Optimal subband Kalman filter for normal and oesophageal speech enhancement

Rizwan Ishaq[*] and Begoña García Zapirain
Deustotech-LIFE, University of Deusto, Bilbao, Spain

Abstract. This paper presents the single channel speech enhancement system using subband Kalman filtering by estimating optimal Autoregressive (AR) coefficients and variance for speech and noise, using Weighted Linear Prediction (WLP) and Noise Weighting Function (NWF). The system is applied for normal and Oesophageal speech signals. The method is evaluated by Perceptual Evaluation of Speech Quality (PESQ) score and Signal to Noise Ratio (SNR) improvement for normal speech and Harmonic to Noise Ratio (HNR) for Oesophageal Speech (OES). Compared with previous systems, the normal speech indicates 30% increase in PESQ score, 4 dB SNR improvement and OES shows 3 dB HNR improvement.

Keywords: Kalman filter, autoregressive, speech enhancement, weighted linear prediction

1. Introduction

The Kalman filter is considered optimal among other signal enhancement methods, such as Wiener filtering, spectral subtraction, wavelet denoising, etc. [1–3]. Inheritance of speech production model and non-stationary signal processing are the advantages of Kalman filtering over other speech enhancement methods [4]. The Kalman filter is first introduced for speech enhancement by Paliwal, providing clean speech Autoregressive (AR) coefficients, and noise variances using conventional Linear Prediction (LP) [5]. The further modification to [5] is done by estimating AR coefficients recursively through Expectation Maximization (EM) algorithms [6] and modeling colored noise as AR process [7–9]. The Kalman filter is used in frequency subbands for efficient and low complex processing with fewer number of AR coefficients using conventional LP [10–12].

The Oesophageal Speech (OES) is a special type of alaryngeal speech used after the treatment of laryngeal cancer for rehabilitation of voice. Several techniques are available for speech restoration, most important being Oesophageal Speech (OES), Tracheo-Esophageal Speech (TES) and Electrolarynx (EL) speech. The OES is most used method because it requires no external device (EL) and surgery (TES). Despite its advantages, OES has low fundamental frequency and intelligibility, due to irregular vibration of esophagus and noise. The Kalman filter is also utilized to enhance the quality of OES such as, fullband Kalman filtering [13–15] and subband Kalman filtering [12]. Kalman filtering has shown significant enhancement over other methods such as source-filter decomposition [16] for enhancing source and filter [17–20], use of LF voicing source [21,22], and statistical methods [23].

---

*Corresponding author: Rizwan Ishaq, Deustotech-LIFE, University of Deusto, Bilbao, Spain. Tel.: +34 94 413 90 03; Fax: +34 94 413 90 03; E-mail: rizwanishaq@deusto.es.

Speech enhancement using Kalman filter needs optimal speech, noise AR coefficients and variances. Previous methods [4–7,12,13] use conventional LP for this purpose, but the conventional LP has the sensitivity problem with additive background noise [24]. The estimation of noise AR coefficients also needs non-speech activity detector. To overcome these problems, this paper presents the optimal subband Kalman filter by providing optimum AR coefficients for speech and noise signals usingWeighted Linear Prediction (WLP) [24] and Noise Weighting Function (NWF) [25]. The proposed system is evaluated, for normal speech by Perceptual Evaluation of Speech Quality (PESQ) score and Signal to Noise Ratio (SNR) improvement, and for OES using Harmonic to Noise Ratio (HNR). The paper outline is as follow; Section 2 provides the detailed description of the system, Section 3 provides the optimal parameter estimation, followed by simulation results in Section 4 and Conclusion in Section 5.

2. System design

The proposed system (KF-P) components are shown in Figure 1, and subsequent sections provide the detail of every component.

