Image Tampering Detection Using Methods Based on JPEG Compression Artifacts: A Real-Life Experiment

Babak Mahdian*
Institute of Information Theory and Automation of the ASCR
Pod Vodarenskou vezi 4
Prague, Czech Republic
mahdian@utia.cas.cz

Stanislav Saic†
Institute of Information Theory and Automation of the ASCR
Pod Vodarenskou vezi 4
Prague, Czech Republic
ssaic@utia.cas.cz

ABSTRACT

In this paper we analyze the synergy between forensic image head data consistency analysis and detection of doubles JPEG compression artifacts. We show that image head consistency testing is an effective method for detecting digital images that have been modified. On the other hand, when it is not combined with other forensic methods such as double JPEG detection, a high number of altered photos remained undetected. The same can also be claimed about double JPEG detection. When employed separately without conjunction with other methods, the majority of altered photos remained undetected. In this paper, a quantitative study on this topic is carried out. We show that combining various image forensic methods is a must.

Categories and Subject Descriptors

V.I.VI. [**Information Interfaces and Representation (HCI)**]: Multimedia Information Systems —*Image/video retrieval*

Keywords

Digital forensics, jpeg forensics , double jpeg, tampering detection, EXIF analysis, image retrieval

1. INTRODUCTION

Verifying the integrity of digital images and detecting the traces of tampering without using any protecting pre–extracted or pre–embedded information have become an important and hot research field of image processing. The growing number of published papers in this field has been shown, for example, in [1].

*Babak Mahdian
†Stanislav Saic

ISABEL '11 October 26-29, Barcelona, Spain

Generally, approached for image genuineness verification can be divided into active and passive–blind approaches. The area of active methods simply can be divided into the data hiding approach [2, 3] and the digital signature approach [4, 5].

By data hiding we refer to methods embedding secondary data into the image. The most popular group of this area belongs to digital watermarks [6, 7]. The digital signature approach consists mainly of extracting unique features from the image at the source side and encoding these features to form digital signatures. Afterwards signatures are used to verify the image integrity.

In this paper, we focus on blind methods. In contrast to active methods, blind methods do not need any prior information about the image. They mostly are based on the fact that forgeries can bring into the image specific detectable changes (e.g., statistical changes). Specifically, we focus on JPEG files and compression artifacts which are brought into the image by JPEG lossy compression.

When altering a JPEG image, typically it is loaded into a photo–editing software (decompressed) and after manipulations are carried out, the image is re–saved (compressed again). The quantization matrix of the unaltered image is called as primary quantization matrix. The quantization matrix of the re–saved image is called as secondary quantization matrix. If the primary and secondary quantization matrix are not identical, then the re–saving (double compressing) operation can bring into the image specific quantization based artifacts. Detecting these artifacts plays a valuable role in identifying image forgeries. Detecting the traces of double compression also is helpful in other research fields such as steganography [8].

In this paper, we will employ a typical method for detecting double compression artifacts [9] and confront the method with real-life photos taken by real Internet users. We will download photos created by thousands of individual users and without any further knowledge about photos history, we apply the method introduced in [9] to these photos.

There are many methods which are based on downloading thousands or millions of image data from various places on Internet. Many of such methods prefer or must employ only original and non-altered image data. In order to collect original data, typically, they use EXIF information to search for traces of software manipulation. If such traces are found, the image is omitted and not employed further. Taking into account the capability of many software packages to modify the image data without leaving any noticeable traces of modification in EXIF, methods might end up with uncorrect calculations. This can be observed in image retrieval when training classifiers or in image forensics when gathering fingerprints reference data.

It can be interesting to see if double jpeg detection methods can be of value in such areas and if they can help to gather better noise–free and original image data when it is needed. This is the main objective of this paper. Furthermore, we would like to gain an estimate of the rate of altered photos detected by double JPEG algorithms and denoted as original by EXIF based methods and vice versa.

It is important to note that detecting the traces of double compression does not necessarily imply the existence of malicious modifications in the image. Often images are re–compressed due to reduce the image storage size or transmission time. Furthermore, the image could undergo only simple image adjustment operations such as contrast enhancing.

