ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 11, November 2017

Speech Enhancement under Different Noisy Environments for Intelligibility Enhancement

[1] A RAGHU VARDHAN BABU

***Abstract*— Speech intelligibility is an important measure of the effectiveness or adequacy of a communication system or to communicate in a noisy environment. In many daily life situations it is important to understand what is being said, for example over a loudspeaker system, and to be able to react to acoustic signals of different kinds. Inherent noise presence in message production process has huge impact. Both production and interpretation noise processes have a fixed signal-to-noise ratio. A simple and effective model of human communication is proposed in this work which has ability to enhance the intelligent speech in noisy environment. Simulation results yields better performance over traditional state-of-methods.**

Keywords— Speech Communication, Noise Environment, Intelligibility Enhancement

I. INTRODUCTION

Speech output – whether from mobile phones, public address systems or simply domestic audio devices – is widely used. In many listening contexts the intelligibility of the intended message might be compromised by environmental noise or channel distortion. Problems can be minimized by increasing output intensity or repeating the message, but these approaches are not ideal for either the listener or the output device (e.g. power consumption, failure). A better approach is to seek ways to modify the speech signal to increase intelligibility in noise.

Speech enhancement refers to the improvement in the quality or intelligibility of a speech signal and the reversal of degradations that have corrupted it. Quality is a subjective measure which reflects on the pleasantness of the speech or on the amount of effort needed to understand the speech material. Intelligibility is an objective measure which signifies the amount of speech material correctly understood. The main objective of Speech Enhancement is to enhance the speech signal to obtain a clean signal with higher quality. Such system has been widely used in long distance telephony applications.

2. BACKGROUND

(A) Basis for the proposed method of speech enhancement

Human beings perceive speech by capturing some features from the high signal-to-noise ratio (SNR) regions in the spectral and temporal domains, and then extrapolating the features in the low SNR regions (Cooper, 1980). Therefore speech enhancement should primarily aim at highlighting the high SNR regions relative to the low SNR regions. Lowering the signal levels in the low SNR regions relative to the signal levels in the high SNR regions may help in reducing the annoyance due to noise without losing the information. The relative emphasis of the features in the high SNR regions over the features in the low SNR regions should be accomplished without causing distortion in speech. Otherwise the enhancement may cause annoyance of a type different from that due to additive noise.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 11, November 2017

(B) Effects of noise on the speech signal

Before we proceed to discuss our approach, let us brief review some characteristics of noisy speech. Speech signal has a large (30±60 dB) dynamic range in the temporal and spectral domains. For example, in the temporal domain some sounds have low signal Speech enhancement deals with processing of noisy speech signals, aiming at improving their perception by human or their correct decoding by machines (Berouti et al 1979). Speech enhancement algorithms attempt to improve the performance of communication systems when their input or output signals are corrupted by noise. The presence of background noise causes the quality and intelligibility of speech to degrade. Here, the quality of speech refers how a speaker conveys an utterance and includes such attributes like naturalness and speaker recognizability. Intelligibility is concerned with what the speaker had said, that is, the meaning or information content behind the words (Hu and Loizou 2007). Therefore, a noisy environment reduces the speaker and listeners ability to communicate. To reduce the impact of this problem speech enhancement can be performed. It is usually difficult to reduce noise without distorting speech and thus, the performance of speech enhancement systems is limited by the tradeoff between speech distortion and noise reduction (Boll 1979).

Efforts to achieve higher quality and/or intelligibility of noisy speech may effectively end up improving performance of other speech applications such as energy, especially during the release of stop sounds and in the steady nasal sounds. Speech signal energy level is also low prior to the release of a stop sound and also in some fricative sounds.

(C) Speech Enhancement and Its Applications

speech coding/compression and speech recognition, hearing aids, voice communication systems and so on. The goal of speech enhancement varies according to specific applications, such as to reduce listener fatigue, to boost the overall speech quality, to increase intelligibility and to improve the performance of the voice communication device. Hence speech enhancement is necessary to avoid the degradation of speech quality and to overcome the limitations of human auditory systems.

