[20] F. Maes, D. Vandermeulen, and P. Suetens, "Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information," *Med. Image Anal.*, vol. 3, pp. 373–86, Dec. 1999.

[21] C. Studholme, D. L. Hill, and D. J. Hawkes, "Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures," *Med. Phys.*, vol. 24, pp. 25–35, Jan. 1997.

[22] J. B. Maintz and M. A. Viergever, "A survey of medical image registration," *Med. Image Anal.*, vol. 2, pp. 1–36, Mar. 1998.

[23] T. S. Yoo, G. D. Stetten, and B. Lorensen, "Medical image registration: Concepts and implementation," in *Insight Into Images: Principles and Practice for Segmentation, Registration, and Image Analysis*, T. S. Yoo, Ed. Wellesley, MA: A K Peters, 2004, pp. 19–45.

[24] Y. J. Zhang, "Quantitative study of 3D gradient operators," *Image Vis. Comput.*, vol. 11, pp. 611–622, Dec. 1993.

[25] T. Hildebrand and P. Rüegsegger, "A new method for the model-independent assessment of thickness in three-dimensional images," *J. Microsc.*, vol. 185, pp. 67–75, Jan. 1997.

[26] B. Hallgrímsson, C. J. Dorval, M. L. Zelditch, and R. Z. German, "Craniofacial variability and morphological integration in mice susceptible to cleft lip and palate," *J. Anat.*, vol. 205, pp. 501–517, Dec. 2004.

[27] F. L. Bookstein, " "Voxel-based morphometry" should not be used with imperfectly registered images," *Neuroimage*, vol. 14, pp. 1454–62, Dec. 2001.

[28] J. Ashburner and K. J. Friston, "Why voxel-based morphometry should be used," *Neuroimage*, vol. 14, pp. 1238–1243, Dec. 2001.

[29] S. N. Khan, K. G. Hufnagle, and R. Pool, "Intrafamilial variability of popliteal pterygium syndrome: A family description," *Cleft Palate J.*, vol. 23, pp. 233–236, Jul. 1986.

[30] S. Kondo, B. C. Schutte, R. J. Richardson, B. C. Bjork, A. S. Knight, Y. Watanabe, E. Howard, R. L. de Lima, S. Daack-Hirsch, A. Sander, D. M. McDonald-McGinn, E. H. Zackai, E. J. Lammer, A. S. Aylsworth, H. H. Ardinger, A. C. Lidral, B. R. Pober, L. Moreno, M. Arcos-Burgos, C. Valencia, C. Houdayer, M. Bahuau, D. Moretti-Ferreira, A. Richieri-Costa, M. J. Dixon, and J. C. Murray, "Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes," *Nat. Genet.*, vol. 32, pp. 285–289, Oct. 2002.

[31] D. Lacombe, J. M. Pedespan, D. Fontan, J. F. Chateil, and A. Verloes, "Phenotypic variability in van der Woude syndrome," *Genet. Couns.*, vol. 6, pp. 221–226, 1995.

[32] T. E. Parsons, E. Kristensen, L. Hornung, V. M. Diewert, S. K. Boyd, R. Z. German, and B. Hallgrímsson, "Phenotypic variability and craniofacial dysmorphology: Increased shape variance in a mouse model for cleft lip," *J. Anat.*, vol. 212, pp. 135–143, Feb. 2008.

Effects of Audio Compression in Automatic Detection of Voice Pathologies

Nicolás Sáenz-Lechón*, Víctor Osma-Ruiz, Juan I. Godino-Llorente, Manuel Blanco-Velasco, Fernando Cruz-Roldán, and Julián D. Arias-Londoño

***Abstract*—This paper investigates the performance of an automatic system for voice pathology detection when the voice samples have been compressed in MP3 format and different binary rates (160, 96, 64, 48, 24, and 8 kb/s). The detectors employ cepstral and noise measurements, along with their derivatives, to characterize the voice signals. The classification is performed using Gaussian mixtures models and support vector machines. The results between the different proposed detectors are compared by means of detector error tradeoff (DET) and receiver operating characteristic (ROC) curves, concluding that there are no significant differences in the performance of the detector when the binary rates of the compressed data are above 64 kb/s. This has useful applications in telemedicine, reducing the storage space of voice recordings or transmitting them over narrow-band communications channels.**