2. RELATED WORK

Yi L. Chen and Chiou T. Hsu [10] proposed a quantization noise model to characterize single and doubly compressed images. In [11], Zhigang Fan and Ricardo Queiroz proposed a method determining whether an image has been previously JPEG compressed. If so, compression parameters are estimated. Specifically, a method for the maximum likelihood estimation of JPEG quantization steps was developed. In [12] Hany Farid proposed a method for detecting composites created by JPEG images of different qualities. The method detects whether a part of an image was initially compressed at a lower quality than the rest of the image. Xiaoying Feng and Gwenael Doerr [13] detect double JPEG images by using periodic artifacts of re–quantization and discontinuities in the signal histogram. Jan Lukáš and Jessica Fridrich [14] presented a method for estimation of primary quantization matrix from a double compressed JPEG image. The paper presents three different approaches from which the Neural Network classifier based one is the most effective. Tomáš Pevný and Jessica Fridrich [8] proposed a method based on support vector machine classifiers with feature vectors formed by histograms of low–frequency DCT coefficients. Dongdong Fu et al. [15] proposed a statistical model based on Benford's law for the probability distributions of the first digits of the block–DCT and quantized JPEG coefficients. Weiqi Luo et al. [16] proposed a method for detecting recompressed image blocks based on JPEG blocking artifact characteristics. Babak Mahdian and Stanislav Saic [9] proposed the a method for detection double compressed JPEG images based on histograms properties of DCT coefficients and support vector machines. Ramesh Neelamani et al. [17] proposed a method to estimate the JPEG compression history. Alin C. Popescu [18] proposed a double JPEG Compression technique by examining the histograms of the DCT coefficients. In [19], Zhenhua Qu et al. formulated the shifted double JPEG compression as a noisy convolutive mixing model to identify whether a given JPEG image has been compressed twice with inconsistent block segmentation.

3. BASICS OF JPEG COMPRESSION

Typically, the image is first converted from RGB to YCbCr, consisting of one luminance component (Y), and two chrominance components (Cb and Cr). Mostly, the resolution of the chroma components are reduced, usually by a factor of two. Then, each component is split into adjacent blocks of 8×8 pixels. Blocks values are shifted from unsigned to signed integers. Each block of each of the Y, Cb, and Cr components undergoes a discrete cosine transform (DCT). Let $f(x, y)$ denotes a 8×8 block. Its DCT is:

$$F(u,v) = \frac{1}{4}C(u)C(v) \sum_{x=0}^{7}\sum_{y=0}^{7} f(x,y) cos\frac{(2x+1)u\pi}{16} cos\frac{(2y+1)v\pi}{16}, \quad (1)$$

where

$$\begin{aligned} &(u, v \in \{0 \cdots 7\}); \\ &C(u), C(v) = 1/\sqrt{2} \quad \text{for} \quad u, v = 0; \\ &C(u), C(v) = 1 \qquad \text{otherwise.} \end{aligned} \quad (2)$$

In the next step, all 64 $F(u, v)$ coefficients are quantized. Then, the resulting data for all blocks is entropy compressed typically using a variant of Huffman encoding.

The quantization step is performed in conjunction with a 64–element quantization matrix, $Q(u, v)$. Quantization is a many–to–one mapping. Thus it is a lossy operation. Quantization is defined as division of each DCT coefficient by its corresponding quantizer step size defined in the quantization matrix, followed by rounding to the nearest integer:

$$F^Q(u,v) = round(\frac{F(u,v)}{Q(u,v)}), \quad u, v \in \{0 \cdots 7\} \quad (3)$$

Generally, the JPEG quantization matrix is designed by taking the visual response to luminance variations into account, as a small variation in intensity is more visible in low spatial frequency regions high spatial frequency regions.

The JPEG decompression works in the opposite order: entropy decoding followed by dequantization step and inverse discrete cosine transform.