3. PROPOSED METHOD

(A) Model with Production and Interpretation Noise

Let the message have a construction noise, representing the normal version in its generation, either for one person or across all talkers. The transmitted signal for dimension k at time I is then

$$X_{k,i} = S_{k,i} + V_{k,i} \qquad (1)$$

Where $V_{k,i}$ the place is production noise. The nonheritable signals fulfill

$$Y_{k,i} = X_{k,i} + N_{k,i} \qquad (2)$$

Where $N_{k,i}$ is environmental noise. Eventually, the obtained symbols area unit taken

$$Z_{k,i} = Y_{k,i} + W_{k,i} \qquad (3)$$

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 11, November 2017

The common data rate between the first multi-dimensional message succession and they got multi-dimensional message grouping depicts the viability of the correspondence procedure. In this first depiction, we expect the procedures to be memory less, which is sensible for time-frequency signal representations. The shared data rate is then equivalent to the shared data between the multi-dimensional images and at a specific time interval. We moreover accept that the individual segment signs of the multi-dimensional arrangement are autonomous. At that point we can compose

Let us consider the conduct of the production and interpretation noises for the speech application. Speech creation is a probabilistic procedure. A discourse sound is never precisely the same. This variability is to a great extent free of the force level at which it is created. That is, the creation SNR is consistent (with , where means desire and where we exclude the time subscript to improve documentation). It takes after that the relationship coefficient between the message signal also, the genuine sign , indicated as , is an altered number on [0,1].

A fixed SNR for the elucidation commotion is additionally sensible. The sound-related framework contains an increase adjustment for each basic band [16], which implies that the exactness of the translation scales with the sign over a significant dynamic reach. In this manner, the understanding SNR and the connection coefficient can be demonstrated as fixed.

The consistent SNR creation and/or elucidation noise has a significant impact on a power compelled communication system. In a conventional communication system with parallel channels (without generation and/or elucidation noise) the best data throughput is gotten by water filling [17]: more signal power is given to correspondence channels with low noise power. Be that as it may, in the present correspondence framework there is for the most part little benefit to having a channel SNR, , that is significantly beyond the generation SNR or the understanding SNR. The convenience of a specific correspondence channel "saturates" close to the generation SNR or the interpretation SNR, whichever is lower.

(B) Tractable Model that Includes Enhancement

We now embed a machine-based enhancement symbol $\mathcal{G}$ in the Markov chain. In the event that we mark by~ all signs influenced by the improvement administrator we get a Markov chain S$\rightarrow X \rightarrow X \rightarrow \tilde{Y} \rightarrow \tilde{Z}$, where . $\tilde{X} = \mathcal{G}(X)$

To detail a tractable enhancement issue, let us make the assumptions that all procedures are together Gaussian, stationary, and memoryless. For simplicity of documentation, we preclude the time i from here-on forward. For the Gaussian case it can be demonstrated that

$$I(S_k; \tilde{Z}_k) = -\frac{1}{2} log(1 - \rho^2_{S_k \tilde{Z}_k}) \quad (5)$$

We can make a few disentanglements. Exploiting the Markov chain property, we see that $\rho_{S_k \tilde{Z}_K} = \rho_{S_k \tilde{X}_K} \rho_{\tilde{Y}_k \tilde{X}_K} \rho_{\tilde{Y}_k \tilde{Z}_K}$. The altered translation SNR infers

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 11, November 2017

$\rho_{\tilde{Y}_k\tilde{Z}_K} = \rho_{Y_kZ_K}$..On the off chance that the improvement symbol $\mathcal{G}$ is a relative capacity for every part signal, then we additionally have $\rho_{S_k\tilde{X}_K} = \rho_{S_kX_K}$.

