Noise-Robust Reverberation Time Estimation Based on Sound HOS Decay

Yosra MZAH[1, 2]; Mériem JAIDANE[2, 1]

[1] CEA LinkLab, Telnet Innovation Labs, 2083 Ariana, Tunisia

[2] Ecole Nationale d'Ingénieurs de Tunis, Université de Tunis El Manar, 1068 Tunis, Tunisia

ABSTRACT

This work is a part of a European project entitled I'CityForAll[3] which aims to improve the mobility of elderly people suffering from presbycusis. In particular, this project proposes algorithms dedicated to public address systems, in order to enhance the intelligibility of vocal announces broadcasted in noisy and reverberant public confined spaces.

As the reverberation time (T60) of such spaces varies with the occupancy level, dynamic estimation of this parameter is needed to optimize the performances of these algorithms. In this paper, we deal with a robust method for the estimation of T60 in noisy conditions. It is an approach which generalizes the standard sound energy decay method based on diffusing of a noise sequence and performing the T60 estimation from its energy decay. However, in this method we rather propose to diffuse a signal having specific statistical properties and to estimate the T60 using the decay of its HOS (Higher Order Statistics).

The performances of this method are evaluated using a set of synthesized and real impulse responses, for realistic signal-to-noise ratios ranging from 5dB to 20dB.

Keywords: Reverberation time (T60), energy decay method, ambient noise, Higher Order Statistics (HOS) decay

1. INTRODUCTION

This work constitutes a part of a European project entitled I'CityForAll (1) which aims to improve the living of elderly persons suffering from sensorineural hearing impairment called presbycusis.

In particular, this project aims to propose algorithms for the enhancement of the intelligibility of vocal messages diffused in large dimension and noisy public confined spaces such as railway stations and airports.

The speech intelligibility in such spaces is degraded not only by the effect of ambient noise but also the effect of reverberation resulting from the propagation of the sound from the source to the receiver.

Therefore, algorithms dedicated to public address systems are proposed in the context of the I'CityForAll project in order to reduce the impact of reverberation on speech comprehension.

Room Reverberation is characterized by different acoustic parameters such as reverberation time (T60), clarity, definition and centre time (2). In particular, the reverberation time is considered as the primordial parameter that objectively quantifies room reverberation.

In a previous study (3) we have demonstrated that the reverberation time varies when the room occupancy level changes. This work was conducted in the context of a university restaurant and confirms the findings of previous studies done in churches and classrooms (4, 5).

That is why, a dynamic estimation of the reverberation time is considered to be important to optimize the performances of real time speech enhancement algorithms according to the space occupancy level (3) as well as the population profile (1).

[1] yosra.mzah@telnet-innovation-labs.net

[2] U2S@enit.rnu.tn

[3] http://www.icityforall.eu/

The reverberation time T60 is commonly defined as the time it takes for a sound to decay by 60dB after the sound source is turned off (2). The estimation of this parameter constitutes a research field that interests engineers and acousticians for several purposes including the setting of algorithms and applications requiring knowledge about the reverberation amount such as de-reverberation (5, 6).

Several methods have been used for the estimation of the reverberation time. Based on Sabine's models (8), this parameter is first estimated from the room geometry and the absorption coefficients of the different materials existing in the room. In (9), Schroeder proposed a technique for the estimation of the T60 based on the energy decay curve derived from the room impulse response.

In practical situations, these informations are usually not available and the T60 must be blindly estimated from the recorded reverberant and noisy signals. Several techniques have been proposed for the blind estimation of the reverberation time. Some of them use the distributions of sound decay curves (10, 11, 12) or the statistical modeling of these curves such that the maximum likelihood (ML) estimator can be used to determine the T60 (13).

Semi-blind methods are also been developed where the room characteristics are learned using neural network approaches (14).

Even though some works have dealt with the estimation of the reverberation time in noisy conditions (15, 16), it still remains a challenging issue. It is the case of the T60 estimation in noisy public confined spaces such as railway stations and airports.