4. DOUBLE JPEG QUANTIZATION AND ITS EFFECT ON DCT COEFFICIENTS

By double JPEG compression we understand the repeated compression of the image with different quantization matrices Q_α (primary quantization matrix) and Q_β (secondary quantization matrix). The DCT coefficient $F(u, v)$ is said to be double quantized if $Q_\alpha(u, v) \neq Q_\beta(u, v)$. The double quantization is given by:

$$F^{Q^\beta}(u,v) = round(\frac{F^{Q^\alpha}(u,v)Q^\alpha(u,v)}{Q^\beta(u,v)}) \quad (4)$$

To illustrate the effect of double quantization, consider a set of random values in the range of $\langle -50, 50\rangle$ drawn from a normal zero–mean distribution (see Figure 1(a)). Figure 1 (b) shows the distribution after being quantized with quantization step $Q^\alpha = 3$. Figure 1(c) shows the same distribution after being double quantized with quantization steps $Q^\alpha = 3$ and $Q^\beta = 3$. In other words, Figure 1 (c) was generated by quantization of the distribution by quantization step $Q^\alpha = 3$. Then obtained values were de–quantized using $Q^\alpha = 3$ (so, now each value of the distribution is a multiple of the 3). In the end, values were quantized again using the quantization step $Q^\alpha = 2$. Apparently, the distribution of the doubly quantized values contains periodic empty bins.

This is caused by the fact that in the second quantization values of the distribution are re–distributed into more bins than in the first quantization.

Generally, the double quantization process brings detectable artifacts like periodic zeros and double peaks. The double quantization effect has been studied in detail, for example, [14].

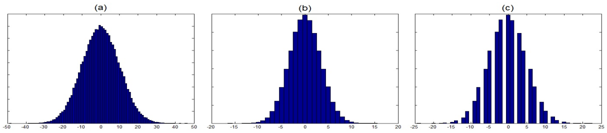

Figure 1: Shown are: (a) the histogram of a non–quantized random values drawn from a zero–mean normal distribution; (b) histogram of the quantized (a) with quantization step $Q^{\alpha} = 3$; (c) histogram of the double–quantized (a) with quantization steps $Q^{\alpha} = 3$ and $Q^{\beta} = 2$.

5. DETECTING DOUBLE JPEG QUANTIZATION ARTIFACTS

Last section briefly described the effect of double quantization. As afore mentioned, in this paper, we employ the method introduced in [9]. The method uses the fact that the histograms of DCT coefficients of a double compressed image contain specific periodic artifacts detectable in the frequency space.

When having a double compressed JPEG image, typically the output of the method applied to the DC component contains a specific clear peak (for example, see Figure 2 (c)). Otherwise, there is no strong peak in the spectrum (Figure 2 (b)). When the method is applied to a singe-quantized AC component, the spectrum has a decaying trend (Figure 2 (e)). Otherwise, in some parts, the spectrum has a local ascending trend (Figure 2 (g)).

As pointed out in [9], the method computes the magnitudes of FFT of the histograms of the DCT coefficients corresponding to low frequencies. Specifically, the following DCT frequencies are employed: $(0, 0)$, $(1, 0)$, $(2, 0)$, $(3, 0)$, $(0, 1)$, $(1, 1)$, $(2, 1)$, $(0, 2)$, $(1, 2)$ and $(0, 3)$. Because of the problem with insufficient statistics for high–frequency DCT coefficients (high frequency DCT coefficients are often quantized to zeros), other frequencies are not considered. Only the first half of the spectrum is considered. We denote the result of this part by $|H_1| \cdots |H_{10}|$, where $|H_1|$ corresponds to DC component and $|H_i|, i = 2 \cdots 10$, correspond to AC components. $|H_i|$ are normalized to have a unit length.

Moreover, before computing the FFT, the margin parts of the histograms are eliminated and not employed for further analysis. The reason is that often images being poor in terms of colors (like scanned documents) and histograms not possible to well approximate by a Gaussian or Laplacian have often a non-typical behavior in the margin parts of leading to false positives.

