervisors most especially Dr. Musa A. Aibinu and Dr. A.J Onumanyi for proof reading this paper. Lastly to my family, my brother (M.Aderinola), my parent and my siblings for their prayers and support. I want to specially recognise my lovely wife and children for their encouragement, prayers and understanding during this graduate studies. Above all else, without her support none of this would have been possible.

REFERENCE

Ascom,*Technical White Paper Series RxQual and voice quality*, Ascom Infrasys AG

Aicha, A.B.. "Perceptual speech quality measures separating speech distortion and additive noise degradations", Speech Communication, 201205.

A.M. Noll, Cepstrum pitch determination, J. Acoust. Soc. Am. 41 (1974) 293–309.

Barile, M., Camarda, P., Dell'Aquila, R., & Vitti, N. (2006, September). Parametric Models for Speech Quality Estimation in GSM Networks. In *Software in Telecommunications and Computer Networks, 2006. SoftCOM 2006. International Conference on* (pp. 204-208). IEEE

Beerends, J., Hekstra, A., Rix, A., & Hollier, M. (1998). *Perceptual Evaluation of Speech Quality (PESQ), the new ITU standard for end-to-end speech quality assessment. Part II− Psychoacoustic model.* Technical report, ITU-T.

BEERENDS, J. G., STEMERDINK, J. A. A perceptual speech quality measure based on a psychoacoustic sound representation. J. Audio Eng. Soc., 1994, vol. 42, p. 115-123, ISSN 1549-4950

Hansen, J. H. and Pellom, B. L., \An effective quality evaluation protocol for speech enhancement algorithms," in Proc. Int. Conf. Spoken Lang. Process., 1998.

Hekstra, A.P. "Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs", 2001 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings (Cat No 01CH37221) ICASSP-01, 2001.

Klatt. et. al, \Prediction of perceived phonetic distance from critical-band spectra: A _rst step," in Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '82, 1982.

Kazuhiro Kondo. "Speech Quality", Signals and Communication Technology, 2012,

Mathworks,'Signal Processing Toolbox™ User's Guide, The MathWorks, Inc. 3 Apple Hill Drive Natick, MA 01760-2098, © COPYRIGHT 1988–2015 by The MathWorks, Inc.

Mahdi, A. E. "Voice Quality Measurement in Modern Telecommunication Networks", 2007 14th International Workshop on Systems Signals and Image Processing and 6th EURASIP Conference focused on Speech and Image Processing Multimedia Communications and Services, 06/2007.

Mahdi, A. E., & Picovici, D. (2009). Advances in voice quality measurement in modern telecommunications. *Digital Signal Processing*, *19*(1), 79-103.

Noll, P. Et. al, \Adaptive quantization in speech coding systems," in lnt. Zurich Seminar on Digital Communication (IEEE), 1976.

Nicolas Côté. "Speech Quality Measurement Methods", Integral and Diagnostic Intrusive Prediction of Speech Quality, 2011.

Peter Pocta. "Impact of fragmentation threshold tuning on performance of voice service and background traffic in IEEE 802.11b WLANs", 20th International Conference Radioelektronika 2010, 04/2010

Peter Pocta "Predictions in Case of Independent and Dependent Losses (in Presence of Receiver-Side Comfort-Noise)", Radioengineering/12102512, 20100401.

Quackenbush, S. R., Barnwell III, T. P., and Clements, M. A., Objective Measures of Speech Quality. Prentice Hall, 1988.

Rakesh Kumar, Sandeep Saini " Measuring Parameters for speech quality in cellular networks" Computer Science and Application Department, Kurukshetra University, Kurukshetra, IndiaEmail:sandeepsaini083@gmail.com, Globalize the Research.,International Journal of Advances in Computer Networks and its Security.

RIX, A. W., HOLLIER, M. P. The perceptual analysis measurement system for robust end-to-end speech quality assessment. In Proceedings of IEEE ICASSP 2000. Istanbul (Turkey), 2000, vol. 3, p. 1515-1518.

Rix, A.W. "Objective Assessment of Speech and Audio Quality—Technology and Applications", IEEE Transactions on Audio Speech and Language Processing, 11/2006

Speech Signal Processing, School of Electronic Information, Wuhan University, Chapter 5 Speech Enhancement.

Tribolet et. al, \A study of complexity and quality of speech waveform coders," in Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '78, 1978.