In this paper, we develop a robust method for the estimation of T60 in noisy conditions. The proposed approach generalizes the standard sound energy decay method based on diffusing a noise sequence and performing the T60 estimation from its energy decay. However, in this method we rather propose to diffuse a signal having specific statistical properties and to estimate the T60 using the decay of its Higher Order Statistics (HOS).

The reminder of this paper is organized as follows:

In Section 2, we review the T60 estimation approach based on the standard sound energy decay and we zoom on the limitations of this method in noisy conditions. In Section 3, we describe the principle of the new estimation method. We evaluate the performances of this method in Section 4 and we conclude in Section 5.

2. LIMITATIONS OF REVERBERATION TIME ESTIMATION USING SOUND ENERGY DECAY METHOD

2.1 Acoustic impulse response model

When propagating from the source to the receiver, the sound makes a series of reflections including the direct sound, the early reflections and the late reflections. Assuming that the late reflections dominate the direct sound and the early reflections, the room impulse response can be modeled according to Polack model (18) as the outcome of a non-stationary random process:

$$h(t) = b(t).e^{-\delta.t} \tag{1}$$

Where $b(t)$ is a zero-mean Gaussian stationary noise and δ is the decay rate which is inversely proportional to the reverberation time (18, 19).

The decay rate δ is related to the slope of the energy decay curve as follows:

$$\log(E(h^2(t))) = -2.\delta.t + \log(\sigma_b{}^2) \tag{2}$$

Where $\sigma_b{}^2$ denotes the variance of $b(t)$

Assuming that the sound field in the room is diffuse and the source-microphone distance is higher than the critical distance (19), the link between the reverberation time T60 and the decay rate δ can be established by the requirement:

$$10\log_{10}\frac{E(h(0)^2)}{E(h(T60)^2)} = 60 \tag{3}$$

The following equation is then obtained.

$$T60 = \frac{3}{\delta.\log_{10}(e)} \qquad (4)$$

2.2 Sound energy decay method

Sound energy decay method consists in diffusing a noise sequence and performing the estimation of the reverberation time from its energy decay.

Let's note $x(t)$ the diffused signal (excitation signal) and $y(t)$ the signal recorded at the receiver. In real situations $y(t)$ is expressed as follows:

$$y(t) = x(t) * h(t) + n(t) \qquad (5)$$

Where $h(t)$ defines the acoustic room impulse response and $n(t)$ refers to the ambient noise.

In noiseless conditions, the temporal decay curve is defined as follows:

$$E(y^2(t)) = \sigma_b{}^2.e^{-2.\delta.t}.\alpha(t) \qquad (6)$$

with $\alpha(t)$ expressed by equation 7, where Th corresponds to the length of $h(t)$.

$$\alpha(t) = \int_{t-Th}^{0} e^{2.\delta.u}.E(x^2(u))\,du \qquad (7)$$

The decay curve derived from the signal $y_d(t)$ is obtained thorough the logarithm of $E(y^2(t))$ (equation 8):

$$\log(E(y^2(t))) = -2.\delta.t + \log(\sigma_b{}^2) + \log(\alpha(t)) \qquad (8)$$

The dynamic range corresponds to the linear portion of the logarithm of the energy decay (equation 8). The T60 is calculated from the slope of this dynamic range estimated based on linear regression. We intuitively notice that the larger is the dynamic range the better is the estimation of the T60.

The comparison between equation 2 and equation 8 shows the bias induced when considering the decay sound of a reverberant signal instead of that of the impulse response for the estimation of the T60. This bias is related to the parameter $\alpha(t)$. The effect of $\alpha(t)$ is more important for non-stationary signals. Figure 1 shows the decay curves corresponding to a white noise and to a speech signal. We notice that the decay curve relative the stationary white noise is smoother and provides thus a more accurate estimation of the T60.

