Performances of The Exponential Sinusoidal Audio Model

Zoran N. Milivojevic[1], Predrag Rajkovic[2] and Sladjana M. Milivojevic[3]

Abstract – **In the first part of this paper TLS and Hankel TLS algorythms for determination of parameters for sinusoidal and exponential sinusoidal model of audio and speech signal are described. In the second part performances of exponential sinusoidal model are determined and a comparative analysis of a model for the case of segment processing with the distinguished and poorly distingushed transiency is performed. In the analysis tabular data and time and frequency diagrams are used.**

Keywords – **Exponential sinusoidal modeling, TLS algorithm.**

I. INTRODUCTION

The sinusoidal model (SM) is suitable for representing the harmonic structure of the speech and audio segments. Special conveniences can be seen in speech analysis/synthesis [1,2], speech modification [3], speech coding [4,5] and audio coding [6,7]. The sinusoidal model for the speech and audio signal *s(n)* can be presented in the following form:

$$s(n) \approx \sum_{k=1}^{K} a_k(n) \sin(2\pi f_k(n) n + \phi_k(n)) . \qquad (1)$$

In the sinusoidal model the signal *s(n)* is presented as a summary of components with time variable amplitude a_k, frequency f_k and phase ϕ_k. These parameters are often invariable or slowly variable in the time of analysis (duration time of an analyzed sequence, i.e. segment). Depending on the signal the length of the quasi-stationary segment varies from several *ms* to several hundreds *ms* [8].

Speech and audio signals often contain segments with superimposed noise as well as segments with transient sound. In such cases the model described by means of (1) does not give satisfying results. In [9] shows the model for presenting the audio signal created by enlarging the model described with (1) by adding the noise $\eta(n)$ and transient segment $\tau(n)$:

$$s(n) \approx \sum_{k=1}^{K} a_k(n) \sin(2\pi f_k(n) n + \phi_k(n)) + \eta(n) + \tau(n) . \qquad (2)$$

In standards for audio signal coding, such as MPEG-1 LI, using of model (2) is not explicitly foreseen. The subcomprehensive coding structure [10] is being used instead. The subcomprehensive coding is efficient for coding signal with superimposed noise in a wide frequency range. However, when coding signals with transient segments, the efficiency is considerably smaller. Generally seen, transient sound is difficult to model by means of the sinusoidal model. More qualitative modeling can be achieved by enlarging the number of model parameters, which reduces coding efficiency. For that reason in some coding schemes detecting of transient segments and selecting the code structure with enlarged resolutions in time domain is done first.

One way of solving this problem is audio signal modeling and coding by using of superposition of sinusoid with time slow exponential changes of amplitudes and quasistationary noise $\eta(n)$:

$$s(n) \approx \sum_{k=1}^{K} a_k(n) \cdot e^{-d_k(n) \cdot n} \sin(2\pi f_k(n) n + \phi_k(n)) + \eta(n) , \qquad (3)$$

where d_k is damping factor of the k-th component. The exponential sinusoidal model (ESM) is described in [11]. Its efficiency in modeling transient segments is presented in [12,13]. Determination of model parameters (amplitude a_k, frequency f_k, phase ϕ_k and damping factor d_k) is numerically complex and demands a lot of calculation time.

In this paper algorithms for determining parameters of the exponential sinuous model are described and their performances are determined.

The organization of this paper is as follows. In Section II TLS-ESM algorithm is described. In Section III Hankel TLS algorithm for forming of model parameters is described. In Section IV results of the comparative analysis of the application of SM and ESM models in transient and non-transient sequences are presented.

II. TLS-ESM ALGORITHM

TLS (Total Least Squares) algorithm is used for determining parameters of the exponential sinuous model [14]. For the inlet segment *s(n), n=1,...,N*, by TLS algorithm parameters of the model of L order ($b(l), l = 1,...,L$) are being determined on the condition of minimizing:

$$\sum_{n=1}^{N} (\hat{s}(n) - s(n))^2 = \sum_{n=1}^{N} (\Delta s(n))^2 , \qquad (4)$$

where:

$$\hat{s}(n) = \sum_{l=1}^{L} b(l)(s(n-l) + \Delta s(n-l)), \; n = (L+1),...,N . \qquad (5)$$

[1]Zoran N. Milivojevic, Technical College, Aleksandra Medvedeva 20, Nis, Serbia, E-mail: milivojevic@bankerinter.net

[2]Predrag Rajkovic, Faculty of Mechanical Engineering, Aleksandra Medvedeva 22, Nis, Serbia.

