Regularized All-Pole Models for Speaker Verification Under Noisy Environments

Cemal Hanilçi, Tomi Kinnunen, Figen Ertaş, Rahim Saeidi, Jouni Pohjalainen and Paavo Alku

***Abstract*—Regularization of linear prediction based mel-frequency cepstral coefficient (MFCC) extraction in speaker verification is considered. Commonly, MFCCs are extracted from the discrete Fourier transform (DFT) spectrum of speech frames. In this paper, DFT spectrum estimate is replaced with the recently proposed regularized linear prediction (RLP) method. Regularization of temporally weighted variants, weighted LP (WLP) and stabilized WLP (SWLP) which have earlier shown success in speech and speaker recognition, is also introduced. A novel type of double autocorrelation (DAC) lag windowing is also proposed to enhance robustness. Experiments on the NIST 2002 corpus indicate that regularized all-pole methods (RLP, RWLP and RSWLP) yield large improvement on recognition accuracy under additive factory and babble noise conditions in terms of both equal error rate (EER) and minimum detection cost function (MinDCF).**

***Index Terms*—Speaker verification, spectrum estimation, linear prediction, regularized linear prediction.**

I. Introduction

SPEAKER verification aims to verify speaker's identity from a given speech signal [1]. A speaker verification system consists of two modules: *feature extraction* (front-end) and *pattern matching* (back-end). In pattern matching, features extracted from a given speech input are compared to the claimed speaker's model. Gaussian mixture models (GMMs) [2] and support vector machines (SVMs) are two popular back-ends, while mel-frequency cepstral coefficients (MFCCs) are commonly used as acoustic features. MFCCs are generally obtained from the discrete Fourier transform (DFT) spectrum of windowed speech frames.

Speaker verification accuracy under clinical and controlled conditions is high but decreases significantly under channel mismatch and in the presence of additive noise. Channel mismatch is the problem of having training and test speech samples from different types of channels or handsets, whereas additive noise refers to other interfering sound sources being added to the speech signal. In literature, several methods have been proposed to tackle channel mismatch and additive noise. These include, for instance, speech enhancement prior to feature extraction and feature normalization using cepstral mean and variance normalization (CMVN). In addition, intersession compensation of speaker models [3] and score normalization [4] are commonly applied.

In [5], the present authors extracted MFCCs from parametric all-pole spectral models based on linear prediction (LP) [6] and its temporally weighted extensions [7]. This led to increased speaker verification accuracy over the standard DFT method under additive noise contamination. A possible explanation for this is that low-order all-pole models, due to smaller number of free parameters in comparison to DFT, exhibit less variations between clean and noisy utterances. In this paper, we would like to explore this further by introducing *regularization* of these all-pole models. In the field of pattern recognition, regularization techniques are commonly used for trading off between training and test errors to enhance classifier generalization [8] but they have been much less studied for feature extraction and speech parameterization [9].

Regularized LP (RLP) [9] is a parametric spectral modeling method motivated from a speech coding point of view for tackling a known problem in that field, over-sharpening of formants. RLP penalizes rapid changes in all-pole spectral envelopes, thereby producing smooth spectra without affecting formant positions. However, RLP has not been applied to any recognition tasks to the best of our knowledge. Intuitively, the use of RLP is justified in speaker verification because it enables computing smooth spectral models and is therefore expected to reduce mismatch between training and test utterances. Since clean speech was used in [9], the present study will address the performance of RLP under additive noise contamination. Moreover, in [9] only boxcar (rectangular) window was used for autocorrelation domain windowing to compute the penalty function. Therefore, we study the effects of different autocorrelation windowing methods on recognition accuracy. Finally, in addition to conventional LP, we extend regularization to the temporally weighted variants of LP, weighted LP (WLP) [5] and stabilized WLP (SWLP) [7].

