USE OF CEPSTRUM-BASED PARAMETERS FOR AUTOMATIC PATHOLOGY DETECTION ON SPEECH

Analysis of Performance and Theoretical Justification

Rubén Fraile, Juan Ignacio Godino-Llorente, Nicolás Sáenz-Lechón, Víctor Osma-Ruiz
Department of Circuits & Systems Engineering, Universidad Politécnica de Madrid
Carretera de Valencia Km 7, 28031 Madrid, Spain
rfraile@ics.upm.es, igodino@ics.upm.es, nicolas.saenz@upm.es, vosma@ics.upm.es

Pedro Gómez-Vilda
Department of Computer Systems' Architecture and Technology, Universidad Politécnica de Madrid
Campus de Montegancedo s/n, Boadilla del Monte, 28660 Madrid, Spain
pedro@pino.datsi.fi.upm.es

Keywords: Speech analysis, Pattern classification.

Abstract: The majority of speech signal analysis procedures for automatic pathology detection mostly rely on parameters extracted from time-domain processing. Moreover, calculation of these parameters often requires prior pitch period estimation; therefore, their validity heavily depends on the robustness of pitch detection. Within this paper, an alternative approach based on cepstral-domain processing is presented which has the advantage of not requiring pitch estimation, thus providing a gain in both simplicity and robustness. While the proposed scheme is similar to solutions based on Mel-frequency cepstral parameters, already present in literature, it has an easier physical interpretation while achieving similar performance standards.

1 INTRODUCTION

Analysis of recorded speech is an attractive method for pathology detection since it is a low-cost non-invasive diagnostic procedure (Boyanov and Hadjitodorov, 1997). Although there is a wide range of causes for pathological voice (functional, neural, laryngeal, etc.) and a correspondingly wide range of acoustic parameters has been proposed for its detection (see (Jackson-Menaldi, 2002) for summarising tables and typical values), these intend to detect speech signal features that may be roughly classified in only three classes (Godino-Llorente et al., 2006b):

- *Short-term frequency perturbations*: both in fundamental frequency and in formants.
- *Short-term amplitude perturbations*.
- *Noise* or, more specifically, speech-to-noise ratio.

Calculation of above-mentioned acoustic parameters requires previous and reliable detection of speech fundamental frequency (pitch) (Deliyski, 1993) (Boyanov and Hadjitodorov, 1997). Nevertheless, pitch detection is not an easy task due to its sensitiveness to noise, signal distorsion, speech formants, etc. (Boyanov et al., 1993).

An alternative approach to speech signal analysis is doing it in cepstral domain, more specifically in Mel-frequency cepstral domain. Such approach, consisting in classifying patterns of so-called Mel-frequency cepstral coefficients (MFCC), does not require prior pitch estimation and has proven to be fairly robust against different kinds of speech distortion (Bou-Ghazale and Hansen, 2000), including that of telephone channel (Fraile et al., 2007), and reasonably independent of the particular way in which computations may be implemented (Ganchev et al., 2005). For these reasons, their application to automatic voice pathology detection has been proposed during the last years (Godino-Llorente and Gómez-Vilda, 2004). Yet, to authors' knowledge, up to now no physical explanation exists on the meaning of MFCC and their relevance on pathology detection.

Within this paper, a new scheme for automatic voice pathology detection is proposed. This lies half-way between usual cepstral domain and Mel-frequency cepstral domain. Namely, it takes profit from the conceptual interpretation of cepstral processing of speech signals (Deller et al., 1993), the pattern separation capability of cepstral distances (Rabiner and Juang, 1993) and the smoother spectrum estimation provided by the filter banks in MFCC calculation (Rabiner and Juang, 1993). The mathematical formulation of both cepstrum and MFCC parameters is revised in section 2, while the newly proposed set

of parameters is introduced in section 3. The results from the application of these features to the detection of pathologies on voices belonging to a commercial database are reported in section 4. Last, the conclusions are presented in section 5.