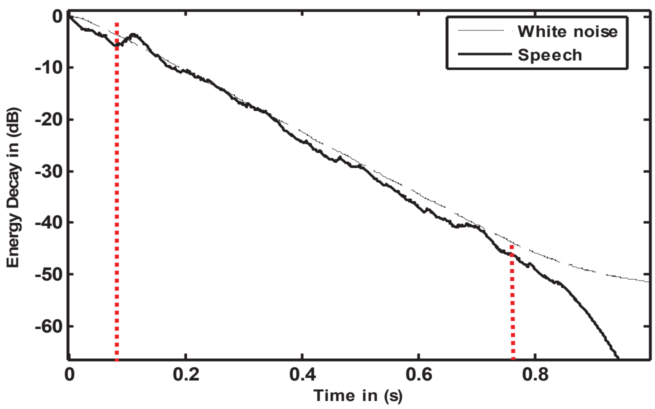

Figure 1 – Energy decay curve corresponding to a white noise signal and to a speech signal. The dynamic range corresponds to the portion of the energy decay delimited with two vertical red dashed lines. In this simulation, an impulse response synthesized according to Polack model with a T60 equal to 1s is used.

In noisy conditions and assuming that the reverberant signal $y(t)$ and the additive noise $n(t)$ are independent random variables, the energy decay corresponding to the noisy reverberant signal is given by:

$$\log(E(y^2(t))) = \log(\sigma_b{}^2.e^{-2.\delta.t}.\alpha(t) + \sigma_n{}^2(t)) \tag{9}$$

The T60 estimation based on the sound decay is disturbed by the effect of both the parameter $\alpha(t)$ (equation 7) and the additive noise variance $\sigma_n{}^2(t)$. According to equation 9, the background noise influence becomes bigger as the reverberant signal vanishes.

The lower is the signal-to-noise ratio (SNR) the more reduced is the dynamic range, as shown in Figure 2 where we illustrate the decay curves obtained with a white noise disturbed with stationary additive noise for an SNR ranging from 5dB to 20dB.

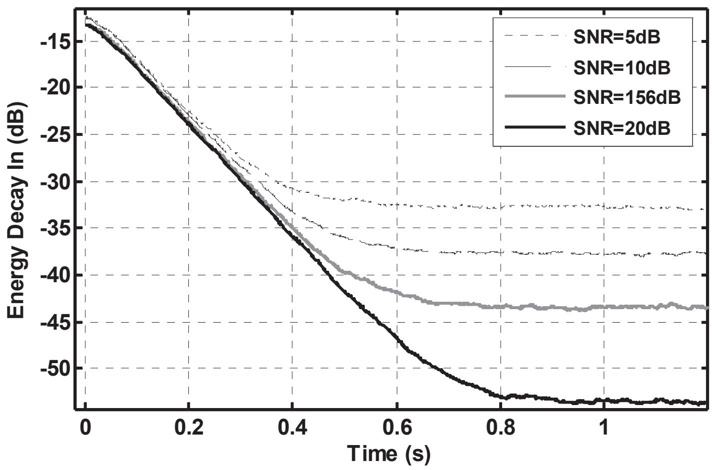

Figure 2 – Reduction of the dynamic range of a white noise energy decay due to the effect of additive noise for SNR levels ranging from 5dB to 20dB. In this simulation, an impulse response synthesized according to Polack model with a T60 equal to 0.7s is used.

3. REVERBERATION TIME ESTIMATION METHOD USING HOS AND ROBUSTNESS TO NOISE

3.1 Principle of the method

The T60 estimation method presented in this paper is also based on the analysis of the sound decay, however it tries to generalize this estimation approach in order to ameliorate its robustness to background noise.

If we note $y(t)$ the reverberant noisy sound decay (equation 5) recorded after the extinction of the sound source $x(t)$. In order to obtain a decay curve that is independent of the effect of additive noise, we propose to use higher order statistics (HOS), in particular the third order moment.

Thus, in this method, instead of using the entity $E(y^2(t))$ for the estimation of the sound decay energy, we propose to use the entity $E(y^2(t).\mathrm{x}(\mathrm{t}-\tau))$.