[3]Sladjana M. Milivojevic, Technical Faculty, Svetog Save 65, Cacak, Serbia, E-mail: milivojevic@bankerinter.net

Equation (5) can be written in the form of:

$$\hat{s}(n)=\sum_{k=1}^{K} a_k(n)\cdot e^{-d_k(n)\cdot n}\sin(2\pi f_k(n)n+\phi_k(n)), \quad (6)$$

where the damping factor d_k can be positive, negative or zero. By comparing of equations (3) and (6) it can be seen that it is possible to apply TLS algorithm for determining parameters of ESM model, i.e. for automatic decomposition of an audio sequence in a certain number of damped sinusoids [15].

III. Hankel TLS Algorithm

Due to its great calculation efficiency, Hankel TLS (HTLS) algorithm is used for solving TLS problems. HTLS algorithm has found its intensive application in nuclear magnetic resonance spectroscopy. In the [15] HTLS algorithm is described that for inlet parameters: a) sequence $s(n)$, $n=1,...,N$; and b) model order K_e; generates parameters of the estimated sinusoids (amplitude $\hat{a}_k$, frequency $\hat{f}_k$, phase $\hat{\psi}_k$, damping factors $\hat{d}_k$). The algorithm consists of the following steps:

Step 1: Out of sequence elements *s(n)* Hankel matrix H with dimensions *m*x*n* is formed.

Step 2: SVD (singular value decomposition) of matrix H is determined:

$$H=USV^H . \quad (7)$$

Step 3: Shortened matrices of *Ke* rank are constructed:

$$\hat{H}=U_{K_e}S_{K_e}V_{K_e}^H , \quad (8)$$

where U_{Ke} contains the first K_e columns of the matrix U, V_{Ke} contains the first K_e columns of the matrix V, whereas S_{Ke} is the submatrix (K_e x K_e) of the matrix S.

Step 4: TLS is calculated for the predefined equation system:

$$\overline{U}_{K_e}\approx \underline{U}_{K_e}E , \quad (9)$$

where $\overline{U}_{K_e}$ is obtained from the matrix U_{Ke} after eliminating the first row, $\underline{U}_{K_e}$ is obtained from the matrix U_{Ke} after eliminating the last rowg. K_e values of E are used for estimation of signal poles:

$$\hat{z}_k=e^{\left(j2\pi\hat{f}_k-\hat{d}_k\right)},\quad k=1,...,K_e . \quad (10)$$

Step 5: The equation of the model is formed on the base of signal poles:

$$\hat{s}(n)=\sum_{k=1}^{K_e} c_k\hat{z}_k^{n-1} , \quad (11)$$

where

$$\hat{c}_k=\frac{\hat{a}_k}{e}e^{j\psi_k} . \quad (12)$$

Taking into consideration that the poles are conjugated complex, the model described with (11) can be presented in a reduced form:

$$\hat{s}(n)=\sum_{k=1}^{K}\hat{a}_k(n)\cdot e^{-\hat{d}_k(n-1)}\sin\left(2\pi\hat{f}_k(n-1)+\hat{\phi}_k\right), \quad (13)$$

where

$$\hat{\phi}_k=\hat{\psi}_k+\pi/2,\quad k=1,...,K . \quad (14)$$

The model described with (13) is equivalent to the ESM model described with (3).

Detailed description of HTLS algorithm can be found in [8, 16].

IV. Performances of ESM model

Performances of ESM model with implemented HTLS algorithm will be determined by means of signal-noise ratio (SNR) that is defined in (8):

$$SNR=10\log_{10}\frac{\sum_{n=1}^{N}s^2(n)}{\sum_{n=1}^{N}(s(n)-\hat{s}(n))^2} . \quad (15)$$

Thus defined SNR represents a measure of precision of the modeled signal in relation to the original signal.

Further analyses were carried out on the archivated speech signal whose sampling frequency is F_S=22.050 kHz, by means of a mathematical packet MatLab. Comparative analyses will be performed by an analysis in time and frequency domains on: a) the original speech signal (s), b) the speech signal modeled by means of an sinusoidal model (s_{SM}) and c) the speech signal modeled by means of an exponential sinusoidal model (s_{ESM}). The following examples relate to two types of sequences: a) with not so outstanding transience (signal sequences where periodicalness is expressed) and b) with an outstanding trasiency.