2.1. Analysis filter bank

The analysis filterbank is used to decompose the input speech signal x(n) into different subbands. The x(n) passes through the filterbank of different Linear Time Invariant (LTI) bandpass filters, each having impulse response of $h_k(n)$, mathematically the frequency subband signal is [26]:

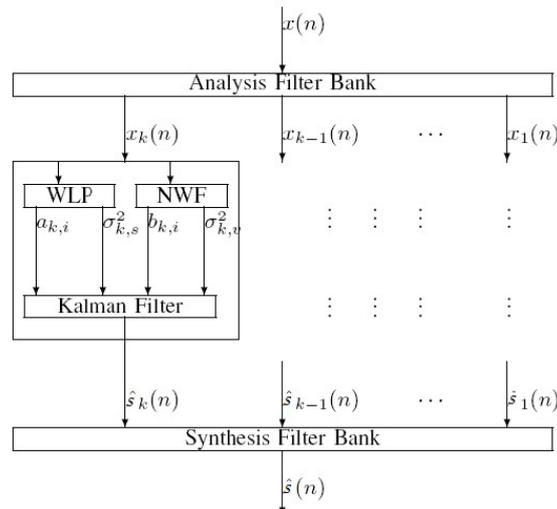$$x_k(n) = x(n) * h_{(k)(n)} \tag{1}$$

where * is the convolution operator.

Fig. 1. Block diagram of proposed system.

2.2. Optimal subband Kalman filter

The subband signal $x_k(n)$ consists of clean speech $s_k(n)$ and noise $v_k(n)$ as follows:

$$x_k(n) = s_k(n) + v_k(n) \tag{2}$$

Both $s_k(n)$ and $v_k(n)$ can be modelled as AR processes of order p and q:

$$s_k(n) = \sum_{i=1}^{p} a_{k,i}(n)s_k(n-i) + \omega_k(n) \tag{3}$$

$$v_k(n) = \sum_{i=1}^{q} b_{k,i}(n)v_{(k)(n-i)} + \Gamma_k(n) \tag{4}$$

where $\omega_k(n)$ and $\Gamma_k(n)$ are uncorrelated additive white Gaussian noises with zero mean and variances $\sigma_{k,\omega}^2$ and $\sigma_{k,v}^2$ respectively. Given $s_k(n) = [s_k(n), \dots, s_k(n-p+1)]$, and $v_k(n) = [v_{(k)}, \dots, v_k(n-q+1)]$, above equations can be written in state-space domain for Kalman filtering:

$$s_k(n) = F_{k,s}s_k(n-1) + G_{k,s}\omega_k(n) \tag{5}$$

$$x_{k,s}(n) = H_{k,s}^T s_k(n) \tag{6}$$

$$v_k(n) = F_{k,v}v_k(n-1) + G_{k,v}\Gamma_k(n) \tag{7}$$

$$x_{k,v}(n) = H_{k,v}^T v_k(n) \tag{8}$$

where state transition matrices $F_{k,s}$ and $F\_(k,v)$ are :

$$F_{k,s} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -a_{k,p} & -a_{k,p-1} & -a_{k,p-2} & \dots & -a_{k,1} \end{bmatrix}_{p,p} \tag{9}$$

$$F_{k,v} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -b_{k,q} & -ba_{k,q-1} & -b_{k,q-2} & \dots & -b_{k,1} \end{bmatrix}_{q,q} \tag{10}$$

and $G_{k,s}, G_{k,v}, H_{k,s}$ and $H\_(k,v)$ are:

$$G_{k,s} = H_{k,s}^T = [0,0,\dots,1]_{px1}, G_{k,v} = H_{k,v}^T = [0,0,\dots,1]_{qx1}$$

Combining the above equations we have:

$$\hat{x}_k(n) = F_k\, \hat{x}_k(n-1) + G_k\, \hat{\omega}_k(n) \tag{11}$$

$$x_k(n) = H_k^T\, \hat{x}_{(k)(n)} \tag{12}$$

where $\hat{x}_k(n), F_k, G_k, H_k,$ and $\hat{\omega}_k(n)$ are given as:

$$\hat{x}_k(n) = [s_k(n), \dots, s_k(n-p+1), v_k, \dots, v_k(n-q+1)]$$

$$F_k = \begin{bmatrix} F_{k,s} & 0 \\ 0 & F_{k,v} \end{bmatrix}, G_k = \begin{bmatrix} G_{k,s} & 0 \\ 0 & G_{k,v} \end{bmatrix}$$

$$H_k^T = [H_{k,s}, H_{k,v}], \qquad \hat{\omega}_k^T(n) = [\sigma_{k,s}^2, \sigma_{k,s}^2]$$