***Index Terms*—Gaussian mixture models, MP3 compression, support vector machines, voice pathology detection.**

I. Introduction

There are several studies in existing literature dealing with the automatic detection of voice disorders [1]–[4], yielding high accuracy rates. These studies are based on the analysis of high-quality voice recordings, characterized by different acoustic parameters, noise measurements, or cepstral coefficients. In order to do this, a set of voices from patients and normal speakers is needed, gathered under controlled conditions, constituting a *voice disorders database*. There are protocols [5] for acquiring the most useful signals for this purpose. These works suggest recording various voice samples from each patient, including the utterances of different vowel sounds, acoustically balanced sentences, and at least 1 min of continuous speech, to allow a good analysis. This requires huge amounts of data involving a considerable storage space (a minute of high quality audio needs about 5 MB) in order to have a statistically representative sample of the population under analysis.

Taken that into account, there is a series of situations in which it is interesting to compress these audio data: first of all, in the daily clinical routine, storing the patients' histories with a minimum of disk space. Besides, the transmission of the database samples over a telecommunications network can sometimes be necessary, in order to perform the analysis in a different place or to share the recordings with other

Manuscript received May 3, 2007; revised January 26, 2008. First published April 22, 2008; current version published December 17, 2008. This work was supported in part by the Ministry of Science and Technology of Spain under Grant TEC2006-12887-C02-02 and in part by the Universidad Politécnica de Madrid, Spain, under Grant AL06-EX-PID-033. *Asterisk indicates corresponding.*

*N. Sáenz-Lechón is with the Department of Ingeniería de Circuitos y Sistemas, Universidad Politécnica de Madrid, 28031 Madrid, Spain (e-mail: nicolas. saenz@upm.es).

V. Osma-Ruiz, and J. I. Godino-Llorente are with the Department of Ingeniería de Circuitos y Sistemas, Universidad Politécnica de Madrid, 28031 Madrid, Spain (e-mail: vosma@ics.upm.es; igodino@ics.upm.es).

M. Blanco-Velasco and F. Cruz-Roldán are with the Department of Teoría de la Señal y Comunicaciones, Universidad de Alcalá, 28871 Madrid, Spain (e-mail: manuel.blanco@uah.es; fernando.cruz@uah.es).

J. D. Arias-Londoño is with the G. de Control y Procesamiento Digital de Señales, Universidad Nacional de Colombia Sede Manizales, Manizales 6460, Colombia (e-mail: judaarlo@gmail.com).

Digital Object Identifier 10.1109/TBME.2008.923769

research groups in a collaborative environment. In the case of rural areas, communications are usually held by conventional telephone lines, with an effective bandwidth of 56 kb/s, which does not allow a fast transmission of high-quality audio files.

The storage problems of the speech databases in the speech recognition field are highlighted in [6], proposing a solution by means of audio compression based on the MP3 standard [7]. This algorithm is very popular for music coding and audio transmission over narrowband channels. MP3 attempts to limit the loss of sound quality, taking into account a series of physical phenomena, eliminating information that is not perceptible by the human auditory system. The experiments showed that binary rates above 32 kb/s have a small influence on the accuracy of the recognition.

Recently, there has been some work on automatic detection of voice pathologies using audio files recorded under nonideal conditions [8]. The authors developed a system with voice registers transmitted over a telephone line. The voice signal was filtered and its bandwidth limited between 300 and 3600 Hz before being transmitted. They employed the typical acoustic features in this context (*jitter, shimmer*, noise parameters, etc.), and the classification was carried out by means of a simple linear discriminant classifier, showing a performance reduction of 14.95% compared to the same system using high-quality voice registers.

In [9], an acoustic analysis of pathological voices compressed in MP3 format is presented, using classical acoustic parameters and showing that dysphonic voices are affected in a different way to normal voices by the compression. However, the authors conclude that registers with binary rates over 96 kb/s present a high fidelity to the original signal and its acoustic properties are not significantly altered, although they clearly state that in the case of severe pathology with seriously damaged harmonic structures, this conclusion may not remain valid.