2 MATHEMATICAL FORMULATION

2.1 Short-time Fourier Transform

As stated in previous section, the variability of speech signal is a key feature for pathology detection. The need for detecting such variability leads to the convenience of employing short-time techniques for speech processing. For this reason, in the following lines the mathematical framework for short-time processing of speech provided in (Deller et al., 1993) is revised.

Let $x[n]$ be a speech signal composed by N samples ($n = 0 \cdots N-1$) obtained at a sampling frequency equal to f_s; then it can be segmented in frames defined by:

$$f[n;m] = x[n] \cdot w[m-n] \tag{1}$$

where $w[n]$ is the framing window:

$$w[n] = 0 \;\; if \;\; n < 0 \;\; or \;\; n \geq L \tag{2}$$

and L is the frame length. Consequently, $f[n;m]$ has non-zero values only for $n \in [m-L+1, m]$. If consecutive speech frames are overlapped a number of l_0 samples, then m may have the following values:

$$m = L + p \cdot (L - l_0) - 1 \tag{3}$$

where p is the frame index and it is an integer such that:

$$0 \leq p \leq \frac{N-L}{L-l_0} \tag{4}$$

Considering the relation between the frame shift m and the frame index p, frames without time shift reference may be renamed as:

$$\begin{aligned} g_p[n] &= f[n+m-L+1;m] = \\ &= f[n+p\cdot(L-l_0);m] = \\ &= x[n+p\cdot(L-l_0)] \cdot w[(L-1)-n] \end{aligned} \tag{5}$$

where $n = 0 \cdots L-1$. From these speech frames, the short-term Discrete Fourier Transform (stDFT) is computed as:

$$S_p(k) = \sum_{n=0}^{N_{DFT}-1} \widetilde{g}_p[n] \cdot e^{-j \cdot \frac{2\pi}{N_{DFT}} \cdot kn} \tag{6}$$

where N_{DFT} is the number of points of the stDFT, $k = 0 \cdots N_{DFT} - 1$ and:

$$\widetilde{g}_p[n] = \begin{cases} g_p[n] & if \quad 0 \leq n < L \\ 0 & otherwise \end{cases} \tag{7}$$

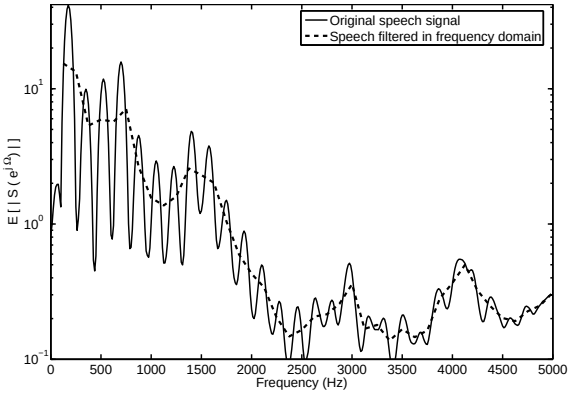

Figure 1: Average modulus of the short-term DFT for one voice record.

thus, if $N_{DFT} \geq L$ then (6) is equal to:

$$S_p(k) = \sum_{n=0}^{L-1} g_p[n] \cdot e^{-j \cdot \frac{2\pi}{N_{DFT}} \cdot kn} \tag{8}$$

The frequency values that correspond to each stDFT coefficient are:

$$f_k = \begin{cases} f_s \cdot \frac{k}{N_{DFT}} & if \quad k \leq \frac{N_{DFT}}{2} \\ f_s \cdot \frac{k-N_{DFT}}{N_{DFT}} & if \quad k > \frac{N_{DFT}}{2} \end{cases} \tag{9}$$