Only the luminance channel is employed to detect the double JPEG compression artifacts. The reason is that the two chrominance channels of a JPEG compressed image are typically down–sampled by a factor of two or four, and quantized using larger quantization steps. Thus, the histograms

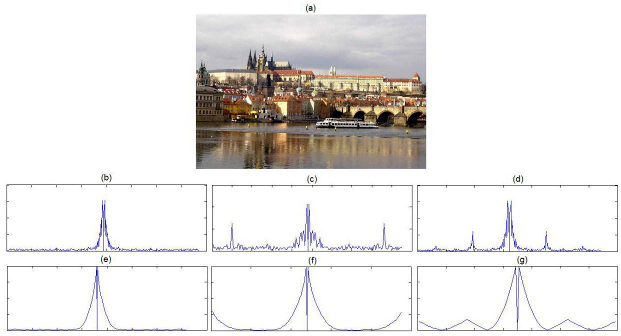

Figure 2: In (a) the test image is shown. (b) and (e) show the magnitudes of Fourier transform of the zero–mean histograms of DCT coefficients corresponding to frequencies $(0, 0)$ and $(1, 1)$ obtained from a single compressed version of (a). Here the image was saved by quality factor 85. (c) and (f) show the same for the double compressed version of (a). Here the image was saved by quality factor 85 followed by quality factor 75. (d) and (g) show the same for double compressed version of (a) with quality factor 85 followed by quality factor 80.

obtaining from these components contain only little information valuable for detecting the presence of double compression.

As mentioned previously, typically, $|H_2| \cdots |H_{10}|$ have a decaying trend. To be able to effectively compare and analyze different histograms, this trend is removed. Method in [9] employed a simple local minimum subtraction operation resulting in removing the decaying trend and preserving local peaks. First, $|H_i|, i = 2 \cdots 10$, are de–noised using an averaging filter. Then, from each frequency f of $|H_i|(f), i = 2 \cdots 10$, the minimum value of its neighbor frequencies is subtracted. Only the neighbor frequencies in direction to the DC component are considered. More formally,

$$|\widetilde{H}_i|(f) = |H_i|(f) - M_i(f), \tag{5}$$

where $M_i(f)$ is the minimum value of $\{|H_i|(f) \cdots |H_i|(f - n)\}$, where $n \in \mathcal{N}_0$ denotes the length of the minimum filter.

Since the histograms of DCT coefficients undergone quantization with a quantization step $Q_1(u, v)$ differ from histograms of DCT coefficients undergone quantization with a step $Q_2(u, v)$ (where $Q_1(u, v) \neq Q_2(u, v)$), the size of the minimum filter n has a different value for different quantization steps. The value n is determined in the training process regarding to the desired detection accuracy and false positives rates.

Before going on, it is important to note that not all combinations of $Q_{\alpha}(u, v)$ and $Q_{\beta}(u, v)$ brings into the DCT histograms double quantization artifacts. If $\frac{Q_{\beta}(u,v)}{Q_{\alpha}(u,v)}$ is an integer value, the specific double quantization artifacts are not introduced into the histograms of DCT coefficients corresponding to frequency (u, v).

So far, we got $|\widetilde{H}_i|$, where $i = 1 \cdots 10$. The quantization step $Q(u, v)$ corresponding to $|\widetilde{H}_i|$ can be determined directly from the quantization table in the JPEG file header. We use this fact and construct one separate classifier for each quantization step of interest The classifier distinguishes between two classes: single compressed and double compressed $|\widetilde{H}|$. When classifying $|\widetilde{H}_i|$, the corresponding classifier is used (the value of $Q(u,v)$ determines the classifier).

Let us assume that we want to build the classifier for a quantization step q, where $q \in \mathcal{N}$. Let us assume that P_q contains normalized positions of peaks in $|\widetilde{H}|$ corresponding to double quantization. Please note that P_q can easily be generated by computing $|\widetilde{H}|$ of a random signal having a uniform distribution and being double quantized with step q and primary step q_α, where $q_\alpha = 1 \cdots n$, $q \neq q_\alpha$.

The feature vector, v_i, corresponding to $|\widetilde{H}_i|$, is constructed by taking the values of $|\widetilde{H}_i|$ in peak positions.

Our training set is consisted of 2000 uncompressed lab images (different kinds of images with narrow, wide, typical, untypical intensity histograms). Half of the images is employed for the training purposes and the second half for testing purposes.

In order to train classifier for quantization step q, we need both single quantized DCT coefficients (with quantization step q) and DCT coefficients double quantized with the secondary quantization step q. To obtain single quantized coefficients, 1000 uncompressed images where compressed using the quantization step q. To obtain double–compressed feature vectors, non-compressed images were first JPEG compressed using the quantization step q_α and the re–compressed using q. Only q_α which brings detectable peaks into $|\widetilde{H}|$ were employed. Only DCT coefficients corresponding to DC component and AC component $(1, 0)$ were used for the training purposes.