In this paper, we are interested in studying the degree to which the MP3 format can affect the efficiency of a voice pathology detection task, bearing in mind that some of the compression procedures can have an important effect on the voice features, especially those related with noise measurements that are typically used to discriminate between normal and pathological voices.

II. Materials and Methods

A. Database

The research was carried out with the *Massachusetts Eye and Ear Infirmary Voice Laboratory* database [10]. Due to the different sampling rates of the recordings, a downsampling was performed when required, in order to adjust every utterance to 25 kHz. All the samples are monophonic and stored with 16 bits of resolution, so the files can be considered of mid-quality [in contrast with compact disk (CD) quality recordings].

The registers contain the sustained phonation of vowel /ah/ from patients with a variety of voice pathologies and were edited to remove the initial and final samples. A subset of 173 pathological and 53 normal registers has been taken, according to those enumerated by Parsa *et al.* [2]. The asymmetry in the amount of normal and pathological records has not been considered a problem due to the fact that pathological recordings are approximately 1 s long, whereas normal recordings last around 3 s. For a more detailed discussion of this database, see [11].

Starting from this corpus of files, six other corpora were developed to carry out the experiments (Table I). They were created by compressing the voice recordings with different qualities (160, 96, 64, 48, 24, and 8 kb/s) with constant bit rate, an output sampling rate of 24 kHz and 16 b of resolution, using the Lame codec, version 3.92 [12]. For subsequent processing, the files were decoded back and stored in WAV format (waveform files with no compression).

TABLE I
Corpora Built for the Experiments

Corpus	MP3 coding		Compression ratio	Effective bandwidth
	Bit rate	Sampling frequency		
1	—	—	—	12.5 kHz
2	160 kbps	24 kHz	2.34:1	12 kHz
3	96 kbps	24 kHz	3.91:1	12 kHz
4	64 kbps	24 kHz	5.87:1	12 kHz
5	48 kbps	24 kHz	7.85:1	11 kHz
6	24 kbps	24 kHz	15.47:1	6 kHz
7	8 kbps	24 kHz	46.59:1	2 kHz

B. Parameterization

The analysis is carried out on a short-time basis, so the first step of the process is the segmentation of the voice signals into frames of 40 ms long, using a Hamming window. The length of each window is enough to contain at least two fundamental periods of any phonation in the database. Consecutive windows are overlapped in a 50% of their length.

Then the windows are parameterized by means of *Mel-frequency cepstrum coefficients* (MFCC) [13], a family of parameters that can be estimated using a nonparametric [fast Fourier transform (FFT) based] approach that allows to model the effects of pathology in both the excitation (vocal folds) and the system (vocal tract) [4]. Another reason for using these parameters is because they are also based on a perceptual representation of the frequency corresponding to the human auditory system response [13]. This matches well with the fact that an experienced speech therapist can often detect the presence of a disorder just by listening to it.

A number of MFCC parameters, between 12 and 20, are extracted for every frame, with the goal of achieving the appropriate dimensionality for the task. The MFCC parameters have been complemented with the energy of the frame and three noise measurements that provide an idea of the voice quality: *harmonics to noise ratio* (HNR) [14], *normalized noise energy* (NNE) [15], and *glottal to noise excitation ratio (GNE)* [16]. These parameters are quite sensitive to any signal manipulation that could increase the noise level, so the inclusion of these features is justified on the basis that MP3 compression has consequences in the fine-grain structure of the sound wave introducing some kind of interharmonic noise and some loss of fidelity [9].

The feature vectors are formed by concatenating the MFCC along with the energy, the noise features, and the first temporal derivative of all of them. In principle, the derivatives provide important information about the dynamic behavior of the temporal sequence of each feature [4]. Every parameter is normalized into the [0, 1] interval before feeding the detector. In the notation followed in this paper to represent the feature vectors, N represents the noise measurements, E is the energy of the frame, L is the number of MFCC features, and $\triangle$ is the set of first derivatives. By way of example, a family of 12 MFCC coefficients is represented with $NE\_MFFC_{12}\triangle$.