II. Spectrum Estimation

A. Baseline FFT and LP Methods

MFCC features are generally obtained from the periodogram of a Hamming-windowed speech frame given by

$$S_{\mathrm{FFT}}(f) = \left| \sum_{n=0}^{N-1} w(n)x(n)e^{-j2\pi nf/N} \right|^2, \qquad (1)$$

where f is the discrete frequency index, $\mathbf{x} = [x(0) \ldots x(N-1)]^T$ is a speech frame and $\mathbf{w} = [w(0) \ \ldots \ w(N-1)]^T$ is the Hamming window. The signal $x(n)$ is assumed to be zero outside of the interval $[0, N-1]$.

Cemal Hanilçi and Figen Ertaş are with the Uludağ University, 16059 Bursa, Turkey (e-mail: chanilci@uludag.edu.tr, fertas@uludag.edu.tr). Tomi Kinnunen is with the University of Eastern Finland, FI-80101, Joensuu, Finland (e-mail: tomi.kinnunen@uef.fi). Rahim Saeidi is with the Radboud University Nijmegen, Netherlands (e-mail: rahim.saeidi@let.ru.nl). Jouni Pohjalainen and Paavo Alku are with the Aalto University, FI-00076, Aalto, Finland (e-mail: jpohjala@acoustics.hut.fi, paavo.alku@aalto.fi).

The work of Tomi Kinnunen and Jouni Pohjalainen are supported by Academy of Finland (projects 132129 and 127345). The work of Rahim Saeidi was funded by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 238803.

LP analysis [6] is based on the assumption that a speech sample, $x(n)$, can be predicted as a weighted sum of its p previous samples, $\hat{x}(n) = -\sum_{k=1}^{p} a_k x(n-k)$, where $x(n)$ is the original speech sample, $\hat{x}(n)$ is the predicted sample and p is the predictor order. Usually, the predictor coefficients $\{a_k\}_{k=1}^{p}$ are obtained by minimizing the energy of the prediction residual, $e(n) = x(n) - \hat{x}(n) = x(n) + \sum_{k=1}^{p} a_k x(n-k)$. In the autocorrelation method, the solution for $\mathbf{a}_{\text{opt}}^{\text{lp}} = [a_1, \ldots, a_p]^{\text{T}}$ is given by

$$\mathbf{a}_{\text{opt}}^{\text{lp}} = -\mathbf{R}_{\text{lp}}^{-1}\mathbf{r}_{\text{lp}}, \tag{2}$$

where $\mathbf{R}_{\text{lp}}$ is the Toeplitz autocorrelation matrix and $\mathbf{r}_{\text{lp}}$ is the autocorrelation vector. Given the predictor coefficients, a_k, the LP spectrum is obtained by

$$S_{\text{LP}}(f) = \frac{1}{|1 + \sum_{k=1}^{p} a_k e^{-j2\pi fk}|^2}. \tag{3}$$

B. Temporally Weighted All-pole Models

In contrast to LP, weighted linear prediction (WLP) [10] determines the predictor coefficients by minimizing a temporally *weighted* energy of the prediction error, $E = \sum_n e^2(n)\Psi_n = \sum_n (x(n) + \sum_{k=1}^{p} b_k x(n-k))^2 \Psi_n$, where Ψ_n is a time-domain weighting function. In matrix notation, the optimum predictor coefficients of WLP are computed by

$$\mathbf{b}_{\text{opt}}^{\text{wlp}} = -\mathbf{R}_{\text{wlp}}^{-1}\mathbf{r}_{\text{wlp}}, \tag{4}$$

where $\mathbf{b} = [b_1, \ldots, b_p]^{\text{T}}$ are the predictor coefficients, $\mathbf{R}_{\text{wlp}} = \sum_n \mathbf{x}(n)\mathbf{x}(n)^{\text{T}}\Psi_n$, $\mathbf{r}_{\text{wlp}} = \sum_n x(n)\mathbf{x}(n)\Psi_n$ and $\mathbf{x}(n) = [x(n-1)\ x(n-2)\ \ldots\ x(n-p)]^{\text{T}}$. Note that $\mathbf{R}_{\text{wlp}}$ and $\mathbf{r}_{\text{wlp}}$ correspond to $\mathbf{R}_{\text{lp}}$ and $\mathbf{r}_{\text{lp}}$, respectively, if and only if $\Psi_n = 1$ for all n. The matrix $\mathbf{R}_{\text{wlp}}$ is symmetric but in general does not have Toeplitz structure.

