PRINCIPAL COMPONENT ANALYSIS IN STRESS DETECTION TECHNIQUE

Milan BOŠTÍK, Doctoral Degree Programme (2)
Dept. of Radio Electronics, FEEC, BUT
E-mail: bostik@feec.vutbr.cz

Supervised by Dr. Milan Sigmund

ABSTRACT

In this study, the analysis of glottal pulse shape was processed. The Iterative Adaptive Inverse Filtering technique was used for obtaining glottal pulses. Four Liljencrant-Fant's parameters of glottal pulse were computed. Both the Principal Component Analysis and Karhunen-Loeve Transformation was applied on these parameters because of parameters decorrelation and both methods were compared together.

1 INTRODUCTION

Speech influenced by psychical stress can be identified e.g. by different time lengths of the phonemes or by different time lengths of speech pauses between two words. The usually used methods for identifying stress start from the time distribution of single phonetic parts of words or sentences [1]. Classifiers, based on the pitch period detection and its variation in time, are also often used. The statistical evaluation to examine e.g. the distribution function of the first two formants or the distribution of time samples is commonly used. All procedures mentioned above have one common factor, namely that a long time recordings have to be processed (for statistical methods the long time recordings are necessary).

In this contribution the method of recognizing emotional stress, based on the analysis of long time recordings is discussed. Method called Iterative Adaptive Inverse Filtering (IAIF) [2] was used for obtaining glottal pulses. Liljencrant-Fant's (LF) model was used for approximating the glottal pulses. The description of the analysis of the speech signal using LF model can be found in [3]. This model estimates four parameters of glottal pulses and can also be used for speech signal synthesis and it is possible to change the parameters of this model in order to imitate the voice of a specific person. Some parameters of glottal pulses, obtained by the LF model, are especially proper for "abnormal" speaker state identification [4, 5]. Both the Principal Component Analysis (PCA) and Karhunen-Loeve Transformation (KLT) were used for LF parameters decorrelation. The methods were applied to sound recordings made at a diploma work defence, under the influence of speakers' examination stress.

2 METHODS

2.1 IAIF TECHNIQUE

It is possible to expect better results if the IAIF [2] is used for estimating the glottal pulses instead of using common inverse filtering [4]. Block diagram of IAIF is shown in Fig. 1.

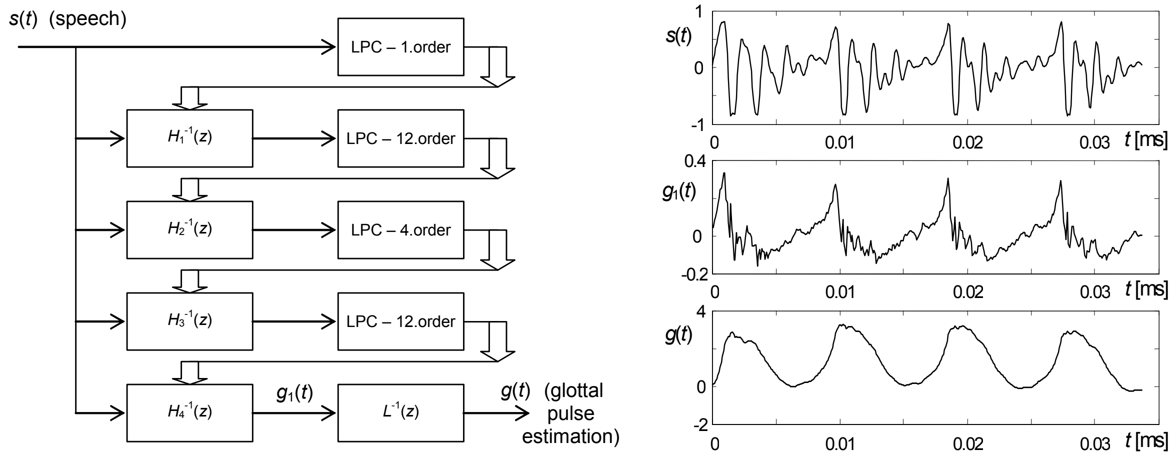

Fig. 1: *IAIF block diagram – left, typical waveforms in block diagram – right.*

Description of blocks from the diagram in Fig. 1 is following:

$H_x^{-1}(z)$ – inversion filter corresponds to Eq. 1, where a_i are LPC coefficients of M^{th} order.