2.2 Short-time Cepstrum

In (Deller et al., 1993), an algorithm for computing the short-time cepstrum from the stDFT is given, under the assumption that $N_{DFT} >> L$:

$$c_p[q] = \frac{1}{N_{DFT}} \cdot \sum_{k=0}^{N_{DFT}-1} \log|S_p(k)| \cdot e^{j \cdot \frac{2\pi k}{N_{DFT}} \cdot q} \tag{10}$$

A physical interpretation of cepstrum can be derived from the discrete-time model for speech production that can also be found in (Deller et al., 1993). This model may be written in frequency domain as:

$$S\left(e^{j\Omega}\right) = E\left(e^{j\Omega}\right) \cdot G\left(e^{j\Omega}\right) \cdot H\left(e^{j\Omega}\right) \tag{11}$$

where $S\left(e^{j\Omega}\right)$ is the speech, $E\left(e^{j\Omega}\right)$ is the impulse train corresponding to the fundamental frequency and its harmonics, $G\left(e^{j\Omega}\right)$ is the glottal pulse waveform that modulates the impulse train and $H\left(e^{j\Omega}\right)$ is, herein, the combined effect of vocal tract and lip radiation. These components can be appreciated in figure 1, which corresponds to the average modulus of the short-term DFT calculated from one of the voice records belonging to the database referred in section 4.1.

The quick impulse-like variations in figure 1 correspond to the pitch harmonics $E\left(e^{j\Omega}\right)$, and the evolution of the impulse amplitude envelope is related

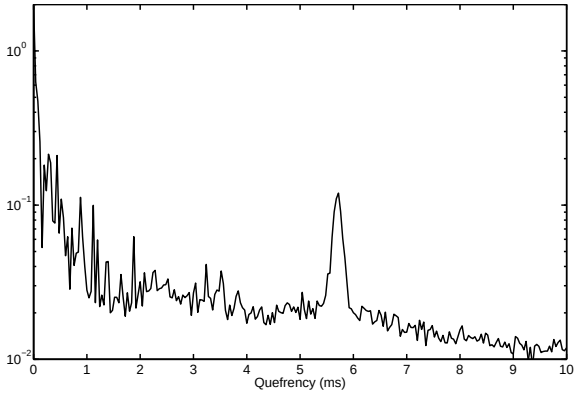

Figure 2: Short term cepstrum averaged for all frames of the same voice record as used for figure 1.

to the glottal waveform $G\left(e^{j\Omega}\right)$ and the formants induced by the vocal tract $H\left(e^{j\Omega}\right)$. These formants correspond to the three envelope peaks with a decreasing level of energy that are centered at 750 Hz, 1375 Hz and 3000 Hz. In fact, these center frequencies are coherent with the range of typical values given in (Jackson-Menaldi, 2002).

The logarithm operation in (10) converts the products in (11) into sums. Consequently, it allows the cepstrum to separate fast from slow signal variations in frequency domain. This widely known fact is illustrated in figure 2, where the peak around 5.7 ms clearly identifies the fundamental frequency (175 Hz) and the values below 2 ms correspond to the spectrum envelope.

2.3 Short-time MFCC

Once the stDFT of a speech signal is available, another option for further processing, as mentioned in section 1, is the calculation of short-time MFCC (stMFCC) parameters. For stMFCC computation, only the positive part of the frequency axis is considered (Rabiner and Juang, 1993), that is, $f_k \geq 0$ and, therefore, $k \leq N_{DFT}/2$. In order to calculate stMFCC coefficients, a transformation is applied to the frequencies so as to convert them to Mel-frequencies f_k^m (Godino-Llorente and Gómez-Vilda, 2004):

$$f_k^m = 2595 \cdot \log_{10}\left(1 + \frac{f_k}{700}\right) \qquad (12)$$

and the stDFT is further processed through band-pass integration along M equally long Mel-frequency intervals, being $M = \lfloor 3 \cdot \log_{10} f_s \rfloor$ ($\lfloor \cdot \rfloor$ means rounding to the previous integer). Namely, the i^{th} interval ($i = 1 \cdots M$) in Mel-domain is defined by:

$$I_i^m = \left[F^m \cdot \frac{i-1}{M+1}, F^m \cdot \frac{i+1}{M+1}\right] \qquad (13)$$

where F^m is the maximum Mel-frecuency:

$$F^m = \max_k f_k^m = 2595 \cdot \log_{10}\left(1 + \frac{f_s/2}{700}\right) \qquad (14)$$

and the interval length in Mel-domain is given by:

$$L\left(I_i^m\right) = \frac{2}{M+1} \cdot F^m \qquad (15)$$

According to previous equations, the N_{DFT} stDFT coefficients are transformed to M frequency components as follows:

$$\widetilde{S}_p(i) = \sum_{f_k \in I_i} \left(1 - \frac{\left| f_k^m - F^m \cdot \frac{i}{M+1} \right|}{L\left(I_i^m\right)/2}\right) \cdot \left|S_p(k)\right| \quad (16)$$

Last, the q^{th} ($q = 1 \cdots Q$) stMFCC of the p^{th} speech frame, where Q is the desired length of the Mel-cepstrum, is given by cosine transform of the logarithm of the smoothed "Mel-spectrum" (Rabiner and Juang, 1993):

$$\widetilde{c}_p[q] = \sum_{i=1}^{M} \log\left|\widetilde{S}_p(i)\right| \cdot \cos\left[q \cdot \left(i - \frac{1}{2}\right) \cdot \frac{\pi}{M}\right] \quad (17)$$

3 CEPSTRAL COEFFICIENTS BASED ON SMOOTHED SPECTRUM

3.1 Justification

As stated in section 1, while MFCC parameters exhibit both good performance and robustness in feature extraction from speech, they lack a clear physical interpretation. On the opposite, cepstrum has a physical meaning (recall section 2.2), yet raw cepstrum coefficients are not as useful for speech parametrisation. In the next paragraphs, the reasons for these facts are exposed.

Cepstrum calculation, as formulated in (10), is based on the spectrum estimate provided by the absolute value of the stDFT. Due to the logarithm, this gives a result that is proportional to the case of periodogram-based spectrum estimation. However, such estimation is very dependent on the specific values of the original speech frame. A more robust spectrum estimate can be obtained by smoothing of the periodogram (Blackman and Tukey method, (Proakis and Manolakis, 1996)). In fact, this is what (16) expresses in the calculation of MFCC. Therefore, filtering of the stDFT may be assumed to be one of the sources of MFCC robustness.

In contrast, an explanation for the lack of clear interpretation of MFCC also lies in the meaning of

(16). According to that equation, stDFT smoothing for MFCC computation is carried out with a variable-length filter, that is, a Bartlett window whose length decreases for lower frequency bands. Moreover, the smoothed stDFT is downsampled to obtain only M samples in the interval $[0, f_s/2]$ that are not uniformly spaced (Rabiner and Juang, 1993). While the downsampling is positive in the sense that it reduces the dimensionality of the problem, its non-uniformness, together with the previous variable-length filtering, obscures the interpretation of the output of the cosine transform in (17).

From the previous reasoning, if stDFT is smoothed with a fixed-length filter and its output is uniformly decimated prior to the logarithm computation, the cepstral coefficients in (10) can be transformed to a more robust parameter set. Moreover, this is achieved while keeping the physical meaning of cepstrum, since the output of the first operation gives an improved spectrum estimate and the second only limits the length of cepstrum in quefrency domain.

3.2 Formulation

Starting from (8), if the stDFT modulus is smoothed with a Bartlett window of constant length equal to Δf then the following output is obtained:

$$S'_p(i) = \sum_{f_k \in I_i} \left(1 - \frac{\left|f_k^m - i \cdot \Delta f/2\right|}{\Delta f/2}\right) \cdot |S_p(k)| \quad (18)$$

where $I_i = [\Delta f \cdot (i-1)/2, \Delta f \cdot (i+1)/2]$ and the Bartlett window has been chosen for similarity with (16). Herein, only the positive part of the frequency axis has been considered, as in section 2.3.