Conventional autocorrelation LP guarantees that the corresponding all-pole model is stable, i.e., a filter whose poles are within the unit circle. For WLP, however, the stability of the all-pole model is not guaranteed. The stability condition of an all-pole model is essential in speech coding and synthesis applications. Besides the coding and synthesis applications, it has been noted that stabilization improves speaker verification performance as well [5]. Thus, stabilized WLP (SWLP) was proposed in [7]. In SWLP, the weighted autocorrelation matrix and the weighted autocorrelation vector are expressed as $\mathbf{R}_{\text{swlp}} = \mathbf{Y}^T\mathbf{Y}$ and $\mathbf{r}_{\text{swlp}} = \mathbf{Y}^T\mathbf{y}_0$, respectively (the original article [7] presents the problem in a slightly different form). The columns of the matrix $\mathbf{Y} = [\mathbf{y}_1\ \mathbf{y}_2\ \ldots\ \mathbf{y}_p]$ are calculated by $\mathbf{y}_{k+1} = \mathbf{B}\mathbf{y}_k$ for $0 \le k \le p-1$, where $\mathbf{y}_0 = [\sqrt{\Psi_1}x(1)\ \ldots\ \sqrt{\Psi_N}x(N)\ 0\ \ldots\ 0]^T$ and $\mathbf{B}$ is a matrix where all the elements are zero outside the subdiagonal and the elements of the subdiagonal, for $1 \le i \le N+p-1$, are

$$\mathbf{B}_{\text{i+1,i}} = \begin{cases} \sqrt{\Psi_{i+1}/\Psi_i}, & \Psi_i \le \Psi_{i+1} \\ 1, & \Psi_i > \Psi_{i+1}. \end{cases} \tag{5}$$

In [10] and [7], short-time energy (STE) was chosen as the weighting function, $\Psi_n = \sum_{i=1}^{M} x^2(n-i)$, where M is the length of the STE window.

C. Regularized Linear Prediction

In regularization, a penalty measure is included in the objective function and the predictor coefficients are calculated by minimizing a modified cost function, $\sum_n (x(n) + \sum_{k=1}^{p} c_k x(n-k))^2 + \lambda\phi(\mathbf{c})$, where $\phi(\mathbf{c})$ is the penalty measure which is a function of the unknown predictor coefficients $\mathbf{c}$ and $\lambda > 0$ is a regularization constant which controls the smoothness of the spectral envelope. In [9], the penalty measure was chosen as

$$\phi(\mathbf{c}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left| \frac{C'(e^{j\omega})}{W(\omega)} \right|^2 d\omega \tag{6}$$

where $1/|W(\omega)|^2$ is a coarse approximation of the spectral envelope and $C'(e^{j\omega})$ is the frequency derivative of the RLP inverse filter, $C(e^{j\omega}) = \sum_{k=0}^{p} c_k e^{-j\omega k}$ with $c_0 = 1$. The advantage of this penalty function is that a closed form non-iterative solution exists and it is computationally efficient. In [9], the coarse spectral envelope $1/|W(\omega)|^2$ was derived from windowed autocorrelation sequence, in which the penalty function was shown to have the following form:

$$\phi(\mathbf{c}) = \mathbf{c}^{\text{T}}\mathbf{DFDc}. \tag{7}$$

Here $\mathbf{c} = [c_1, \ldots, c_p]^{\text{T}}$ are the predictor coefficients, $\mathbf{D}$ is a diagonal matrix where each diagonal element is the corresponding row number and $\mathbf{F}$ is a Toeplitz matrix corresponding to the autocorrelation sequence, $f(m) = r(m)v(m)$, where $r(m)$ is the original autocorrelation sequence, $r(m) = \sum_{n=0}^{N-1} x(n)x(n-m), m = 0, \ldots, p-1$, and $v(m)$ is a window function. The matrix $\mathbf{F}$ represents the denominator term, $W(\omega)$ in (6). The matrix $\mathbf{F}$ is equal to conventional Toeplitz autocorrelation matrix $\mathbf{R}_{\text{lp}}$ when using boxcar (rectangular) window. The optimum predictor coefficients are now given by

$$\mathbf{c}_{\text{opt}}^{\text{rlp}} = -(\mathbf{R}_{\text{lp}} + \lambda\mathbf{DFD})^{-1}\mathbf{r}_{\text{lp}}. \tag{8}$$