Our classifiers are soft–margin support vector machines (SVM) with the the Gaussian kernel $k(x, y) = exp(-\gamma||x - y||^2)$. The false positive rate was controlled to be 1 percent. In our experiments we trained classifiers for quantization steps $1 \cdots 25$.

To test the method, 1000 images were compressed resulting in single JPEG compressed images (using quality factor Q_α and JPEG standard quantization matrix). Then each single compressed image was re–compressed using a quality factor Q_β, resulting in a double JPEG compressed image. Detection accuracies of the classifier are reported in Table 1.

Table 1: Detection accuracy [%] as a function of different JPEG compression factors.

$Q_\beta \backslash Q_\alpha$	60	65	70	75	80	85	90	95
60	–	98	98	93	69	97	69	27
65	96	–	93	98	62	92	35	35
70	100	86	–	98	95	71	44	1
75	100	100	98	–	95	95	81	2
80	100	100	100	100	–	95	93	1
85	100	100	100	100	100	–	96	95
90	100	100	100	100	100	100	–	99
95	100	100	100	100	100	100	100	–

As is apparent from Table 1, the detection accuracies are higher for $Q_\beta > Q_\alpha$. When $Q_\beta < Q_\alpha$, generally, the DCT coefficient histograms have a shorter support. Furthermore, the introduced periodic properties have a larger period (due to the fact that $Q_\beta(u,v) > Q_\alpha(u,v)$). In some cases (for example, $Q_\alpha = 95$ and $Q_\beta = 70$), the detection accuracy is almost zero. This is because the fact that $\frac{Q_\beta(u,v)}{Q_\alpha(u,v)}$ is an integer value for employed DCT frequencies.

Generally, the content of the image also has an important impact on detecting the traces of double JPEG. Images containing heavy textures or images containing large uniform regions have different properties in their histograms of DCT coefficients comparing to natural images. Unfortunately, we do not have available the test set used in [18] allowing a direct comparison.

6. FLICKR EXPERIMENT

In order to carry out the main experiment of this paper, we needed to download and collect a large number of digital images. Keeping at disposition a variety of popular photo–sharing servers from which photos can by downloaded, we have opted for Flickr, one of the most popular photo sharing sites. We randomly selected 100.000 Flickr users and downloaded one photo per user resulting in 100.000 photos.

First we analyzed the quality factor of JPEG images and divided photos into three quality factor groups (see Table 2). JPEG quality factor has been estimated using quantization tables of JPEG files. After that, all photos have been divided into two further groups: original and modified photos. This has been done using EXIF examination. Whenever any traces of software manipulation has been found in the EXIF data, the photo has been denoted as modified. We employed typical ways of EXIF examination like searching for a software tag or inconsistencies original and modification dates, etc.

The population of different quality factors associated with original and modified images in our set is shown in Table 2 as well. The method described in the previous section has been applied to all EXIF denoted original and modified images. Results of detected JPEG images having traces of double compression are shown in the same table. For example, the last columns in Table 2 says that 1% of images in our image set have been denoted as original and 4% as modified based on EXIF examination. 7% of those 1% denoted as original by EXIF have been found to be altered by applying the double compression detection method to them. The last cell says that 27% of those 4% denoted as modified by EXIF also have been found to be altered using the method detecting double compression artifacts.

Table 2: Population of different JPEG qualities in our test set and corresponding rate of detected altered images using EXIF analysis and double JPEG test. For example, first cell in the first row shows that 37% of JPEG images of quality factor of 85% - 100% in our set were denoted as original by EXIF test. In the same time, 12% of those 37% have been found being double JPEG compressed (altered) by using the double JPEG detection method.

$Type \backslash QF$	85% – 100%	70% – 84%	55% – 69%
EXIF Original	[37%] (12%)	[8%] (9%)	[1%] (7%)
EXIF Modified	[36%] (17%)	[14%] (26%)	[4%] (27%)

All experiments have been done in Matlab. In order to simply and effectively read JPEG head data and DCT coefficients, we employed a free and easy to use library provided by VerifEyed^{TM} (can be downloaded from www.verifeyed.com).