The Kalman filter provides optimal estimate $\hat{s}_k(n)$ of $x_k(n)$, with Kalman gain $K_k(n)$, prediction error covariance $P_k(n|n-1)$ and filtering error covariance $P_k(n)$ [5,12] :

$$\hat{\hat{x}}_k(n) = F_k\hat{x}_{k(n-1)} + K_k(n)\big[x_k(n) - H_k^T F_k \hat{\hat{x}}_k(n-)\big] \tag{13}$$

$$K_k(n) = P_k(n|n-1)H_k^T\big[H_k P_k(n|n-1)H_k^T\big]^{-1} \tag{14}$$

$$P_k(n|n-1) = F_k P_k(n-1)F_k^T + G_k Q_k(n)G_k^T \tag{15}$$

$$P_k(n) = \big[\,I - K_k(n)H_k^T\big]P_k(n|n-1) \tag{16}$$

Where $\hat{\hat{x}}_k(n)$ is the estimated state vector. The covariance matrix of $\hat{\omega}_k(n)$ is:

$$Q_k(n) = E\big\{\hat{\omega}_k(n)\, \hat{\omega}_k^T(n)\big\} = diag\big(\sigma_{k,s}^2, \sigma_{k,v}^2\big)$$

The desired $\hat{s}_k(n)$ is given as:

$$\hat{s}_k(n) = H_{k,s}^T \hat{\hat{x}}_k(n) \tag{17}$$

2.3. Synthesis filter bank

The filterbank summation method is used to reconstruct the enhanced fullband signal $\hat{s}(n)$, using the modified frequency subbands $\hat{s}_k(n)$ [26]:

$$\hat{s}(n) = \sum_{k=1}^K \hat{s}_k(n) \tag{18}$$

3. Optimal parameter estimation

The optimal estimation of speech, noise AR coefficients $a_{k,i}, b_{k,i}$ and their respective variances $\sigma_{k,s}^2, \sigma_{k,v}^2$ are necessary for the optimal results of Kalman filtering. This section provides the optimal AR coefficients and variance for speech using WLP [27], and colored noise optimal parameters by computing noise signal through NWF [25].

3.1. Weighted Linear Prediction (WLP)

The conventional Linear Prediction (LP) sensitivity to additive background noise produces poor AR coefficients [27]. The Conventional LP AR coefficients are estimated using Minimum Mean Square Error (MMSE) criterion. The prediction error $\varepsilon_k^2(n)$ partial derivative with respect to AR coefficients $a_{k,i}$ is set to zero [16,27]:

$$\frac{\partial}{\partial a_{k,i}} \varepsilon_k^2(n) = \frac{\partial}{\partial a_{k,i}} \left(\sum_n (x_k(n) - \sum_{l=1}^p a_{k,l} x_k(n-l))^2 \right) = 0 \tag{19}$$

Considering $r_{k,x}(l,i) = \sum_n x_k(n-l)x_k(n-i)$, the solution to Eq. (19) is:

$$\sum_{i=1}^p a_{k,i} r_{k,x}(l,i) = r_{k,x}(l,0), \quad l = 1,2,\dots,p \tag{20}$$