D. Extending Regularization for Other All-pole Models and Autocorrelation Lag Windows

Regularization can be imposed on LP, WLP or SWLP methods by using corresponding autocorrelation matrix and vector pair ($\mathbf{R}_{\text{lp}}$ and $\mathbf{r}_{\text{lp}}$; $\mathbf{R}_{\text{wlp}}$ and $\mathbf{r}_{\text{wlp}}$; $\mathbf{R}_{\text{swlp}}$ and $\mathbf{r}_{\text{swlp}}$). As λ increases, the spectral envelope gets smoother and as $\lambda \to 0$, it reduces to conventional LP, WLP or SWLP depending on the way the autocorrelation is computed.

We consider different window functions to compute $\mathbf{F}$ matrix. In [11] and [9] the authors used, respectively, Blackman and boxcar windows to compute $\mathbf{F}$ matrix. We compare these two windows and, additionally, also the Hamming window in speaker verification. In [12], [13], [14], it was shown that the so-called *double* autocorrelation (DAC) sequence can be used for robust estimation of spectral envelope in the presence of additive noise. Thus, besides the different window functions, we use DAC sequence, $f(t) = \sum_{m=0}^{p-1} r(m)r(m-t)$, $t = 0, \ldots, p-1$, to compute $\mathbf{F}$. Differently from [14], we use the first p autocorrelation coefficients $(r(0) - r(p-1))$ when computing DAC sequence.

Figure 1 shows the RLP spectra computed using different windowed autocorrelations $f(m)$ of a voiced speech frame

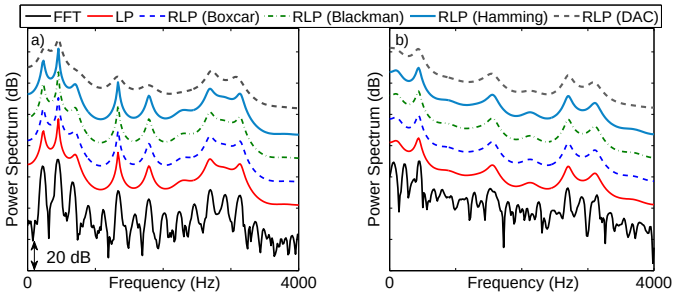

Fig. 1. Short-term spectra of a (a) clean speech frame taken from NIST 2002 SRE and (b) its factory noise corrupted (0 dB SNR) counterpart. The spectra in each plot have been shifted by 10 dB for better visualization. ($\lambda = 10^{-7}$ is used for RLP (DAC) and $\lambda = 10^{-4}$ is used for the RLP with boxcar, Blackman and Hamming windows.)

taken from the NIST 2002 SRE corpus and its 0 dB noisy counterpart. As seen from the figure, regularized methods give a smoother spectrum compared to conventional FFT and LP methods. Different window functions do not show large differences on spectra but estimating $\mathbf{F}$ from DAC does. Dynamic range differences between original and noisy spectra for DAC are smaller compared to conventional LP or RLP with boxcar, Blackman and Hamming windows. We will demonstrate that this leads to considerable improvements in speaker verification accuracy.

III. Experimental setup

Speaker recognition experiments are carried out on the NIST 2002 SRE corpus which consists of conversational telephone speech sampled at 8 kHz and transmitted over different cellular networks. It involves 330 target speakers (139 males and 191 females) and 39259 verification trials (2982 targets and 36277 impostors). For each target speaker, approximately two minutes of training data is available whereas duration of the test utterances varies between 15 seconds and 45 seconds.

Gaussian mixture model with the universal background model (GMM-UBM) [2] is used as the classifier. Test normalization (Tnorm) [4] is applied on the log-likelihood scores for score normalization. Two gender-dependent background models and cohort models for Tnorm with 512 Gaussians are trained using the NIST 2001 SRE corpus.