C. Classification

The set of pathological and normal feature vectors is used to adjust the parameters of a *Gaussian mixture model* (GMM) for each class and to train a *support vector machine* (SVM). These detectors were chosen

TABLE II
PERFORMANCE OF THE DIFFERENT CORPORA USING GMM AND SVM

Corpus	Rate	Features	Frame basis		File basis	
			Efficiency ± σ	AUC ± SE	Efficiency ± σ	AUC ± SE
1	—	NE_MFCC$_{14}\Delta$; GMM M=3	91.74% ± 5.1	0.96 ± 0.02	95.04% ± 4.3	0.98 ± 0.01
		NE_MFCC$_{12}\Delta$; SVM C=10^{3}, $\gamma=10^{-4}$	93.52% ± 3.9	0.97 ± 0.01	95.04% ± 3.8	0.98 ± 0.01
2	160 kbps	NE_MFCC$_{12}\Delta$; GMM M=6	90.85% ± 2.8	0.97 ± 0.01	94.35% ± 1.7	0.98 ± 0.01
		NE_MFCC$_{16}\Delta$; SVM C=10^{4}, $\gamma=10^{-2}$	93.35% ± 4.8	0.97 ± 0.01	94.90% ± 3.6	0.98 ± 0.01
3	96 kbps	NE_MFCC$_{12}\Delta$;GMM M=6	94.35% ± 1.7	0.98 ± 0.01	93.25% ± 2.2	0.98 ± 0.01
		NE_MFCC$_{16}\Delta$; SVM C=10^{4}, $\gamma=10^{-2}$	93.01% ± 3.6	0.97 ± 0.01	94.60% ± 3.9	0.98 ± 0.01
4	64 kbps	NE_MFCC$_{14}\Delta$; GMM M=6	89.68% ± 3.2	0.96 ± 0.02	92.98% ± 2.4	0.97 ± 0.02
		NE_MFCC$_{14}\Delta$; SVM C=10^{4}, $\gamma=10^{-2}$	90.75% ± 3.8	0.96 ± 0.01	93.52% ± 4.3	0.97 ± 0.01
5	48 kbps	NE_MFCC$_{14}\Delta$; GMM M=6	88.41% ± 3.2	0.95 ± 0.01	91.32% ± 2.5	0.97 ± 0.01
		NE_MFCC$_{16}\Delta$; SVM C=10^{4}, $\gamma=10^{-2}$	89.34% ± 3.8	0.96 ± 0.01	91.46% ± 4.1	0.97 ± 0.01
6	24 kbps	NE_MFCC$_{12}\Delta$; GMM M=8	87.89% ± 2.9	0.92 ± 0.01	89.67% ± 2.0	0.97 ± 0.01
		NE_MFCC$_{16}\Delta$; SVM C=10^{4}, $\gamma=10^{-4}$	85.42% ± 3.9	92 ± 0.02	86.09% ± 4.9	0.93 ± 0.02
7	8 kbps	NE_MFCC$_{12}\Delta$; GMM M=8	84.36% ± 3.1	0.91 ± 0.02	85.67% ± 1.43	0.94 ± 0.01
		NE_MFCC$_{12}\Delta$; SVM C=10^{4}, $\gamma=10^{-4}$	85.37% ± 4.3	0.91 ± 0.03	87.05% ± 4.3	0.95 ± 0.02

The efficiency is calculated at the Minimum Cost threshold.

on the basis of the modelling capabilities they present. The non-linear mapping carried out by the SVMs maximizes the generalization capabilities of the classifier [17]. On the other hand, the GMMs fit the distribution of the observed data by means of a set of weighted Gaussian functions. The advantages of using a GMM are that it is computationally inexpensive, with the robustness and smoothness of the Gaussian parametric model, and yet it is capable of modeling complex statistical distributions [18]. The modeling ability of GMM and SVM for detection of voice pathologies has already been demonstrated in earlier works [4], displaying a superior performance to other techniques.

The training of the SVM involves adjusting the parameter of the kernel γ and the penalty parameter C. The aim is to identify the best (γ, C) pairs using a subset of the voice registers (the *training* set), so the classifier can accurately predict unknown data (the *test* set). The output value given by the SVM for an input feature vector can be interpreted as the likelihood that the vector belongs to a specific class (normal and pathologic). The logarithm of this likelihood is calculated for every frame or feature vector and is called *score* henceforth.