If the filtered stDFT is decimated so as to keep only the outputs of consecutive windows with a 50% overlap, this is equivalent to decimation by a factor $D = \lfloor \Delta f \cdot N_{DFT}/(2 \cdot f_s) \rfloor$. The modified cepstrum then becomes:

$$c'_p[q] = \frac{D}{N_{DFT}} \cdot \sum_{k=0}^{\frac{N_{DFT}}{2 \cdot D}} \log\left|S'_p(k \cdot D)\right| \cdot \cos\left((k-1) \cdot \frac{2\pi D}{N_{DFT}} \cdot q\right) \quad (19)$$

where only the positive frequencies have been considered, hence computing the inverse DFT as a cosine transform as in (17). $c'_p[q]$ has the twofold advantage over $c_p[q]$ of being based on a smoother spectrum estimate $S'_p(i)$ and having a period length that has been reduced by a factor D, thus providing some dimensionality reduction.

3.3 Cepstral distances

Differences in cepstrum can be used for speech signal classification. An example of such usage is the definition of the cepstral distance in (Rabiner and Juang, 1993) as the norm of the vector resulting form substraction of the two cepstra to be compared. This, if directly applied to pathology detection, would result in comparing the cepstrum of consecutive speech frames so as to assess the variability of the signal. Mathematically:

$$d_p^2 = \sum_{q=0}^{\frac{N_{DFT}}{D}-1} \left|c'_{p+1}[q] - c'_p[q]\right|^2 \quad (20)$$

However, bearing in mind the physical interpretation of cepstrum, this definition has the drawback of mixing pitch variations with formant and glottal pulse variations. To overcome this problem an individual frame-to-frame cepstral parameter variation analysis is proposed:

$$d_p[q] = \left|c'_{p+1}[q] - c'_p[q]\right| \quad (21)$$

This way, analysis of the distribution of $d_p[q]$ related to speech formant and glottal pulse variability (low values of q) can be isolated from pitch changes associated to values of q around the pitch period.

4 APPLICATION AND RESULTS

For the purpose of performance analysis, the modified cepstral parameters presented in previous section have been applied to the problem of automatic pathology detection on recorded voice. The results have been compared to those produced by MFCC. Within this section, first the voice database is presented, second the used parameter set is specified, third the classifier is described and, last, the results are shown and commented.

4.1 Database

The voice records used in this investigation are the same as in (Godino-Llorente et al., 2006a). They belong to a database distributed by the company Kay Elemetrics (Kay Elemetrics Corp., 1994). The recorded sounds correspond to sustained phonations (1-3 s long) of the vowel /ah/ from patients with either normal or disordered voice. Such voice disorders belong to a wide variety of organic, neurological, traumatic and psychogenic classes. Sampling rate of speech records has been made uniform for all of them and equal to 25 kHz, while the coding has a resolution

of 16 bits. The subset taken from the database contains 53 normal and 173 pathological speakers which are uniformly distributed in age and gender (Godino-Llorente et al., 2006a).

4.2 Parameter sets

For each speech record, cepstrum-based coefficients, as defined in (19), have been calculated. Namely, a filter length $\Delta f = 200\ Hz$ has been chosen for sfDFT smoothing. As a consequence, a cepstrum length of $(f_s - \Delta f/2)/(\Delta f/2) = 124$ samples results. The choice of Δf is consistent to the approximate length of the low-band filters used for MFCC calculation (recall (16)). At first sight, however, it has the drawback of loosing pitch information of the signal spectrum. This is illustrated in figure 1 where the filtered DFT has been plotted with a dashed line. Nevertheless, such filtered spectrum contains information on both harmonic-to-noise ratio (HNR) and glottal pulse waveform (Murphy and Akande, 2005) and HNR is a useful parameter for pathology detection that is closely related to both frequency and amplitude perturbations of pitch (Jackson-Menaldi, 2002).