Power spectral subtraction (as described in [15]) is used as a pre-processing step in the signal domain to suppress additive noise. The MFCC features are extracted from 30 ms Hamming windowed speech frames every 15 ms. Magnitude spectrum estimation method differs depending on the method. Our baseline system uses the FFT magnitude spectrum of windowed frames. For all-pole methods and their regularized versions, the predictor coefficients and short-time spectra are computed as described in Section II. All the all-pole methods use $p = 20$ as in [5]. WLP and SWLP are computed as in [5] by utilizing the STE window function with $M = 20$. The regularization factor λ is 10^{-7}, 10^{-10} and 10^{-10} in RLP, RWLP, and RSWLP, respectively. For the Blackman, boxcar and Hamming windowed RLP the regularization factor λ is fixed to 10^{-4}. The λ value for each method was optimized based on the smallest equal error rate criterion on clean data.

The spectra are processed through a 27-channel triangular filterbank and logarithmic filterbank outputs are converted into MFCCs using the discrete cosine transform (DCT). After RASTA filtering the 12 MFCCs, their first and second order time derivatives (Δ and $\Delta\Delta$) are appended. The last two steps are energy-based voice activity detector (VAD) followed by cepstral mean and variance normalization (CMVN).

As the performance criteria, we consider both equal error rate (EER) and minimum detection cost function (MinDCF). EER is the threshold value at which false alarm rate (P_{fa}) and miss rate (P_{miss}) are equal and MinDCF is the minimum value of a weighted cost function which is given by $0.1 \times P_{\text{miss}} + 0.99 \times P_{\text{fa}}$. Detection error tradeoff (DET) curves are also presented to show full behavior of the proposed methods.

For additive noise contamination, we use *factory2* (which we refer to as *"factory noise"*) and *babble* noises from NOISEX-92[1]. In the noisy experiments, the target speaker models, background models and Tnorm cohort models are trained using the original data and noise is added to test samples with five different average segmental signal-to-noise-ratios (SNRs): $\text{SNR} \in \{\text{clean}, 20, 10, 0, -10\}$ dB, where *clean* refers to the original NIST samples.

IV. Experimental results

We first examine the effect of different window functions, $v(m)$, to compute $\mathbf{F}$ matrix in RLP method as described in Section II. The EER and MinDCF values for different window functions are given in Table I. As seen from the table, different window functions do not show large differences on recognition accuracy as expected from Fig. 1. However, using the DAC sequence to compute $\mathbf{F}$ matrix improves recognition accuracy extensively.

Next, we analyze regularization of the temporally weighted all-pole methods, RWLP and RSWLP, using the DAC sequence. The results are given in Table II. Figure 2 shows the DET plots of each regularized and unregularized all-pole method in comparison to the baseline FFT method for babble noise at SNR level of -10 dB. Recognition accuracy of all methods degrades under additive noise as expected. The following observations can be made:

- In **clean** condition, LP, RLP and WLP methods slightly outperform the baseline FFT technique.
- For **factory noise** contamination, RLP outperforms other methods at low SNR levels (0 dB and -10 dB). RWLP and RSWLP show minor improvements over all-pole methods at high SNR levels (20 dB and 10 dB). In terms of MinDCF, RLP outperforms the other methods at low SNRs (0 dB and -10 dB) while RWLP wins at high SNRs (10 dB and 20 dB)
- For **babble noise**, RLP achieves the smallest EER in nearly all cases (WLP is slightly better at 20dB). In terms of MinDCF, WLP gives smaller MinDCF values at high SNR levels. In the noisier cases, RLP yields the smallest values.