IV.A. Sequencies with not so outstanding transienceitle

Examples of sequences of speech and audio signals with not so outstanding transience are presented in Fig. 1 where the original signal s is shown and modeled signals s_{SM} and s_{SM} for Ke=32. In Fig. 2 the same signals for Ke=128 are shown. In these sequences the signal periodicalness can be seen (pronunciation of vowels, musical signal etc.).

IV.A. Sequences with outstanding transience

Sequences of speech signal with an outstanding effect of transience are modeled for some values of model order K_e (K_e=4,8,16,32,64,128). Time forms of signals are presented in

Fig. 3 (K_e=32) and Fig. 5 (K_e=128). The signal spectra are determined by means of FFT and presented in Fig.4 (K_e=32) and Fig. 6 (K_e=128).

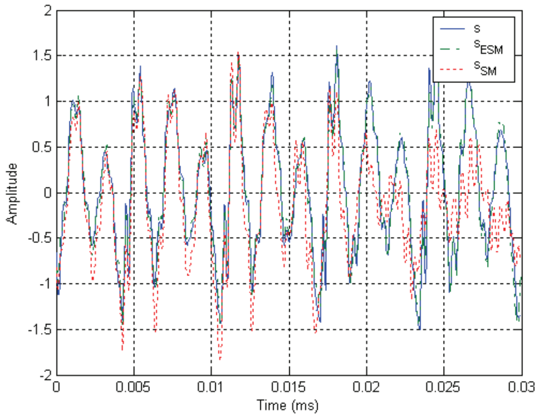

Fig. 1. The sequence of the speech signal for the word 'five' with not so outstanding transience: a) ***s*** the original signal, b) $\boldsymbol{s_{ESM}}$ reconstructed signal on the base of the estimated parameters of ESM model and c) $\boldsymbol{s_{SM}}$ reconstructed signal on the base of the estimated parameters of SM model (*Fs=22.050 kHz,*K_e*=32*).

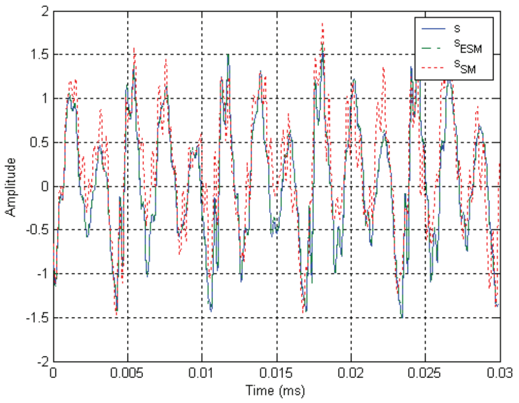

Fig. 2. The sequence of the speech signal for the word 'five' with not so outstanding transience: a) ***s*** the original signal, b) $\boldsymbol{s_{ESM}}$ reconstructed signal on the base of the estimated parameters of ESM model and c) $\boldsymbol{s_{SM}}$ reconstructed signal on the base of the estimated parameters of SM model (*Fs=22.050 kHz,*K_e*=128*).

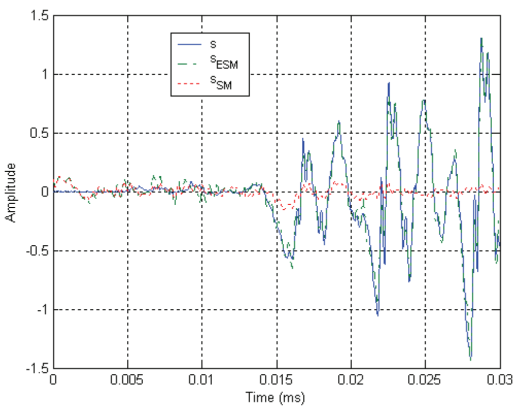

Fig. 3. Transient segment of the speech signal for the word 'five': a) ***s*** the original signal, b) $\boldsymbol{s_{ESM}}$ reconstructed signal on the base of the estimated parameters of ESM model and c) $\boldsymbol{s_{SM}}$ reconstructed signal on the base of estimated parameters of SM model (*Fs=22.050 kHz,*K_e*=32*)