If the decay sound $y(t)$ is observed at the moment indexed t, $x(t-\tau)$ corresponds to the previous value of the excitation signal according to a time gap τ.

If we consider that $h(t)$ denotes the room impulse response following Polack model (equation 1) and we assume that the sound source $x(t)$ is a zero mean random variable and that $h(t)$ and $x(t)$ are independent random variables.
Under the previous assumptions, the third order decay is expressed as follows:

$$E(y^2(t).x(t-\tau)) = \sigma_b{}^2.e^{-2.\delta.t}.\beta(t) \tag{10}$$

Where the entity $\beta(t)$ reads:

$$\beta(t) = \int_{t-Th}^{0} e^{2.\delta.u}.E(x^2(u).x(t-\tau))\,du \tag{11}$$

The decay curve is then given by:

$$\log(E(y^2(t).x(t-\tau))) = -2.\delta.t + \log(\sigma_b{}^2) + \log(\beta(t)) \tag{12}$$

Where *Th* denotes the length of the acoustic impulse response $h(t)$.
The comparison between equation 9 and equation 12 shows that the decay curve obtained with the proposed method is independent of the effect of additive noise.
The equation 12 is close to equation 8 that corresponds to the case of undisturbed decay and provides thus a larger dynamic range for the estimation of the reverberation time.

3.2 Design of the excitation signal

The reverberation time estimation method proposed in this paper is based on the diffusion of an excitation signal with specific statistical properties, so that the HOS decay (equation 10) is independent of the additive ambient noise $n(t)$ and provides thus a larger dynamic range.
In order to achieve this objective, the excitation signal $x(t)$ must verify the following conditions:

- The amplitude values of $x(t)$ must be highly correlated, in order to avoid that the integral $\beta(t)$ (equation 11) approaches zero and yields a third order decay equal to zero.
- The third order moment of $x(t)$ must not be equal to zero. Thus, the amplitude values of $x(t)$ must not follow a symmetric distribution.

In this paper, for the assessment of the proposed T60 estimation method, we use a sample of excitation signal, synthesized in the aim to validate the conditions listed above.
For this, we consider a zero mean random signal whose amplitude values follow an asymmetric distribution (skewness coefficient=1).
Such distribution is obtained as the sum of two Gaussian distributions weighted using a Bernoulli series (20). The first Gaussian distribution has a standard deviation equal to 2 and the second one has a standard deviation equal to 0.3.
The realizations of the two Gaussian distributions were also highly correlated in order to achieve the first condition to be verified for the excitation signal.
In Figure 3 we show the temporal variations of the proposed excitation signal as previously designed, the distribution of its amplitude values and its normalized auto-correlation function.
This excitation signal is used in this paper in order to assess the performances of the proposed reverberation time estimation method based on HOS decay.