Since cepstrum contains information on total signal energy and its distribution among formants, the whole sequence is used as part of the parameter set. As well as the cepstrum, information on its variability is used as an input for the pathology detector. More specifically, the mean and variance of $d_p[q]$ for each value of q are used as descriptors of the cepstrum variability. Therefore, on the whole, a parameter vector of 124×3 elements is produced.

For the sake of comparison, another classifier based on a parameter vector consisting of $M = \lfloor 3 \cdot \log_{10} f_s \rfloor = 13$ MFCC coeficients averaged for all signal frames has also been tested.

4.3 Classifier description

For both classification schemes, a Multilayer Perceptron (MLP) with two hidden layers, each consisting of 4 neurons, and a single-neuron output layer has been used as a classifier. All neurons have logistic activation functions. An MLP with a single hidden layer having 50 neurons was utilised in (Godino-Llorente and Gómez-Vilda, 2004). The structure herein proposed, in contrast, has less free parameters, thus allowing a faster learning, and the reduced number of neurons is compensanted by the introduction of an additional hidden layer that permits learning of more complex relations (Haykin, 1994).

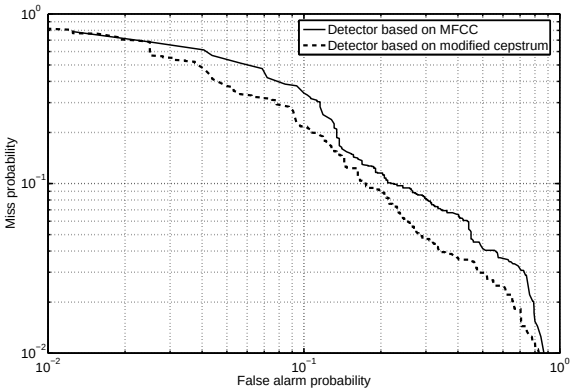

Figure 3: DET plot for MFCC based and modified cepstrum based classifiers.

4.4 Results

The MLP classifier has been trained with 70% of available speech records in such a way that its output is expected to be "1" for pathological voices and "0" for normal voices. The remaining 30% of records have been used for testing. The experiment has been repeated 20 times, each of them with different, randomly chosen, training sets. The average results for both MFCC and herein presented cepstrum-based parameters are drawn in the DET plot (Martin et al., 1997) of figure 3.

Plotted results indicate that the performance of the classifier based on the newly proposed set of parameters is in the same order of magnitude than that of MFCC parameters. To be specific, in terms of equal error rate (EER), that is, for false alarm rate equal to miss rate, the MFCC-based classification yields an experimental error probability of 15 % while the cepstrum-based classification error probability for the same conditions is 14 %. Considering that within this experiment the task of fine-tuning the classifier has not been carried out and that the MLP has been chosen as a standard for comparison, the difference in the results is not significant.

In order to acquire a deeper understanding of the reasons for these results, an analysis of the relevance of cepstrum-based parameters for speech classification as either pathological or not has been realised. Such analysis is based on the evaluation of the Fisher criterion (Duda et al., 2001) for each individual parameter. The results, differentiated for the three subsets of parameters (modified cepstrum, variance of differences and average of absolute differences) are plotted in figure 4.

According to this plot, the most relevant cepstral parameters for pathology detection maybe roughly classified into two groups:

- The modified cepstrum values with lowest indices

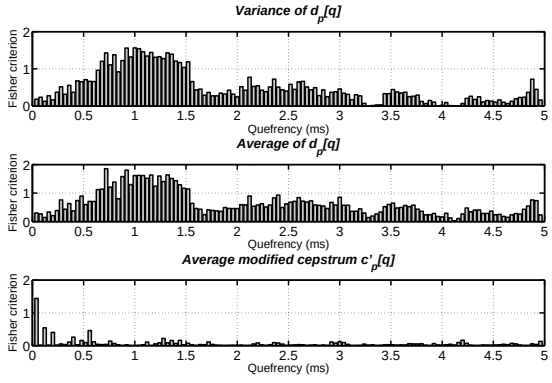

Figure 4: Value of Fisher criterion for each cepstral parameter.