Next, we consider how the hypothesis is influenced if the signal is translated in its auditory representation. In Section II-A we portrayed a mapping from the acoustic to the auditory representation. Inside each ERB band various Gaussian variables are combined in one method. Our model without upgrade inside a specific ERB band with record comprises of i) the era of an arrangement of variables$S_k, k \in k_m$, ii) The expansion of autonomous noise variable $U_k = V_k + N_k + W_k$to each created variable iii) The summation (in the ear) of all variables to the single ERB band arbitrary variable:$Z_m = \sum_{k\in k} S_k + U_k$. Assuming $\rho^2_{S_kS_n+U_k,}$ is constant for $k \in k_m$,it can then be demonstrated that

$$I(\{S_k\}_{k\in k_m}; Z_m) = -\frac{1}{2}log(1 - \rho^2_{S_kS_n+U_n,}), n \in k_m \quad (6)$$

Which is same as (5) preceding the enhancement symbol is included. In this manner, we have found that from the mentioned assumptions the above hypothesis persists to the situation where the last recipient is the human auditory system, which incorporates within the speech signal.

(C) Optimizing Information Throughput

Our objective is to optimize the effectiveness of the communication method by choosing a good enhanced operator G. Let us contemplate a typical time-frequency illustration such as that obtained with a paraunitary physicist or DCT filterbank. For this illustration, the belief of a memoryless stationary process is cheap. We tend to contemplate a memoryless linear and time-invariant operator

$(g(X))_k = \sqrt{b_k} \cdot X_k$, that is affine, and redistributes signal power by multiplying every frequency channel with a

The measure (4) is identified with existing heuristically-determined measures. In the event that we compose the channel SNR a $\xi_k = \frac{\sigma^2_{N_k}}{\sigma^2_{X_k}}$ and $\rho_{0,k} = \rho_{S_k}, X_k, \rho_{Y_k}, Z_k$, we can utilize (5) to revise (4) as

$$I(S; \tilde{Z}) = -\sum_{K\in k} \frac{1}{2} log\left(\frac{(1-\rho^2_{0,k})\xi_k+1}{\xi_k+1}\right) \quad (7)$$

Using $I_k = -\frac{1}{2} log\left((1 - \rho^2_{0,k})\right)$ and the sigmoid $A_k(\xi_k) = \frac{log\frac{(1-\rho^2_{0,k})\xi_k+1}{\xi_k+1}}{log\left((1-\rho^2_{0,k})\right)}$ we get

$$I(S; \tilde{Z}) = \sum_{k\in k} I_k A_k(\xi_k) \quad (8)$$

If we recognize I_k as the scaled band-importance function and $A_k(\cdot)$ as the weighting function the shared data can be deciphered as the scaled verbalization file (AI), e.g., [20], [21], or the scaled speech intelligibility file (SII) [22], [23]. While the sigmoid $A_k(\xi_k)$, varies from the heuristically chose bends utilized as a part of AI and SII, the likeness is well inside the precision of the reasoning used to touch base at the AI and SII plan. Along these lines, (8) forms a theoretical justification for this established work on speech intelligibility

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 11, November 2017

gain $\sqrt{b_k}$. The distribution is subject to AN overall signal power preservation constraint. The understandability optimization problem is currently,

$$\max_{\{b_k\}} I(S;\tilde{Z}) \; subject\; to \sum_{k\in k} b_k\sigma_{X_k}^2 - B = 0 \; and \; b_k \geq 0, \forall_k \; (9)$$

The ß above is the power of the vector X. This drawback can be solved by KKT (Karush-Kuhn-Tucker) conditions.

The correlation coefficients are ρ S_kX_k and ρ Y_kZ_k and those correlation coefficient are fixed. Another correlation coefficient ρ X_kY_k varies with the coefficient b_k as given below,

$$\rho_{\bar{X}_k\bar{Y}_k} = \frac{1}{\sqrt{1+\frac{\sigma_{N_k}^2}{b_k\sigma_{X_k}^2}}} \quad (10)$$

Which may be a bell-shaped improvement downside as the objective operates is formed. From (11) we construct the Lagrangian,

$$\max_{\{b_k\}} \sum_{k\in k} \frac{1}{2} log\left(\frac{b_k\sigma_{X_k}^2+\sigma_{N_k}^2}{(1-\rho_{0,k}^2)b_k\sigma_{X_k}^2+\sigma_{N_k}^2}\right) subject\; to \sum_{k\in k} b_k\sigma_{X_k}^2 - B = 0 \; and b_k \geq 0, \forall_k \; (11)$$

$$\mathcal{L}(\{b_k\},\lambda,\{\mu_k\}) = \sum_{k\in k}\frac{1}{2} log\left(\frac{b_k\sigma_{X_k}^2+\sigma_{N_k}^2}{(1-\rho_{0,k}^2)b_k\sigma_{X_k}^2+\sigma_{N_k}^2}\right) + \lambda b_k\sigma_{X_k}^2 + \mu_k b_k \; (12)$$