To overcome the sensitivity of LP to additive background noise, WLP introduced weighting function $\zeta_k(n)$ using the Short Time Energy (STE) of size M, which provides better estimation of AR coefficients by focusing on high SNR region [24]:

$$\zeta_k(n) = \sum_{i=1}^M x_k^2(n-i) \tag{21}$$

Using $\zeta_k(n)$, the AR coefficients can be obtained, considering $r_{k,x}(l,i) = \sum_n \zeta_k(n) x_k(n-l)x_k(n-i)$:

$$\begin{bmatrix} a_{k,1} \\ a_{k,2} \\ \vdots \\ a_{k,p} \end{bmatrix} = \begin{bmatrix} r_{k,x}(1,1) & r_{k,x}(1,2) & \dots & r_{k,x}(1,p) \\ r_{k,x}(2,1) & r_{k,x}(2,2) & \dots & r_{k,x}(2,p) \\ \vdots & \vdots & \vdots & \vdots \\ r_{k,x}(p,1) & r_{k,x}(p,2) & \dots & r_{k,x}(p,p) \end{bmatrix}^{-1} \begin{bmatrix} r_{k,x}(1,0) \\ r_{k,x}(2,0) \\ \vdots \\ r_{k,x}(p,0) \end{bmatrix} \tag{22}$$

The speech variance $\sigma_{k,s}^2$ is [28]:

$$\sigma_{k,s}^2 = r_{k,x}(0,0) - \sum_{i=1}^p a_{k,i} r_{k,x}(0,i) \tag{23}$$

3.2. Noise Weighting Function (NWF)

The noise signal $v_k(n)$ is necessary for the sub-optimum AR coefficients $b_{k,i}$ and noise variance $\sigma_{k,v}$. The NWF estimates $v_k(n)$ using the ratio of noise to speech $\kappa_k(n)$ [25]:

$$\kappa_k(n) = \min\left\{ \left(\frac{\gamma_k(n)}{\eta_k(n)+\epsilon} \right)^{\rho_k}, L_k \right\} \tag{24}$$

where ρ_k and L_k are gain rise exponent and limiting factor respectively. The $\eta_k(n)$ and $\gamma_k(n)$ are:

$$\eta_k(n) = \alpha_k \eta_k(n-1) + (1-\alpha_k)|x_k(n)| \tag{25}$$

$$\gamma_k(n) = \begin{cases} \eta_k(n) & if\ \eta_k(n) \leq \gamma_{(k)}(n-1) \\ (1+\beta_k)(\gamma_k(n-1)) & otherwise \end{cases} \tag{26}$$

where $\alpha_k = \frac{1}{f_{k,s}T_{k,a}}$ and $\beta_k \frac{1}{f_{(k,s)}T_{k,b}}$ are forgetting factor constant and positive constant respectively. The $f_{k,s}$ is subband sampling frequency, and $T_{k,b}$ and $T_{k,a}$ are time constants controlling the noise level. The noise signal $v_k(n)$ is obtained by multiplying $\kappa_k(n)$ to $x_k(n)$:

$$v_k(n) = \kappa_k(n).x_k(n) \tag{27}$$

The AR coefficients $b_{k,i}$ can be calculated by solving following equation:

$$\sum_{i=1}^{q} b_{k,i} r_{k,v}(l,i) = r_{k,v}(l,0), \quad l = 1,2,\dots,q \tag{28}$$

Where $r_{k,v}(l,i) = \sum_n v_k(n-l)v_k(n-i)$. The noise variance $\sigma_{k,v}^2$ is [28]:

$$\sigma_{k,v}^2 = r_{k,v}(0,0) - \sum_{i=1}^{q} b_{(k,i)} r_{k,v}(0,i) \tag{29}$$

4. Evaluation results

4.1. Speech material and system setting

The performance of proposed system (KF-P) for normal speech, is tested by male (5 speakers) and female (5 speakers) speech signals of sampling frequency 16000 Hz [29], which are corrupted by Factory Noise (FN) and Engine Noise (EN) at different Signal to Noise Ratio (SNR) levels (-10, -5, 0, 5 ,10 dB). For OES, Spanish OES vowels \a\, \e\, \i\, \o\, and \u\are used, which are recorded from pa-

thology speech rehabilitation association (6 male speakers uttered each vowel 3 times and center does not have any female speaker).

The system uses 16 subbands filterbank [29]. Each subband uses segment and overlap size of 30 ms and 15 ms respectively. The prediction orders p and q for speech and noise are 12 and 6 respectively.