$$H^{-1}(z) = 1 + \sum_{i=1}^{M} a_i z^{-i} \tag{1}$$

Block LPC-1.order – estimating the contribution of glottal pulses to final speech signal. Filter $H_1^{-1}(z)$ is proper filter for obtaining this contribution in condition that the pulses are slow component of speech. Block LPC-12.order (upper block) – signal that inputs into this block is speech without glottal pulses information. Now it is possible to estimate characteristics of vocal tract better than from raw speech. The vocal tract characteristics are represented by the filter $H_2^{-1}(z)$ of 12. order. For better estimate, next iteration is processed by LPC-4.order, $H_3^{-1}(z)$, LPC-12.order (lower block), $H_4^{-1}(z)$. Filter $L^{-1}(z)$ – by this filter lips radiation is removed. Filter corresponds to Eq. 2, where $\lambda = (0.7\text{-}1)$.

$$L^{-1}(z) = \frac{1}{1 - \lambda \cdot z^{-1}} \tag{2}$$

2.2 LF MODEL

Glottal pulse approximation using the LF model uses, as the approximation curve, the exponential function combined with harmonic function as can be seen in Fig. 2. Functions $v_{g1}(t)$ and $v_{g2}(t)$ are two parts of the approximation curve and together they form approximation function $v_g(t)$.

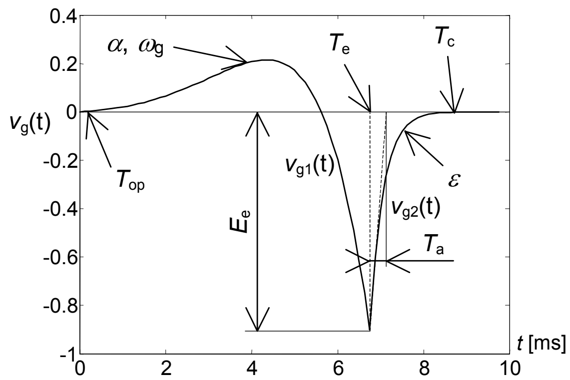

Fig. 2: *Approximation function and the meaning of individual parameters.*

Variables T_{op}, T_e, T_c and time interval T_a are important parameters and their meaning can be clear from Fig. 2. The remaining variables E_e, ω_g, α and ε are the LF parameters sought.

2.3 PRINCIPAL COMPONENT ANALYSIS

The mathematical technique used in PCA is called eigen analysis. It is solved for the eigenvalues and eigenvectors of a square symmetric matrix with sums of squares and cross products. The eigenvector associated with the largest eigenvalue has the same direction as the first principal component. Let $x_v^i(k)$ is i^{th} component of k^{th} feature vector $\mathbf{x}_v$ of class "v", where i = 1-N and N is number of features. Dispersion matrix $\mathbf{T}_v$ can be computed as follows

$$T_v(i,j) = \frac{1}{K_v}\sum_{k=1}^{K_v}\left[\left(x_v^i(k) - \mu_v^i\right)\cdot\left(x_v^j(k) - \mu_v^j\right)\right] \tag{3}$$

where K_v is number of feature vectors for class "v" and μ_v^i is mean value of i^{th} feature. Let $\mathbf{V}_v$ is matrix, where each column contains eigenvector of dispersion matrix $\mathbf{T}_v$. PCA of feature vector $\mathbf{x}_v$ can be expressed as follows

$$\hat{\mathbf{x}}_v = \mathbf{x}_v \cdot \mathbf{V}_v \tag{4}$$

where $\hat{\mathbf{x}}_v$ is new feature vector after PCA.

2.4 KARHUNEN-LOEVE TRANFORMATION

The KLT was used for glottal pulse evaluation. Let $x_v^i(k)$ is i^{th} component of k^{th} feature vector $\mathbf{x}_v$ of class "v", where i = 1-N and N is number of features. Covariance matrix $\mathbf{C}_v$ can be computed as

$$C_v(i,j) = \frac{1}{K_v}\sum_{k=1}^{K_v}\left[x_v^i(k)x_v^j(k)\right] - \mu_v^i\mu_v^j \tag{5}$$

where K_v is number of feature vectors for class "v" and μ_v^i is mean value of i^{th} feature. Let $\mathbf{V}_v$ is matrix, where each column contains eigenvector of covariance matrix $\mathbf{C}_v$. Karhunen-Loeve transformation of feature vector $\mathbf{x}_v$ can be processed as

$$\hat{\mathbf{x}}_v = \mathbf{x}_v \cdot \mathbf{V}_v \tag{6}$$

where $\hat{\mathbf{x}}_v$ is new feature vector after KLT.