For the GMM, the number of Gaussian components M, the weights, the means, and the covariance matrices were estimated for each of the target classes (normal and pathological voice) using a training set. Once a GMM is adjusted, it can produce an estimation of the *a posteriori probability* that a given test feature vector would have been drawn from the model. For every input feature vector, the probabilities produced by the two models are divided, yielding a *likelihood ratio* or, in the logarithmic domain, a *score*.

The scores given by the classifiers for pathological and normal voices are used to plot the true and false score curves, respectively. The decision about the presence or absence of pathology is taken, establishing a threshold T in a point called *minimum cost point* (MCP) that corresponds to the minimum average error rate. Once T is chosen, the frames with scores greater or equal to T are assigned to the pathological class, whereas the samples with scores lower than T are labeled as normal.

In order to compare the results between the classifiers, the values of the scores given by the detectors are normalized into the [0, 1] interval according to [19]. This normalization allows us to consider the normalized scores as *a posteriori* estimations of the probability that belongs to each class.

The final score assigned to each record is calculated by averaging in time the total number of frames of the record. Then, two different accuracies can be calculated: file (number of files well classified) and frame accuracy (number of frames well classified). The file accuracy is expected to be equal or better than the frame accuracy.

D. Evaluation

In order to allow comparisons, the methodology proposed in [11] has been used for the evaluation of the system. According to this methodology, the generalization abilities of the system have to be tested, following a cross-validation scheme, with different sets for training and validation (*k-folds*). The results are presented by means of frame and file accuracies and two curves plotted using the scores given by each classifier that show the performance of the proposed architecture: the *detector error tradeoff* (DET) [20], and the *receiver operating characteristic* (ROC) [21]. The *area under the* ROC *Curve* (AUC) and its standard error (SE) are also interesting estimators of the performance.

III. EXPERIMENTS AND RESULTS

The database has been parameterized for each corpus changing the number of MFCC parameters, ranging from 12 to 20. The experiments consisted of searching for the most appropriate values of the SVM and GMM parameters (γ and C for the SVM, and M for the GMM) to achieve the best possible accuracies for each corpus. With respect to the SVM, the optimum working point has been searched for inside the grid $C = [10^3, 10^6]$ and $\gamma = [10^{-4}, 10^{-2}]$. Regarding the GMM, the parameter M has been evaluated into $M = [1, 8]$. The evaluation of the system has been carried out by means of a cross-validation strategy with 11 folds. Table II presents a summary of the best results achieved for each corpus. Regarding the number of MFCC parameters, the best results for every corpus were found in the interval from 12 to 16 parameters. Within this range, small but, due to the confidence intervals, not significant variations in the performance of the system have been found.

For the GMM model and regarding the number of mixtures, the systems trained with compressed recordings needed a larger number of Gaussian components to obtain the best results. The explanation for this fact could lie in the lower resolution of the feature vectors produced by the compression, causing the appearance of isolated clusters; so more Gaussians would be needed to model the more complex feature space. A similar behavior appeared for the SVM model, where the parameter