The μ_k are always positive and ≯ is nonpositive (as the mutual information is monotonically increasing as a perform of b_k). Differentiating the Lagrangian to the b_k and setting the results to zero results in the stationarity conditions of the KKT conditions:

$$0 = \frac{1}{2}\frac{\sigma_{X_k}^2}{\sigma_{X_k}^2+\sigma_{N_k}^2} - \frac{1}{2}\frac{(1-\rho_0^2)\sigma_{X_k}^2}{(1-\rho_0^2)b_k\sigma_{X_k}^2+\sigma_{N_k}^2} + \lambda\sigma_{X_k}^2 + \mu_{k,}\forall_k \; (13)$$

Multiplying with the denominators leads to a quadratic in b_k ,

$$\alpha b_k^2 + \beta b_k + \gamma = 0 \; (14)$$

$$\gamma = \frac{1}{2}\rho_0^2,\sigma_{X_k}^2,\sigma_{N_k}^2, + (\lambda\sigma_{X_k}^2+\mu_k)\sigma_{N_k}^4 \; (15)$$

$$\beta = (\lambda\sigma_{X_k}^2+\mu_k)(2-\rho_{0,k}^2)\sigma_{X_k}^2,\sigma_{N_k}^2 \; (16)$$

$$\alpha = (\lambda\sigma_{X_k}^2+\mu_k)(1-\rho_{0,k}^2)\sigma_{X_k}^4 \; (17)$$

We have to study the behavior of the equation (14) which is quadratic. We can say that roots are real only if the condition $\beta^2 - 4\alpha\gamma \geq 0$. we will study that what will happen if $\mu_k = 0$. we may see that $4\alpha\gamma$ may include two terms as given $\frac{1}{2}\rho_0^2\sigma_{X_k}^2\sigma_{N_k}^2\alpha$,which is negative for $\mu_k = 0$ and (≯ $\sigma_{X_k}^2$+μ_k) $\sigma_{N_k}^4\alpha$,which is positive for $\mu_k = 0$. If we have letter term < β^2 we may conclude that b_k has real roots and we can say that only if

$$4(1-\rho_{0,k}^2) \leq (2-\rho_{0,k}^2)^2 \quad (18)$$

This is always true since $\rho_{0,k}^2$ ε [0,1] . The roots might, however, both be negative and during this case the term $\mu_k b_k$ should be sufficiently negative to force the basis to $b_k = 0$ .This leads to the standard KKT resolution. a straightforward line search formula for the that provides the proper overall power is:

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 11, November 2017

(1) Choose $\prec$;

(2) Solve (14) with for all;

(3) set any negative to zero;

(4) Check if the facility $\sum_{k\in K} b_k \sigma_{X_k}^2$ is sufficiently on the point of the desired overall power B. If not, then fits $\prec$ to be more negative if the facility is just too high and additional positive if the facility is just too low.

The formula is well extended to a bi-section formula. It will currently be seen that, in distinction to the case wherever the assembly and interpretation noise aren't thought of, increasing a single will either decrease or increase b_k .From the standard quadratic root formula it follows that for a given ρ_0^2 and $\sigma_{X_k}^2$ the modification in worth for depends on the term in the root. Contemplate once more. The behavior depends on whether the positive term $-\frac{1}{2}\ \rho_0^2 \sigma_{X_k}^2 \sigma_{N_k}^2 \alpha$ or the negative term $-(\prec \sigma_{X_k}^2 + \mu_k)\ \sigma_{N_k}^2 \alpha$ is larger. The primary term being larger corresponds to the "saturated" case mentioned at the end of the introduction and also the case wherever the second terms is larger to the "unsaturated" case.