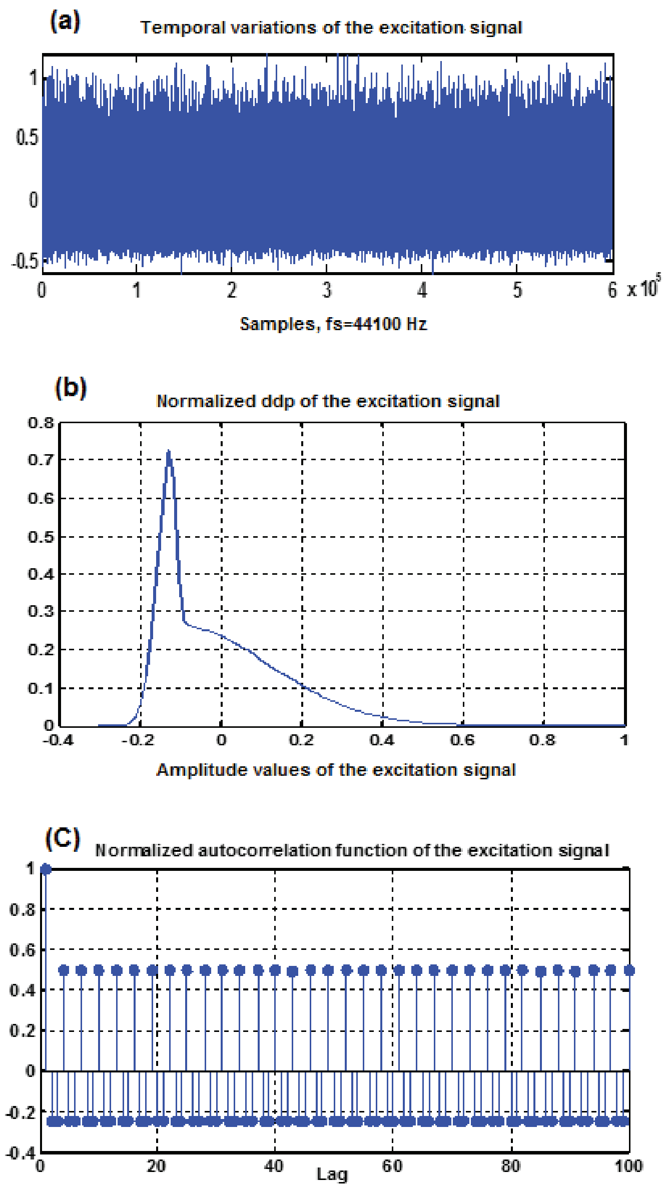

Figure 3 – (a) temporal variations of a sample realization of the excitation signal (b) Normalized distribution of the amplitudes values of the excitation signal samples (c) Normalized auto-correlation function of the excitation signal samples (lags=100)

3.3 Example of T60 estimation based on HOS method

For the estimation of the third order decay, we use a recursive structure defined as

$$E_3(k) = \lambda.E_3(k-1) + (1-\lambda).(y_n{}^2(k).x(k-\tau)) \tag{13}$$

With smoothness parameter $0 < \lambda < 1$

This recursive structure can be adequate for real time T60 estimation, however it presents some limitations. In fact, for an efficient estimation, it is necessary to take into account the convergence time of the estimator which is dependent on the initialization value as well as the smoothing parameter value. The dynamic range of the decay to be used for the estimation of the T60 is dependent on the estimator settings (21).

In Fig.4 we show the decay curve obtained with the HOS based method using the excitation signal described in the previous section.
An impulse response synthesized according to Polack model (T60=0.5s) is used. The reverberant signal is disturbed with an additive white noise (SNR=10dB). A smoothing parameter of 10^{-3} is used for the estimation of the decay.

The third order decay is larger than the energy decay under the same conditions (impulse response and SNR). The useful portion for the estimation of the T60 corresponds to the part of the decay located above a threshold around -25dB for the HOS method and -19dB for the energy decay method.

In Fig.5 we show the decay curves obtained with the HOS based method for the same excitation signal and a Polack impulse response (T60=0.5s) for three SNR levels (10dB, 15dB and 20dB).We notice that the decay is nearly independent of the SNR and the dynamic range remains the same.