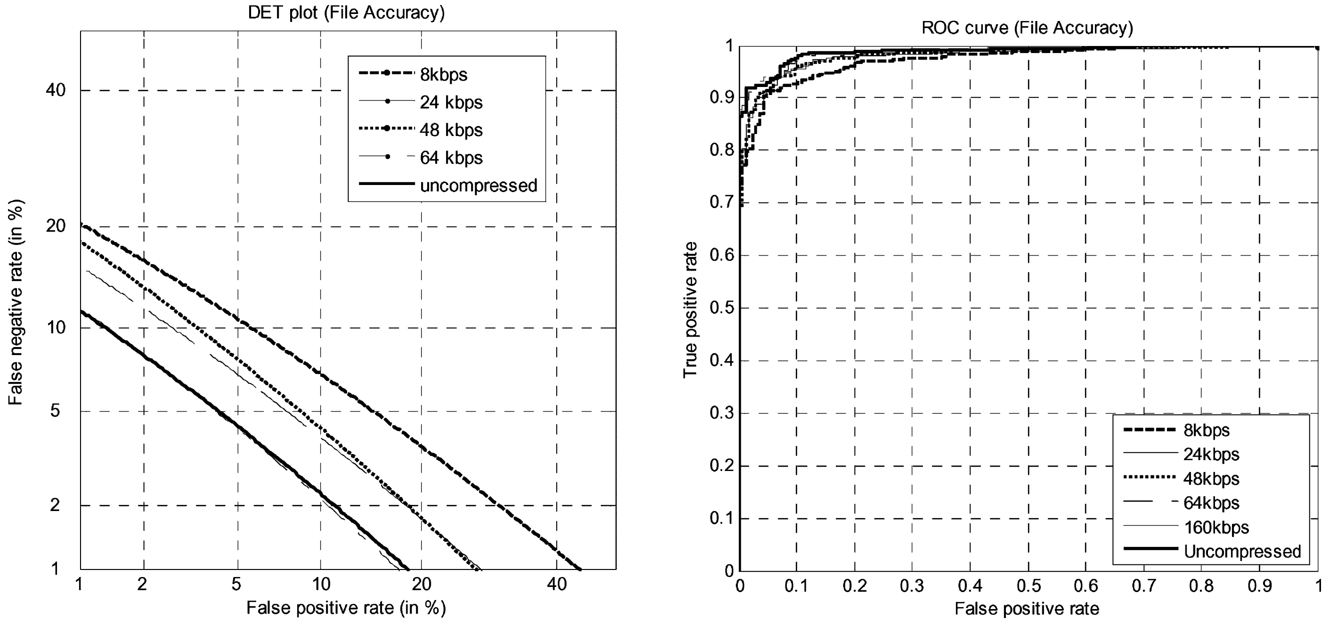

Fig. 1. DET and ROC plots of the file accuracy for the different corpora using a GMM-based detector. The DET plot has been interpolated with a quadratic function.

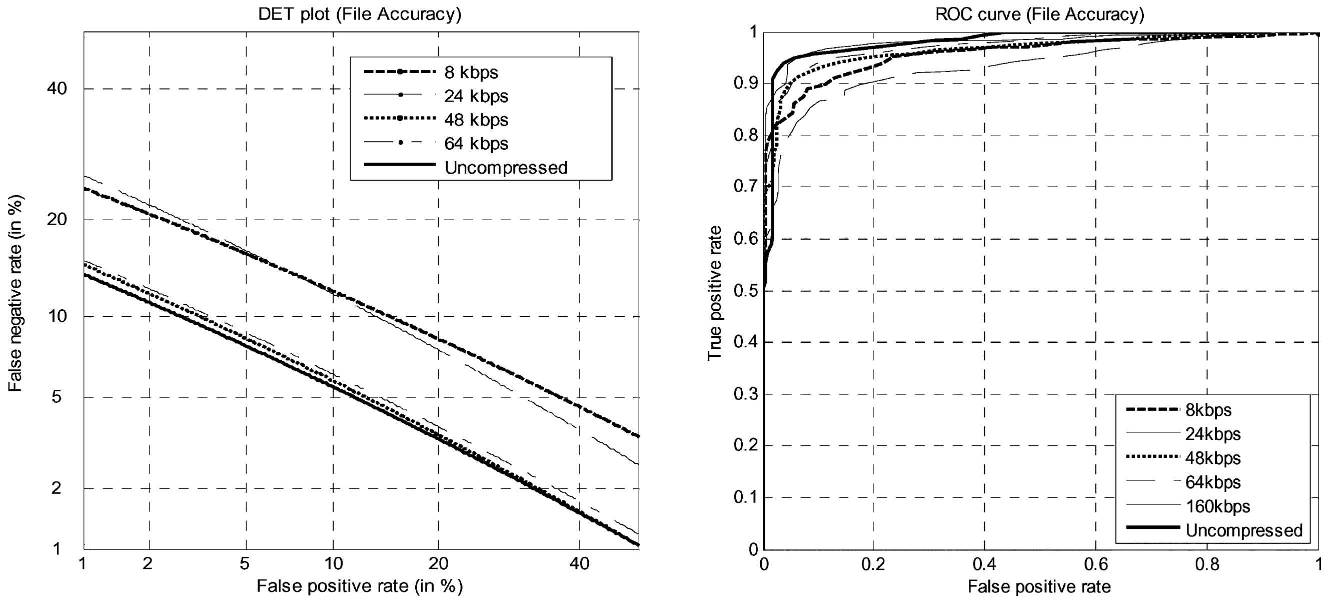

Fig. 2. DET and ROC plots of the file accuracy for the different corpora using a SVM-based detector. The DET plot has been interpolated with a quadratic function.