3 RESULTS

Described algorithms were applied to speech data from 12 male speakers. About 200 segments (segment length was 20 ms) of speech were extracted from each of them.

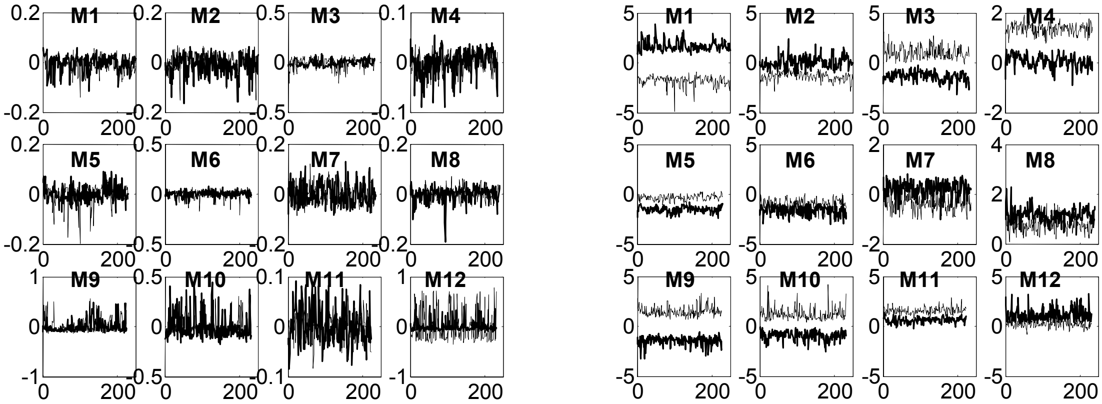

Fig. 3: *Comparing the feature obtained by PCA – left, by KLT – right, vertical value means feature (the component of vector $\hat{\mathbf{x}}_v$). Bold solid line is for normal state, dotted line is for stressed state of speaker.*

All four LF parameters (feature vector $\{x_v^i(k)\}_{i=1,2,3,4}=\{\alpha(k), \omega_g(k), E_e(k), \varepsilon(k)\}$) were used for both PCA and KLT analysis that result in one parameter (only third eigenvector was chosen from matrix $\mathbf{V}_v$ for computing the new feature vector $\hat{\mathbf{x}}_v$ that contains only one parameter). Comparing the results can be seen in Fig. 3. Feature obtained by PCA is shown in Fig. 3 – left and feature obtained by KLT is shown in Fig. 3 – right.

4 DISCUSSION

Analysed features need to have a special distribution in feature space, because both the PCA and KLT analysis is only proper for decorrelating the features having different principal component for different recognizing classes. Big differences between PCA and KLT can be seen in Fig. 3. For further work, it could be better to change LF approximation function for another one, e.g. beta-function model with two parameters [2].

5 CONCLUSION

By comparing the diagrams in Fig. 3 it was found that the PCA technique is less proper for speaker stress detection than KLT technique. Parameters obtained by PCA for normal and stressed state were mixed together – there were no differences between them (Fig. 3 - left). Parameters obtained by KLT for normal and stressed state were quite separated (Fig. 3 - right). Though the reliability of using KLT for stress detection is only about seventy percents (eight positive cases of twelve).

ACKNOWLEDGEMENT

This work was supported by the grant GA CR No. 102/03/H109.

REFERENCES

[1] Vloeberghs, C. at al. *The Impact of Speech Under „Stress“ on Military Speech Technology.* Technical Report 10, RTO, Canada Communication Group Inc. 2000.

[2] Waters, J., Nunn, S., Gillcrist, B., Voncolln, E. *Detection of Stress by Voice: Analysis of the Glottal Pulse*. Technical Report 1652, RTO, Canada Communication Group Inc. 1994.

[3] Iseli, M. R. a Alwan, A. Inter- and Intra-speaker Variability of Glottal Flow Derivative using the LF Model. In *Proceedings of ICSLP-2000*, vol.1, pp. 477-480, Beijing, China, 2000.

[4] Boštík, M. Influence of Speaker Stress on Glottal Pulses Form. In *Proceedings of Radioelektronika 2003*. Brno 2003, p. 451-454.

[5] Boštík, M., Sigmund, M. Speaker Stress Detection by Analysis of Glottal Excitation. In *Proceedings of the 3rd International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications – MAVEBA 2003.* Firenze 2003, p. 87-90