The proposed system (KF-P) for normal speech, is compared with following available subband based
Kalman filtering algorithms:
– (KF-1){subband Kalman filtering with optimal parameter estimated recursively using conventional
– Linear prediction [28]};
– (KAF-2){Subband Kalman filtering, where parameters estimated in modulation domain with conventional linear prediction [12]};
– (KF-3){ subband Kalman filtering where speech and noise parameters estimated using LMS algorithm [10]}.
The proposed algorithm (KF-P) for OES, is compared with following systems:
– (KF-1){fullband Kalman filter originally used for OES [13]};
– (KF-2){Subband Kalman filter in modulation domain using conventional linear prediction [12]};
– (KF-O){modification to [13] using poles stabilization [15]}.

4.2. PESQ

Table 1 and 2 show the PESQ scores for female and male speech, corrupted by engine noise and factory noise. The proposed system (KF-P) outperforms all other available subband Kalman filtering methods, particularly at low SNR.

4.3. Signal to Noise Ratio (SNR) improvement

The Table 3 and 4 has shown SNR improvement for male and female speech signals corrupted by factory and engine noises. The proposed system (KF-P) has shown around 3-4 dB improvement over the other methods.

Table 1

PESQ Scores for Engine Noise Corrupted Speech

SNR(dB)	-10	-5	0	5	10
Original	0.52	0.90	1.73	1.84	2.48
KF-P	1.34	1.48	2.53	3.18	3.91
KF-1	0.93	1.02	2.48	2.78	3.19
KAF-2	0.78	0.99	1.38	2.72	3.62
KF-3	0.28	0.69	1.59	2.81	3.71
Male Speaker					
Original	0.65	0.80	1.53	1.94	2.78
KF-P	1.17	1.67	2.83	3.28	3.47
KF-1	0.91	1.22	2.18	2.98	3.09
KAF-2	0.96	1.19	2.07	2.52	3.10
KF-3	1.03	1.07	2.09	2.91	3.26
Female Speaker					

Table 2

PESQ Scores for Factory Noise Corrupted Speech

SNR(dB)	-10	-5	0	5	10
Original	0.32	0.89	1.63	2.43	2.97
KF-P	1.20	1.68	2.03	2.78	3.70
KF-1	0.34	0.91	1.98	2.12	2.99
KAF-2	0.98	1.10	1.78	2.52	3.12
KF-3	0.78	0.99	1.89	2.11	3.31
Male Speaker					
Original	0.42	0.79	1.53	2.63	2.45
KF-P	1.40	1.88	2.53	3.21	3.50
KF-1	0.67	1.01	2.11	2.12	2.79
KAF-2	0.98	1.27	2.18	2.75	3.12
KF-3	1.18	1.69	2.39	2.94	3.01
Female Speaker					

Table 3

Signal to Noise Ratio (SNR) improvement for Engine Noise Corrupted Speech

SNR(dB)	-10	-5	0	5	10
KF-P	2.231	3.31	2.84	5.12	8.97
KF-1	0.33	1.98	2.03	3.98	7.3
KAF-2	1.94	2.21	2.38	4.12	7.09
KF-3	0.98	1.89	2.78	3.52	7.12
Male Speaker					
KF-P	3.342	4.42	3.95	6.23	9.89
KF-1	1.44	2.99	3.14	4.99	6.41
KAF-2	2.83	3.32	3.49	4.23	6.10
KF-3	1.09	2.90	3.89	4.63	5.23
Female Speaker					

Table 4

Signal to Noise Ratio (SNR) improvement for Factory Noise Corrupted Speech

SNR(dB)	-10	-5	0	5	10
KF-P	2.321	4.61	5.12	7.02	9.17
KF-1	1.33	2.98	3.03	4.98	5.3
KAF-2	1.54	2.51	3.38	4.52	6.09
KF-3	1.98	2.89	4.78	5.52	7.12
Male Speaker					
KF-P	2.34	4.12	4.95	8.23	9.90
KF-1	1.54	2.89	3.45	5.69	7.21
KAF-2	1.93	3.12	3.19	5.13	6.10
KF-3	1.99	3.90	4.19	6.13	7.93
Female Speaker					