γ (that represents the spreading of the kernel) is smaller for the lowest binary rates.

In view of Table II, we may say that the performance of the system that uses the voices compressed with 160 and 96 kb/s is very similar to that of the uncompressed registers. In fact, the area under the ROC curve of the former is even higher than for the uncompressed database. This implies that with these compression rates, the results of the detection system are not degraded even though the storage space has been reduced.

Even considering the files compressed with 64 kb/s, it is possible to observe that the loss of information with respect to the original files is not too relevant. In the remaining cases, there is a bigger information loss due to, among other reasons, the fact that the bandwidth of the signals has been considerably reduced. For the case of 8 kb/s, the performance of the detector is reduced by almost 10% due to the effect of the MP3 encoding and the important reduction of the signal bandwidth (2 kHz). Despite the encoding and the significant reduction of the bandwidth, there is not an abrupt drop in the performance because the signals still maintain the most important part of the harmonic structure, and this is the part of the spectrum where most of the frequency and amplitude perturbation is encoded.

Figs. 1 and 2, respectively, show the DET and ROC curves for some of the corpora using the GMM and SVM detectors presented in Table II. The curves corresponding to the uncompressed files and those with 64, 96, and 160 kb/s are very similar, either for an SVM or for a GMM detector, so the latter has been intentionally removed for the sake of clarity. The degradation of the performance can also be seen as the compression ratio increases.

IV. Conclusion

In the last years, an increasing interest to develop pathological voice databases for research purposes has emerged. These databases are the first step for developing automatic detectors of pathologies, voice teletherapy systems, evaluation of voice quality, training experts in acoustic analysis, and so on. The volume of the recordings is considerable by now, representing a problem for storing and specially for transmitting them in voice tele-health applications over narrow-band channels. The compression of the audio samples is a possible solution for these problems, and we think that it can also be useful in different situations, including the ones mentioned before.

For this reason, our goal was to study if MP3 audio compression was a possibility for the detection/evaluation of voice pathologies. The results of this study suggest that it is possible to reduce the size of the recordings about four times (or even more if the original files had more quality than those used in this paper), without compromising the validity of the conclusions and the quality of the voice registers.

Binary rates above 160 kb/s were not taken into account because their quality was considered enough for the detection of voice disorders without loss of efficiency. In view of the results, this assumption was confirmed: a good tradeoff is to use a binary rate of 96 kb/s because it gives similar results to those of the corpus with no compression.

Furthermore, the performance is reduced as the binary rate is decreased: the accuracy of the system working with the original uncompressed WAV files is similar to that using the MP3 compression with 160 or 96 kb/s. Below 96 kb/s, the fidelity of the speech samples is decreased, introducing more important alterations in the voice signal that affect the efficiency of the detector. Working at low bit rates, if the encoder runs out of bits, it will not encode some bands with the required fidelity, which will have consequences in the fine-grain structure of the sound wave.

We can conclude that there are no clear differences between the first three corpora (uncompressed, 160 kbps, and 96 kbps), either with a GMM or with an SVM-based detector. This is also supported by the fact that the harmonic structure of the recordings remains almost untouched in the low-frequency bands when the binary rate is 96 kb/s or higher.

Consequently, it is possible to consider the use of registers with compression rates over 96 kb/s. On the other hand, the discriminative capabilities of the detectors are seriously affected for binary rates below 64 kb/s. These tests open the possibility of using MP3 compression at high binary rates to store and/or transmit the speech recordings used for the automatic detection of voice disorders, with no alteration of the efficiency of the system.

Regarding the parameterization approach, the MFCC parameters can be considered to be adequate for our purposes because they are based on perceptual grounds similar to that of MP3 standard. The MP3 coding transforms the energy in bands in a similar way to the MFCC parameters. The loss of information due to the compression is not meaningful for the detection of pathology. The performance of the compressed corpora is lower in most of the cases than for the original files, but conversely, they need far less free space on disk for their storage and can be transmitted easily through low-speed networks.