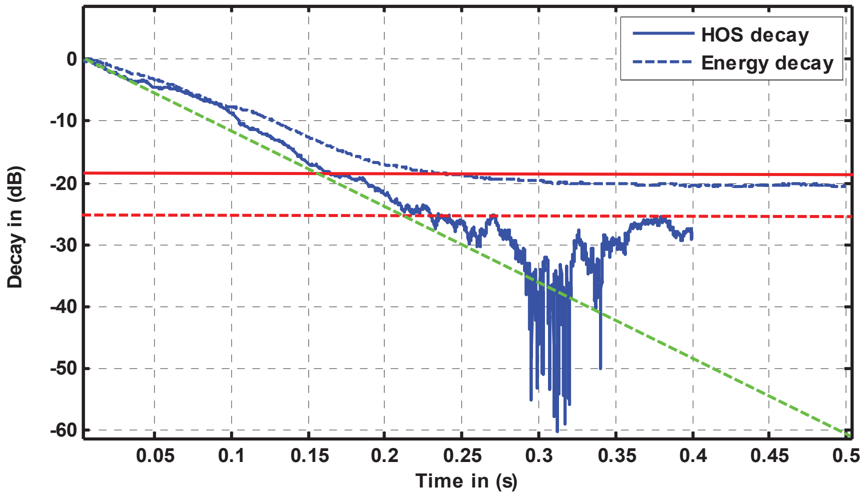

Figure 4 – Comparison between the third order decay and the energy decay. An impulse response synthesized according to Polack model (T60=0.5s) is used. The reverberant signal is disturbed with an additive white noise (SNR=10dB). The dotted red curve marks the end of the dynamic range relative to the HOS method and the solid red curve marks the end of the dynamic range relative to the energy decay method. The green curve corresponds to the undisturbed sound decay (decay with the true reference slope).

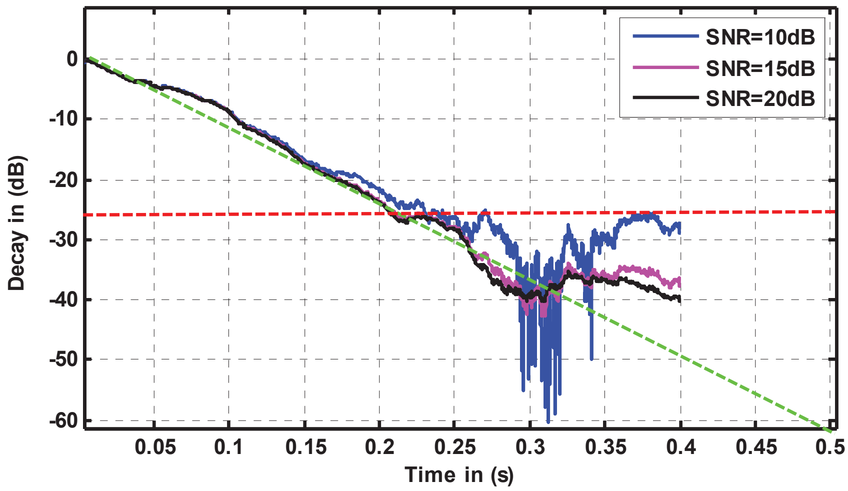

Figure 5 – Robustness of the HOS method to additive noise. An impulse response synthesized according to Polack model (T60=0.5s) is used. The reverberant signal is disturbed with an additive white noise (SNR=10, 15 and 20dB). The dotted red curve marks the end of the dynamic range relative to the HOS method. The green curve corresponds to the undisturbed sound decay (decay with the true reference slope).

4. PERFORMANCES OF THE REVERBERATION TIME ESTIMATION METHOD BASED ON HOS

The effectiveness of the proposed T60 estimation based on HOS decay is tested with impulse responses synthesized according to Polack model (1).
We consider four values of T60: 0.5s, 0.75s, 1s and 2s. For each T60 value, 10 different realizations of impulse responses are considered. The reverberated signal is disturbed with white noise.
Three SNR levels are used. For each impulse response and SNR level configuration, the estimation error is calculated as the difference between the estimated T60 and the target value.

The test results illustrated in Fig.6 show that for the considered values of T60, the median of the estimated values presents an estimation error inferior to 0.1 s.
However, some probable extreme values with estimation error that can reach 0.4s are observed especially for the higher T60 tested value (T60=2). It is noticeable also that the estimation accuracy remains nearly the same when changing the SNR level.