4.4. Harmonic to Noise Ratio (HNR)

The HNR parameter is used, extensively by OES research community for quality measurement [30]. The HNR parameter is calculated using the freely available speech analysis software VoiceSauce [31], according to following settings: segment length and overlap are 30 milliseconds and 15 milliseconds respectively, fundamental frequency estimated using STRAIGHT [32] method in the range of 50 to
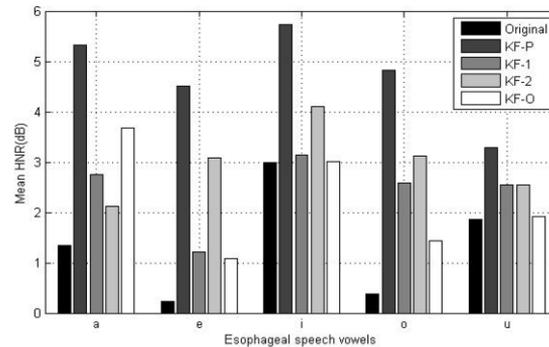
Fig. 2. Mean harmonic to Noise Ratio (HNR) for Spanish vowels.

120 Hz (OES fundamental frequency falls in this range), and prediction order is set to 12. The Figure 2 has shown the mean HNR improvement of 3dB over previously available methods, particularly for vowels \a\, \i\.

5. Conclusion

The system successfully implemented the subband Kalman filtering, by providing optimized parameters for speech and noise. The optimal AR coefficients for speech and its variance are estimated utilizing Weighted Linear Predication (WLP). On the other hand, noise AR coefficients are estimated by calculating Noise Weighting Function (NWF) for subband signals. The method has outperformed all the available subband Kalman filtering both for normal and Oesophageal Speech (OES) signals objectively. The normal speech has shown its superiority through improved PESQ scores and SNR improvement, while for OES signals, improvement has been shown by improved HNR.

Acknowledgement

References

[1]　M.H. Hayes, Statistical Digital Signal Processing and Modelling, John Wiley & Sons, Inc., New York, USA, 1st edi., 1996.

[2]　Z. Ji and F. Wang, Application of the dual-tree complex wavelet transform in biomedical signal denoising, Bio-Medical Materials and Engineering **24** (2014), 109–115.

[3]　B.P. Hu, Y. Li, D.Y. Qiao and T. He, Non-contact physiological signal detection using continuous wave doppler radar, Bio-Medical Materials and Engineering **24** (2014), 993–1000.

[4]　S. So and K.K. Paliwal, Suppressing the influence of additive noise on the kalman gain for low residual noisespeech enhancement, Elsever, Speech Communication **53** (2011), 355–378.

[5]　K.K. Paliwal and A. Basu, A speech enhancement method based on Kalman filtering, IEEE Int. Conf. Acoust., Speech, Signal Processing **12** (1987), 177–180.

[6]   D. Weixiu and P. Driessen, Speech enhancement based on Kalman filtering and EM algorithm, IEEE Pacific. Rim. Conf. on Communication, Computers and Signal Processing **1** (1991), 142–145.

[7]   D.J.G.B. Koo and S.D. Gray, Filtering of colored noise for speech enhancement and coding, IEEE Trans. on Signal Processing **8** (1991), 1732–1742.

[8]   D.C. Popescu and I. Zeljkovic, Kalman filtering of colored noise for speech enhancement, Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing **2** (1998), 997–1000.

[9]   H. Puder, Kalman-filter in subbands for noise reduction with enhanced pitch-adaptive speech model estimation, European Transactions on Telecommunications **13** (2002), 139–148

[10] W.-R. Wu and P.-C. Chen, Subband Kalman filtering for speech enhancement, circuits and systems II, IEEE Transactions on Analog and Digital Signal Processing **45** (1998), 1072–1083.

[11] S. So and K.K. Paliwal, Modulation-domain kalman filtering for single-channel speech enhancement, Speech Commun. **53** (2011), 818–829.