The use of a commercial database for evaluating the results ensures that the conclusions presented can be reproduced by other authors. Nevertheless, the results presented here should be validated using another database of voice disorders in order to better generalize them.

References

[1] L. Gavidia-Ceballos and J. H. L. Hansen, "Direct speech feature estimation using an iterative EM algorithm for vocal fold pathology detection," *IEEE Trans. Biomed. Eng.*, vol. 43, no. 4, pp. 373–383, Apr. 1996.

[2] V. Parsa and D. G. Jamieson, "Identification of pathological voices using glottal noise measures," *J. Speech Lang. Hear. Res.*, vol. 43, no. 2, pp. 469–485, 2000.

[3] S. Hadjitodorov and P. Mitev, "A computer system for acoustic analysis of pathological voices and laryngeal disease screening," *Med. Eng. Phys.*, vol. 24, no. 6, pp. 419–429, 2002.

[4] J. I. Godino-Llorente, P. Gómez-Vilda, and M. Blanco-Velasco, "Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 10, pp. 1943–1953, Oct. 2006.

[5] P. H. Dejonckere, P. Bradley, P. Clemente, G. Cornut, L. Crevier-Buchman, G. Friedrich, P. H. Van de Heyning, M. Remacle, and V. Woisard, "A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques," *Eur. Arch. Oto-Rhino-Laryngol.*, vol. 258, no. 2, pp. 77–82, 2001.

[6] P. Sirum and I. Sanches, "The influence of audio compression on speech recognition systems," in *Proc. SPECOM'04*, St. Petersburg, Russia, pp. 128–131.

[7] *ISO-MPEG Audio Layer-3. Information technology—Coding of moving pictures and associated audio for digital storage media up to 1.5 Mbit/s. Part 3: Audio*, ISO/IEC Standard 11172-3, 1993.

[8] R. J. Moran, R. B. Reilly, P. deChazal, and P. D. Lacy, "Telephony-based voice pathology assessment using automated speech analysis," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 3, pp. 468–477, Mar. 2006.

[9] J. González, T. Cervera, and M. J. Llau, "Acoustic analysis of pathological voices compressed with MPEG system," *J. Voice*, vol. 17, no. 2, pp. 126–139, 2003.

[10] Massachusetts Eye and Ear Infirmary, *Voice Disorders Database, Version 1.03, [CD-ROM]*. Lincoln Park, NJ: Kay Elemetrics Corp., 1994.

[11] N. Sáenz-Lechón, J. I. Godino-Llorente, V. J. Osma-Ruiz, and P. Gómez-Vilda, "Methodological issues in the development of automatic systems for voice pathology detection," *Biomed. Signal Process. Control*, vol. 1, no. 2, pp. 120–128, 2006.

[12] Lame MP3 Encoder. (2007). *The Lame Project* [Online]. Available: http://lame.sourceforge.net

[13] J. R. Deller, J. G. Proakis, and J. H. L. Hansen, *Discrete-Time Processing of Speech Signals*. New York: Macmillan, 1993.

[14] G. de Krom, "A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals," *J. Speech Hear. Res.*, vol. 36, no. 2, pp. 254–266, 1993.

[15] H. Kasuya, S. Ogawa, K. Mashima, and S. Ebihara, "Normalized noise energy as an acoustic measure to evaluate pathologic voice," *J. Acoust. Soc. Amer.*, vol. 80, no. 5, pp. 1329–1334, 1986.

[16] D. Michaelis, T. Gramss, and H. W. Strube, "Glottal-to-noise excitation ratio—A new measure for describing pathological voices," *Acustica/Acta Acustica*, vol. 83, pp. 700–706, 1997.

[17] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2000.

[18] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digit. Signal Process.*, vol. 10, pp. 19–41, 2000.

[19] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.

[20] A. F. Martin, G. R. Doddington, T. Kamm, M. Ordowski, and M. A. Przybocki, "The DET curve in assessment of detection task performance," in *Proc. Eurospeech 1997*, Rhodes, crete, vol. 4, pp. 1895–1898.

[21] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, pp. 29–36, 1982.