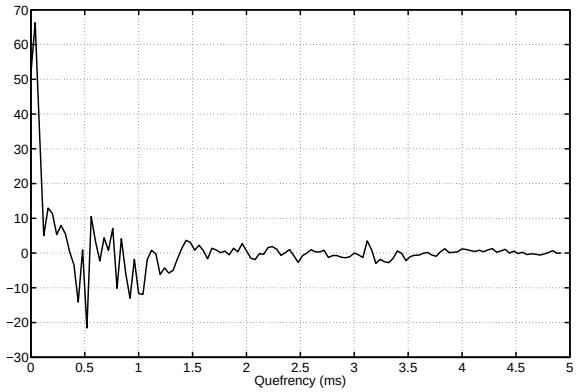

Figure 5: 124 modified cepstral parameters from one of the database's voice records.

(plot at the bottom of figure 4): these are related to the slowest components of the spectrum envelope in figure 1, which, on their side, are associated to spectral noise levels and HNR (Murphy and Akande, 2005).

- The frame-to-frame variations in cepstrum-based coefficients whose quefrecies are within the interval $[0.5, 1.5]$ miliseconds approximately: coefficients within that interval correspond to the short frequency range components of the spectrum envelope. These components, as justified in section 2.2, are related to glottal waveform and speech formants. However, this information itself does not help to discriminate the presence of pathology, as indicated by the low values of the Fisher criterion in the bottom plot of figure 4. Instead, frame-to-frame variations of these factors are much more relevant, as depicted in the other two plots of the same figure.

To be more specific, since the voice records of the database used for this experiment correspond to sustained vowel phonations, it can be assumed that the vocal tract has very little variations, hence formants do not change and the second group of parameters should be more closely related to changes in the glottal waveform. As for the limits of the quefrency interval in which parameters from the second group are relevant, the lower limit of 0.5 ms corresponds to the quefrency band that separates slow components of the spectrum envelope (first group of parameters) from faster components (associated to the second set); on the other hand, the upper limit of 1.5 ms corresponds to the highest quefrency range at which the modified cepstrum $c'_p[q]$ has significant values. This is shown in figure 5, where a plot of the frame-averaged modified cepstrum of one voice record is depicted.

5 Conclusions

Speech parametrisation in cepstral domain is a useful technique for automatic pathology detection. Specifically, MFCC have been successfully used for this purpose. While the computation of these parameters has an intrinsic robustness due to its independency from pitch extraction and the spectrum filtering, their physical interpretation is obscure because of the non-linear Mel-frequency transformation.

Within this paper an alternative set of cepstrum-based parameters has been proposed. Such parameters share the robustness of MFCC since they do not require pitch estimation and filtering of the estimated speech spectrum is also performed. In contrast to MFCC, the calculation of these newly proposed parameters does not involve any non-linear frequency transformation and, consequently, their physical interpretation remains clear. Namely, their values have been shown to be related to the amount of noise energy present in speech and the glottal waveform variability. Both factors are directly associated to laringeal pathologies.

Finally, the performance of the proposed cepstral parameters for pathology detection has been tested using a MLP classifier and results have been compared to those of MFCC. The obtained misclassification rates indicate that the performances of both sets of parameters are similar. Moreover, a deeper analysis on the individual impact of each parameter on the classification task has revealed that the most relevant parameters are those more closely linked to the abovementioned two factors: noise energy and glottal wave variations.

ACKNOWLEDGEMENTS

This research was carried out within projects funded by the Ministry of Science and Technology of Spain (TEC2006-12887-C02) and the Universidad Politécnica de Madrid (AL06-EX-PID-033).