V. Conclusion

Regularization of all-pole models was introduced for robust speaker verification. The proposed methods outperformed standard DFT and LP techniques under two different additive

[1]http://www.speech.cs.cmu.edu/comp.speech/Section1/Data/noisex.html

TABLE I
EFFECT OF AUTOCORRELATION DOMAIN WINDOW FUNCTION USED FOR COMPUTING THE **F** MATRIX IN RLP

	SNR (dB)	Equal error rate (%) Boxcar	Blackman	Hamming	DAC	MinDCFx100 Boxcar	Blackman	Hamming	DAC
	clean	7.57	7.52	**7.37**	7.38	3.07	**3.02**	3.03	3.03
Factory	20	7.81	**7.78**	8.04	7.84	3.18	3.18	**3.16**	3.19
	10	8.75	8.85	8.85	**8.38**	3.57	3.55	3.57	**3.45**
	0	10.29	10.02	10.16	**9.41**	4.17	4.16	4.16	**3.81**
	-10	15.02	15.08	15.45	**13.61**	6.10	6.15	6.06	**5.81**
Babble	20	7.81	7.81	**7.78**	7.90	3.19	3.15	**3.14**	3.30
	10	8.92	8.51	8.68	**8.35**	3.44	3.41	**3.37**	3.46
	0	10.94	11.05	11.20	**9.61**	4.32	4.27	4.26	**3.96**
	-10	20.12	20.92	20.73	**16.93**	7.55	7.76	7.65	**6.63**

TABLE II
SPEAKER RECOGNITION PERFORMANCE UNDER ADDITIVE NOISE (DAC SEQUENCE IS USED FOR REGULARIZED METHODS). FOR A GIVEN NOISE TYPE AND SNR LEVEL, ALL THE DIFFERENCES ARE STATISTICALLY SIGNIFICANT WITH 95% CONFIDENCE ACCORDING TO MCNEMAR'S TEST.

	SNR (dB)	Equal error rate (%) FFT	LP	RLP	WLP	RWLP	SWLP	RSWLP	MinDCFx100 FFT	LP	RLP	WLP	RWLP	SWLP	RSWLP
	clean	7.65	7.44	**7.38**	7.48	8.10	7.81	7.94	3.07	3.05	3.03	**2.99**	3.33	3.08	3.41
Factory	20	8.08	7.83	7.84	7.81	**7.75**	8.22	7.85	3.25	3.22	3.19	**3.12**	3.14	3.21	3.24
	10	9.32	8.50	8.38	8.79	**8.32**	9.11	8.50	3.64	3.56	3.45	3.57	**3.32**	3.62	3.45
	0	10.46	9.93	**9.41**	10.34	9.62	10.06	9.59	4.13	4.21	**3.81**	4.19	3.92	4.17	3.92
	-10	15.35	14.96	13.61	15.19	13.86	14.35	**13.32**	6.63	6.14	**5.81**	6.19	6.03	5.94	5.87
Babble	20	7.83	7.78	7.90	**7.71**	8.21	8.11	8.17	3.14	3.12	3.30	**3.09**	3.35	3.19	3.44
	10	8.85	8.58	**8.35**	8.70	8.48	8.78	8.65	**3.44**	3.48	3.46	3.46	3.53	3.56	3.64
	0	11.62	11.23	**9.61**	11.47	10.29	10.93	9.99	4.53	4.34	**3.96**	4.49	4.35	4.38	4.27
	-10	21.27	20.35	**16.93**	21.02	18.40	19.69	17.64	8.05	7.67	**6.63**	7.90	7.22	7.65	7.04