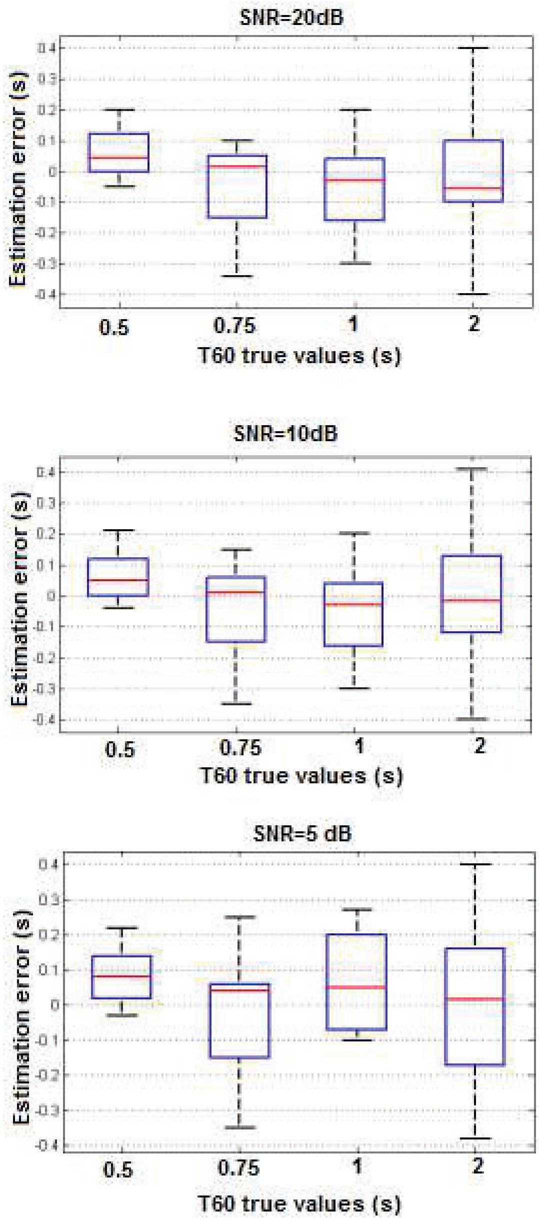

Figure 6 – Evaluation of the HOS estimation method accuracy for a set of synthetic impulse responses having a T60 is equal to 0.5 s, 1s and 2s for three levels of SNR 5dB, 10dB and 20dB. For each T60, 10 different realizations of impulse responses are considered.

The proposed T60 estimation method based on HOS decay is also tested with real impulse responses. In table1, we present the obtained results for three examples of real impulse responses with a T60 ranging from 0.9s to 4s. The reverberant signal is disturbed with additive stationary ambient noise for an SNR equal to 10dB.

According to these three examples, we notice that the estimations of T60 provides by the HOS decay method is always inferior to the true value. The estimation error is the most important for the higher T60 value. These results show that the robustness of the proposed HOS decay method for the estimation of T60 in real conditions needs to be more investigated. This is mainly related to the estimation parameters that need to be adapted to the real situations.

Table 1 – T60 estimation based on HOS method for real impulse response (SNR=10dB)

	Lecture room	Concert hall	Church
True T60	0.9 s	2.1 s	4s
Estimated T60	0.8s	1.5s	1s

5. CONCLUSIONS

In this paper we describe a new method for the estimation of the T60 in noisy conditions. The proposed approach generalizes the estimation method based on the sound energy decay; however it is based on HOS decay and needs the use of an excitation signal with specific statistical properties.

The dynamic range of the decay provided by this method is shown to be larger and nearly independent of the additive noise level. The performance of this method was evaluated using a set of synthetic impulse responses with a T60 ranging from 0.5s to 2s, for an SNR ranging from 5dB to 20dB. The obtained results show that the proposed method performs well for synthetic impulse responses, even though some extreme estimation values with an error that may exceed 0.4s are observed.

However, for the case of real impulse responses, the estimation error may be more important. Thus, the robustness of the method needs to be more assessed. More tests with real impulse responses and background noise must be done as well as a fine analysis of the different parameters used for the HOS decay estimation.

6. ACKNOWLEDGEMENTS

The authors would like to thank Professor Olivier Sicard for his assistance in making available his R-code for the synthesis of signals with asymmetric distribution and high autocorrelation.