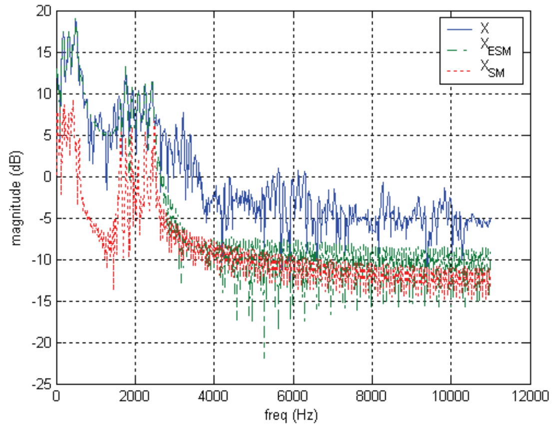

Fig. 4. Spectrum of the transient segment of the speech signal for the word 'five': a) ***s*** the original signal, b) $\boldsymbol{s_{ESM}}$ reconstructed signal on the base of the estimated parameters of ESM model and c) $\boldsymbol{s_{SM}}$ reconstructed signal on the base of the estimated parameters of SM model (*Fs=22.050 kHz,*K_e*=32*)

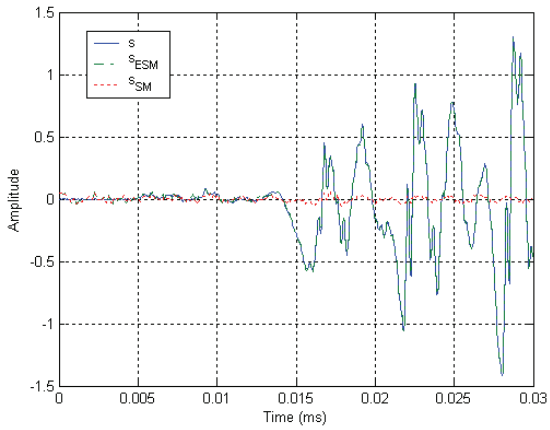

Fig. 5. Transient segment of the speech signal for the word 'five': a) ***s*** the original signal, b) $\boldsymbol{s_{ESM}}$ reconstructed signal on the base of the estimated parameters of ESM model and c) $\boldsymbol{s_{SM}}$ reconstructed signal on the base of estimated parameters of SM model (*Fs=22.050 kHz,*K_e*=128*).

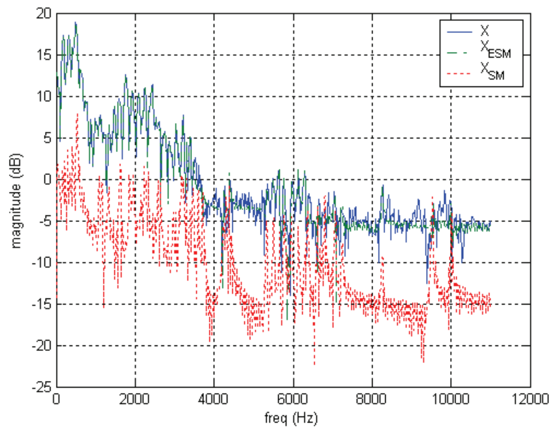

Fig. 6. Spectrum of the transient segment of the speech signal for the word 'five': a) ***s*** the original signal, b) $\boldsymbol{s_{ESM}}$ reconstructed signal on the base of the estimated parameters of ESM model and c) $\boldsymbol{s_{SM}}$ reconstructed signal on the base of the estimated parameters of SM model (*Fs=22.050 kHz,*K_e*=128*).

In Table 1 results of SNR at the sinusoidal and exponential model for the transient and weakly present transience are presented.

TABLE I

ADVANTAGES OF SNR FOR A) TRANSIENT AND B) NOT SO OUTSTANDING TRANSIENT SEGMENT FOR THE CASE OF THE APPLICATION OF SINUOUS AND EXPONENTIAL SINUOUS MODEL, DEPENDING ON THE MODEL ORDER

K_e	Transient		Not so outstanding transient	
	SNR_{SM}	**SNR_{ESM}**	**SNR_{SM}**	**SNR_{ESM}**
4	0.1690	3.0173	6.0962	6.2688
8	0.4863	6.3888	9.8830	10.415
16	0.6693	8.3856	6.8996	11.7304
32	0.2837	14.637	4.4636	15.9509
64	0.1115	22.5617	3.5876	19.7519
128	0.1650	27.9215	5.2516	27.8634
mean values	0.3141	13.8186	6.03	15.33

On the base of time and frequency diagrams, as well as on the base of tabular data for SNR it should be concluded that the ESM model is superior in relation to the SM model. It special advantage is in regard to modeling of signals in transient sequences. In the transient sequence the relation of the mean values is 13.8186/0.311=43.99, whereas in the sequence with not so outstanding transience the relation is 2.54.