REFERENCES

Bou-Ghazale, S. E. and Hansen, J. H. L. (2000). A comparative study of traditional and newly proposed features for recognition of speech under stress. *IEEE Transactions on Speech and Audio Processing*, 8(4):429–442.

Boyanov, B. and Hadjitodorov, S. (1997). Acoustic analysis of pathological voices. A voice analysis system for the screening of laryngeal diseases. *IEEE Engineering in Medicine and Biology*, 16(4):74–82.

Boyanov, B., Ivanov, T., Hadjitodorov, S., and Chollet, G. (1993). Robust hybrid pitch detector. *IEE Electronics Letters*, 29(22):1924–1926.

Deliyski, D. D. (1993). Acoustic model and evaluation of pathological voice production. In *Proceedings of the 3rd Conference on Speech Communication and Technology (EUROSPEECH'93)*, pages 1969–1972, Berlin (Germany).

Deller, J. R., Proakis, J. G., and Hansen, J. H. L. (1993). *Discrete-time processing of speech signals*. Macmillan Publishing Company, New York (USA).

Duda, R. O., Hart, P. E., and Stork, D. G. (2001). *Pattern classification*. John Wiley & sons, New York (USA), 2nd edition.

Fraile, R., Godino-Llorente, J. I., Sáenz-Lechón, N., Osma-Ruiz, V., and Gómez-Vilda, P. (2007). Analysis of the impact of analogue telephone channel on mfcc parameters for voice pathology detection. In 8th *INTERSPEECH Conference (INTERSPEECH 2007)*, pages 1218–1221, Antwerp (Belgium).

Ganchev, T., Fakotakis, N., and Kokkinakis, G. (2005). Comparative evaluation of various MFCC implementations on the speaker verification task. In *Proceedings of the* 10th *International Conference on Speech and Computer (SPECOM 2005)*, pages 191–194, Patras (Greece).

Godino-Llorente, J. I. and Gómez-Vilda, P. (2004). Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors. *IEEE Transactions on Biomedical Engineering*, 51(2):380–384.

Godino-Llorente, J. I., Gómez-Vilda, P., and Blanco-Velasco, M. (2006a). Dimensionality reduction of a pathological voice quality assessment system based on gaussian mixture models and short-term cepstral parameters. *IEEE Transactions on Biomedical Engineering*, 53(10):1493–1953.

Godino-Llorente, J. I., Sáenz-Lechón, N., Osma-Ruiz, V., Aguilera-Navarro, S., and Gómez-Vilda, P. (2006b). An integrated tool for the diagnosis of voice disorders. *Medical Engineering & Physics*, 28(3):276–289.

Haykin, S. (1994). *Neural Networks: a comprehensive foundation*. Macmillan College Publishing Company, New York (USA), 1st edition.

Jackson-Menaldi, M. C. A. (2002). *La voz patológica*. Editorial Médica Panamericana, Buenos Aires (Argentina).

Kay Elemetrics Corp. (1994). Disordered voice database.version 1.03.

Martin, A., Doddington, G., Kamm, T., Ordowski, M., and Przybocki, M. (1997). The DET curve in assessment of detection task performance. In *Proceedings of the* 5th *Conference on Speech Communication and Technology (EUROSPEECH'97)*, pages 1895–1898, Rhodes (Greece).

Murphy, P. J. and Akande, O. O. (2005). Quantification of glottal and voiced speech harmonics-to-noise ratios using cepstral-based estimation. In *Proceedings of the 3th International Conference on Non-Linear Speech Processing (NOLISP'05)*, pages 224–232, Barcelona (Spain).

Proakis, J. G. and Manolakis, D. G. (1996). *Digital Signal Processing. Principles, Algorithms and Applications*. Prentice-Hall International, New Jersey (USA), 3rd edition.

Rabiner, L. and Juang, B. H. (1993). *Fundamentals of speech recognition*. Prentice-Hall, Englewood Cliffs (USA).