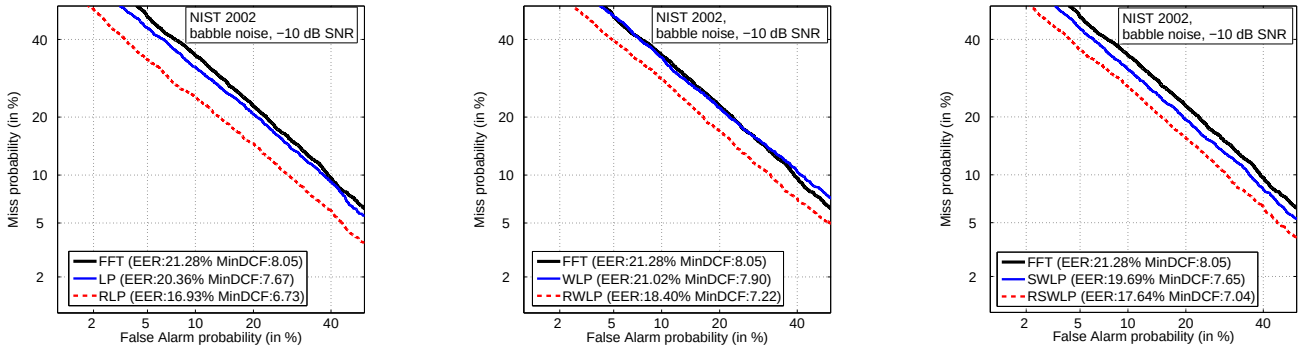

Fig. 2. DET plots for different spectrum estimators under -10 dB SNR babble noise (DAC sequence is used for regularized methods).

noise types, factory and babble noises. In general, regularization using the DAC sequence yielded considerable improvement on the recognition performance especially at low SNRs for conventional and temporally weighted all-pole methods. In summary, the regularized LP based spectrum estimation holds promise for speaker verification in noisy conditions. Adaptive selection of λ based on estimated SNR level or fundamental frequency (as in [9]) is a potential area of future studies.

REFERENCES

[1] T. Kinnunen and H. Li, "An overview of text-independent speaker recognition: from features to supervectors," *Speech Comm.*, vol. 52, no. 1, pp. 12–40, Jan. 2010.

[2] D.A. Reynolds, T.F. Quatieri, and R.B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Dig. Sig. Proc.*, vol. 10, no. 1, pp. 19–41, Jan. 2000.

[3] P. Kenny, G. Boulianne, P. Ouellet, and P. Dumochel, "Joint factor analysis versus eigenchannels in speaker recognition," *IEEE Trans. Audio, Speech and Lang. Proc.*, vol. 15, no. 4, pp. 1435–1447, May 2007.

[4] R. Auckenthaler, M. Carey, and H. Lloyd-Thomas, "Score normalization for text-independent speaker verification systems," *Dig. Sig. Proc.*, vol. 10, no. 1-3, pp. 42–54, Jan. 2000.

[5] R. Saeidi, J. Pohjalainen, T. Kinnunen, and P. Alku, "Temporally weighted linear prediction features for tackling additive noise in speaker verification," *IEEE Sig. Proc. Lett.*, vol. 17, no. 6, pp. 599–602, June 2010.

[6] J. Makhoul, "Linear prediction: a tutorial review," *Proc. of the IEEE*, vol. 64, no. 4, pp. 561–580, Apr. 1975.

[7] C. Magi, J. Pohjalainen, T. Bäckström, and P. Alku, "Stabilized weighted linear prediction," *Speech Comm.*, vol. 51, no. 5, pp. 401–411, April 2009.

[8] Trevor Hastie, Robert Tibshirani, and Jerome Friedman, *The Elements of Statistical Learning*, Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2001.

[9] L. A. Ekman, W. B. Kleijn, and M. N. Murthi, "Regularized linear prediction of speech," *IEEE Trans. Audio, Speech and Lang. Proc.*, vol. 16, no. 1, pp. 65–73, Jan. 2008.

[10] C. Ma, Y. Kamp, and L. Willems, "Robust signal selection for linear prediction analysis of voiced speech," *Speech Comm.*, vol. 12, no. 1, pp. 69–81, March 1993.

[11] M. N. Murthi and W. B. Kleijn, "Regularized linear prediction all-pole models," in *IEEE Speech Coding Workshop*, 2000, pp. 96–98.

[12] D. Mansour and B.H. Juang, "The short-time modified coherence representation and noisy speech recognition," *IEEE Trans. Acoust. and Sig. Proc.*, vol. 37, no. 6, pp. 795–804, Jan. 1989.

[13] T. Shimamura and N. D. Nguyen, "Autocorrelation and double autocorrelation based spectral representations for a noisy word recognition systems," in *Interspeech*, 2010, pp. 1712–1715.

[14] H. Kobatake and Y. Matsunoo, "Degraded word recognition based on segmental signal-to-noise ratio weighting," in *ICASSP*, 1994, pp. 425–428.

[15] P. C. Loizou, *Speech Enhancement: Theory and Practice*, CRC Press, 2007.