4. RESULTS

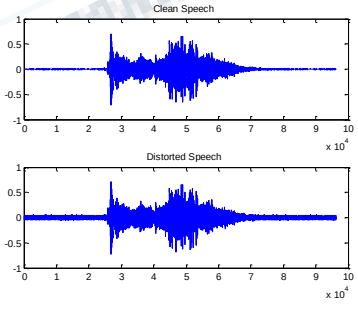

Figure 1: Clean speech and distorted speech (Babole noise)

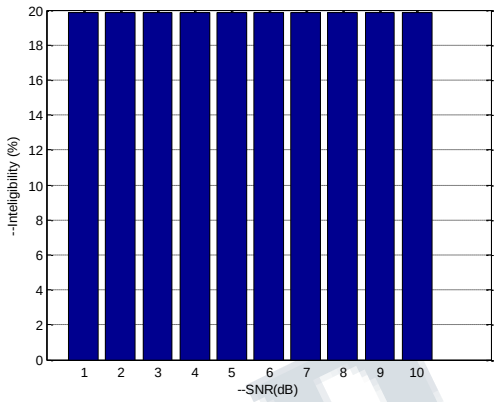

Figure 2: SNR VS Intelligibility

5. CONCLUSION

A simple information-theory based model of speech communication suffices for state-of-the-art enhancement of the intelligibility of speech played out in a noise environment. The model makes the plausible assumption that both the production and the interpretation process in the speech communication chain are subject to noise that scales with the signal level. Our approach can be refined in a number of aspects. Regularization can be applied to reduce intelligibility enhancement if no noise is present.

REFERENCES

[1] P. Loizou, Speech Enhancement Theory and Practice. Boca Raton, FL, USA: CRC Press, 2007.

[2] S. Srinivasan, J. Samuelsson, and W. B. Kleijn, "Codebook-based Bayesian speech enhancement for nonstationary environments," IEEE Trans. Audio, Speech Lang. Process., vol. 15, pp. 441–452, Feb. 2007.

[3] R. C. Hendriks, T. Gerkmann, and J. Jensen, DFT-Domain based single-microphone noise reduction for speech enhancement-a Survey of the State of the Art. San Rafael, CA, USA: Morgan & Claypool, 2013.

[4] V. Grancharov, J. H. Plasberg, J. Samuelsson, and W. B. Kleijn, "Generalized postfilter for speech quality enhancement," IEEE Trans. Audio, Speech Lang. Process., vol. 16, no. 1, pp. 57–64, Jan. 2008.

[5] R. Niederjohn and J. Grotelueschen, "The enhancement of speech intelligibility in high noise levels by high-pass filtering followed by rapid amplitude compression," IEEE Trans. Acoust. Speech, Signal Process., vol. ASSP-24, no. 4, pp. 277–282, 1976.

[6] B. Sauert, G. Enzner, and P. Vary, "Near end listening enhancement with strict loudspeaker output power constraining," in Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC), Paris, France, 2006.

[7] B. Sauert and P. Vary, "Near end listening enhancement optimized with respect to speech intelligibility index and audio power limitations," in Proc. European Signal Processing Conf. (EUSIPCO), Aalborg, Denmark, Aug. 2010, pp. 1919–1923.

[8] T. C. Zorilă, V. Kandia, and Y. Stylianou, "Speech-in-noise intelligibility improvement based on spectral shaping and dynamic range compression," in ISCA Interspeech, Portland, OR, USA, 2012.

[9] P. N. Petkov, G. E. Henter, and W. B. Kleijn, "Maximizing phoneme recognition accuracy for enhanced speech intelligibility in noise," IEEE Trans. Audio, Speech Lang. Process., vol. 21, no. 5, pp. 1035–1045, 2013.

[10] C. Taal, J. Jensen, and A. Leijon, "On optimal linear filtering of speech for near-end listening enhancement," IEEE Signal Process. Lett., vol. 20, no. 3, pp. 225–228, 2013.

Author Profile

He received his M.Tech degree in DSCE from Jaya Prakash narayan College of Engineering, B.Tech degree in Electronics and Communication Engineering from Jaya Prakash narayan College of Engineering, and His interested areas are Image Processing and communications.