[12] R. Ishaq, B.G. Zapirain, M. Shahid and B. Lovstrom, Subband modulator kalman filtering for single channel speech enhancement, IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, 7442–7446.

[13] B. Garcia and A. Mendez, Oesophageal speech enhancement using poles stablization and Kalman filtering, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2008, 1597–1600.

[14] B. Garcia, J. Vicente, A. Alonso and E. Loyo, Esophageal voices: Glottal flow restoration, Acoustics, Speech and Signal Processing **4** (2005), 141–144.

[15] O.R. Ibon, B. Garcia and Z.M. Amaia, New approach for oesophageal speech enhancement, 5th International Symposium on I/V Communications and Mobile Network (ISVC) **5** (2010), 225–228.

[16] J. Makhoul, Linear prediction: A tutorial review, Proceedings of the IEEE **63** (1975), 561–580.

[17] R.H. Ali and S.B. Jebara, Esophageal speech enhancement using excitation source synthesis and formant structure modification, Signal Processing for Image Enhancement and Multimedia Processing (SITIS) **31** (2006), 615–624.

[18] Y.Y. Qi, Replacing tracheoesophageal voicing source using lpc synthesis, Acoustical Society of America **5** (1990), 1228–1235.

[19] M. Alfredo, P.M. Hector, T. Jorge and O. Patricia, Analysis and recognition of esophageal speech, Symposium on Signal Processing and Information Technology **5** (2006), 101–106.

[20] Y.Y. Q., W. Bernd and B. Ning, Enhancement of female esophageal and tracheoesophageal speech, Acoustical Society of America **98** (1995), 2461–2465.

[21] K. Matsui and N. Hara, Enhancement of esophageal speech using formant synthesis, Proceedings of IEEE International Conference on Acoustic, Speech, and Signal Processing **1** (1999), 81–84.

[22] R. Ishaq and B.G. Zapirain, Adaptive gain equalizer for improvement of esophageal speech, IEEE International Symposium on Signal Processing and Information Technology **1** (2012), 153–157.

[23] H. Doi, K. Nakamura, T. Toda, H. Saruwatari and K. Shikano, Statistical approach to enhancing esophageal speech based on gaussian mixture models, 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP) **10** (2010), 4250–4253.

[24] Y. Kamp and Willems. L. F. Ma, C, Robust single selection for linear prediction analysis of voiced speech, Speech Communication **2** (1983), 69–81.

[25] M.D.N. Westerlund and I. Claesson, Adaptive gain equalizer for speech enhancement, research report, Blekinge Institute of Technology, Karlskrona, 2002.

[26] C.P. Clark, Effective coherent modulation filtering and interpolation of long gaps in acoustic signals, M.S. Dissertation, University of Washington, 2008.

[27] T.K.J. Pohjalainen, R. Saeidi and P. Alku, Extended weighted linear prediction (xlp) analysis of speech and its application to speaker verification in adverse conditions, Interspeech, 2010, 1477–1480.

[28] W.-R. Wu, P.-C. Chen, H.-T. Chang and C.-H. Kuo, Frame-based subband Kalman filtering for speech enhancement, ICSP' 98 International Conference on Signal Processing Proceedings **1** (1998), 682–685.

[29] Pascal Clark Les Atlas and Steven Schimmel. Modulation toolbox version 2.1 for matlab, http://isdl.ee.washington.edu/projects/modulationtoolbox/, University of Washington, September 2010.

[30] R.E. Hillman and Y.Y. Qi, Temporal and spectral estimations of harmonics-to-noise ratio in human voice signals, Acoustic Society of America **102** (1997), 537–543.

[31] A.A. Voicesauce, A program for voice analysis (2012), available at: http://www.seas.ucla.edu/spapl/voicesauce/, May 2014.

[32] I.M.-K.H. Kawahara and A.D. Cheveigne, Restructuring speech representation using a pitchadative time-frequency smoothing and an instantaneous-frequency-based f0 extraction, Speech Communication **27** (1999), 187–207.