REFERENCES

1. European AAL project Age Sensitive ICT Systems for Intelligible City For All, ICityForAll, AAL 2011-4-056.
2. International Standards Organization, "Acoustics Measurement of room acoustic parameters", 2009.
3. Y. Mzah, S. Maarfi, R. Ghozi and M. Jaidane, "Non-Monotonic Impact of Occupancy Level on Reverberation Indicators: Case of a Public Confined Eating Establishment", AES 60th International Conference, Leuven, Belgium, February 35, 2016.
4. V. Desarnaulds, A. P.O. Carvalho and G.Monay, "Church Acoustics and the Influence of Occupancy", Building Acoustics, vol.9, no. 1, pp. 29-47, 2002.
5. M. Hodgson and E. M Nosal, "Effect of noise and occupancy on optimal reverberation times for speech intelligibility in classrooms", J. Acoust. Soc. Am, vol. 111, no 2, p. 931-939 111(2), 2002.
6. K. I. Furuya and A. Kataoka, "Robust speech dereverberation using multichannel blind deconvolution with spectral subtraction", Audio, Speech, and Language Processing, IEEE Transactions on, vol.15, no 5, p.1579- 1591, 2007.
7. K. Kinoshita, M. Delcroix, T. Nakatani and M. Miyoshi, "Suppression of late reverberation effect on speech signal using long-term multiplestep linear prediction", Audio, Speech, and Language Processing, IEEE Transactions on, vol.17, no 4, p.534-545, 2009.
8. W. C. Sabine, Collected Papers on acoustics (Originally 1921), Peninsula Publishing, 1993.
9. M. R. Schroeder, "New method of measuring reverberation time", J.Acoust. Soc. Am., vol. 37, pp. 409412, 1965.
10. K. Lebart, J. M. Boucher and P. N. Denbigh, "A new method based on spectral subtraction for speech de-reverberation", Acta Acustica united with Acustica, 87(3), 359-366, 2001.
11. J. Y. Wen, E. A. P. Habets and A. P. Naylor, "Blind estimation of reverberation time based on the distribution of signal decay rates", Acoustics, Speech and Signal Processing ICASSP IEEE International Conference on (pp. 329-332). IEEE, 2008.
12. R.Talmon and E. Habets, "Blind reverberation time estimation by intrinsic modeling of reverberant speech", Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on. IEEE, 2013.
13. R. Ratnam, D. Jones, L. Wheeler, B. C., O Brien Jr, W. D., Lansing, C. R., Feng, A. S. (2003), "Blind estimation of reverberation time", The Journal of the Acoustical Society of America, 114(5), 2877-2892.
14. T. J. Cox, F. Li and P. Darlington, "Extracting room reverberation time from speech using artificial neural networks", Journal of the audio engineering society, vol. 49, no 4, p. 219-230, 2001.
15. H. W. Llmann, P. Vary, "Estimation of the reverberation time in noisy environments", In Proceedings of International Workshop on Acoustic Echo and Noise Control, 2008.
16. J. Eaton, N. D Gaubitch and P. A. Naylor, (2013, May), "Noise-robust reverberation time estimation using spectral decay distributions with reduced computational cost", In Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on (pp. 161-165). IEEE.
17. M. Karjalainen, P. Ansalo, A. Mäkivirta, T. Peltonen and V. Välimäki, "Estimation of modal decay parameters from noisy response measurements", Journal of the Audio Engineering Society, 50(11), 867-878, 2002.
18. H. Kutruff, Room acoustics, Taylor and Francis, London, fourth edition, 2000.
19. J.-D. Polack, La transmission de l'énergie sonore dans les salles, Ph.D. thesis, Université du Maine, 1988.
20. O. Sicard, La loi normale asymétrique: définition, premières propriétés, estimation des paramètres (http://irem.univ-reunion.fr/IMG/pdf/normale_asymetrique__sicard_2013.pdf), Mars 2013.
21. Y. Mzah and M. Jaidane, "Dynamic Range Variability of Sound Energy Decay for Reverberation Time Estimation in Railway Noisy Environments", International Conference on Advanced Technologies for Signal&Image Processing ATSIP'2016, Monastir-Tunisia, March 2016.