V. CONCLUSION

In this paper the exponential sinusoidal audio model with the implemented HTLS algorithm is described. In the second part of this paper the results of testing the application of the sinusoidal and exponential model in modeling the speech signal are presented. Modeling was performed for various operating parameters of the model. As a measure of successfulness, i.e. of precision of modeling, SNR was used. Analysis of the obtained results points to the greater efficiency of ESM in relation to SM in all the values of the model order. In addition to that, the results demonstrate great efficiency in transient segments, i.e., in the concrete example, 43.99 times in relation to SM model. The results testify that the application of ESM model for the speech and audio signal compression is justified in archivating and transition by communication.

REFERENCES

[1] E.B. George, M.J.T. Smit, *Speech analysis/synthesis and modification using an analysis-by-synthesis/overlap-add sinusoidal model*, IEEE Trans. Speech Audio Process. 5 (5) (September 1997) 389–406.

[2] R.J. McAulay, T.F. Quatieri, *Speech analysis-synthesis based on a sinusoidal representation*, IEEE Trans. Acoustics, Speech and Signal Processing 34 (4) (August 1986) 744–754.

[3] Quatieri, R.J. McAulay, *Speech transformations based on a sinusoidal representation*, IEEE Trans. Acoustics, Speech and Signal Processing 34 (6) (August 1986) 1449–1464.

[4] R.J. McAulay, T.F. Quatieri, *Speech Coding and Synthesis*, Elsevier, Amsterdam, 1995, pp. 121–173.

[5] L.B. Almeida, F.M. Silva, *Variable-frequencysy nthesis: an improved harmonic coding scheme*, in: Proceedings of the International Conference on Acoustics, Speech and Signal Processing, San Diego, CA, 1984, pp. 27.5.1–27.5.4.

[6] J. Jensen, R. Heusdens, *A comparison of differential schemes for low-rate sinusoidal audio coding*, in: Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, October 2003, pp. 205–208.

[7] H. Purnhagen, B. Edler, C. Ferekidis, *Object-based analysis/synthesis audio coder for very low bit rates*, in: Proceedings of the 104th AES Convention, Amsterdam, The Netherlands, May1998, Convention paper 4747.

[8] K. Hermus, W. Verhelst, P. Lemmerling,P. Wambacq, S. Huffel, *Perceptual audio modeling with exponentially damped sinusoids*, Signal Processing 85 (2005) 163–176.

[9] S.N. Levine, *Audio representations for data compression and compressed domain processing*, Ph.D. Thesis, Stanford University, December 1998.

[10] T. Painter, A. Spanias, *Perceptual coding of digital audio*, Proceedings of the IEEE, vol. 88(4), April 2000, pp. 451–513.

[11] K. Hermus, W. Verhelst, P. Wambacq, P. Lemmerling, *Total least squares based subband modelling for scalable speech representations with damped sinusoids*, in: Proceedings of the International Conference on Spoken Language Processing, Beijing, China, October 2000, pp. 1129–1132.

[12] J. Jensen, S.H. Jensen, E. Hansen, *Harmonic exponential modeling of transitional speech segments*, in: Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Istanbul, Turkey, June 2000, pp. 1439–1442.

[13] P. Lemmerling, I. Dologlou, S. Van Huffel, *Speech compression based on exact modeling and structured total least norm optimization*, in: Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Seattle, WA, May1998, pp. 353–356.

[14] S. Van Huffel, J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis*, Frontiers in Applied Mathematics, vol. 9, SIAM, Philadelphia, PA, 1991.

[15] K. Hermus, W. Verhelst, P. Wambacq, *Perceptual audio modeling based on Total Least Squares algorithms*, AES 112^{TH} Convention, Munich, Germany, May 10-13, 2002.

[16] S. Van Huffel, H. Chen, C. Decanniere, P. Van Hecke, *Algorithm for time-domain NMR data fitting based on total least squares*, J. Magn. Reson. A 110 (1994) 228–237.