7. DISCUSSION

Results obtained in the experimental part are interesting and show that only relying on image head consistency check when original image data are needed leads to a high number of undetected altered images. On the other hand, it also was shown that when detection of altered image data is carried out only by analyzing double compression artifacts, many of altered photos remain undetected. To summarize, apparently combining various image forensics methods is a must.

8. ACKNOWLEDGMENTS

This work has been supported by the Czech Science Foundation under the project no. GACR P202/10/P509 and project no. GACR 102/08/0470. Furthermore, by the Ministry of the Interior of the Czech Republic under the project no. MVCR VG20102013064.

9. REFERENCES

[1] B. Mahdian and S. Saic, "A bibliography on blind methods for identifying image forgery," *Image Commun.*, vol. 25, no. 6, pp. 389–399, 2010.

[2] H. T. Sencar, M. Ramkumar, and A. N. Akansu, *Data Hiding Fundamentals and Applications: Content Security in Digital Multimedia*. Orlando, FL, USA: Academic Press, Inc., 2004.

[3] M. Wu and B. Liu, *Multimedia Data Hiding*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2002.

[4] M. Schneider and S. F. Chang, "A robust content based digital signature for image authentication," in *IEEE International Conference on Image Processing (ICIP'96)*, 1996. [Online]. Available: citeseer.ist.psu.edu/schneider96robust.html

[5] C.-H. Tzeng and W.-H. Tsai, "A new technique for authentication of image/video for multimedia applications," in *Proceedings of the 2001 workshop on Multimedia and security*. New York, NY, USA: ACM Press, 2001, pp. 23–26.

[6] M. Arnold, M. Schmucker, and S. D. Wolthusen, *Techniques and Applications of Digital Watermarking and Content Protection*. Norwood, MA, USA: Artech House, Inc., 2003.

[7] N. Nikolaidis and I. Pitas, "Robust image watermarking in the spatial domain," *Signal Processing*, vol. 66, no. 3, pp. 385–403, May 1998.

[8] J. Fridrich and T. Pevny, "Detection of double–compression for applications in steganography," *IEEE Transactions on Information Security and Forensics*, vol. 3, no. 2, pp. 247–258, June 2008.

[9] B. Mahdian and S. Saic, "Detecting double compressed jpeg images," in *The 3rd International Conference on Imaging for Crime Detection and Prevention (ICDP-09)*, London, UK, December 2009.

[10] Y.-L. Chen and C.-T. Hsu, "Detecting doubly compressed images based on quantization noise model and image restoration," in *IEEE Workshop on Multimedia Signal Processing*, 2009.

[11] Z. Fan and R. L. de Queiroz, "Identification of bitmap compression history: Jpeg detection and quantizer estimation," *IEEE Transactions on Image Processing*, vol. 12, no. 2, pp. 230–235, 2003.

[12] H. Farid, "Exposing digital forgeries from jpeg ghosts," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 4, pp. 154–160, 2009.

[13] X. Feng and G. Doerr, "Jpeg recompression detection," in *SPIE Conference on Media Forensics and Security*, 2010.

[14] J. Fridrich and J. Lukas, "Estimation of primary quantization matrix in double compressed jpeg images," in *Proceedings of DFRWS*, vol. 2, Cleveland, OH, USA, August 2003.

[15] D. Fu, Y. Q. Shi, and W. Su, "A generalized benford's law for jpeg coefficients and its applications in image forensics," in *SPIE Electronic Imaging: Security, Steganography, and Watermarking of Multimedia Contents*, San Jose, CA, USA, January 2007.

[16] W. Luo, Z. Qu, J. Huang, and G. Qiu, "A novel method for detecting cropped and recompressed image block," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, Honolulu, HI, USA, April 2007, pp. 217–220.

[17] R. Neelamani, R. L. de Queiroz, Z. Fan, S. Dash, and R. G. Baraniuk, "Jpeg compression history estimation for color images," *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1365–1378, 2006.

[18] A. C. Popescu, "Statistical tools for digital image forensics," Ph.D. dissertation, Ph.D. dissertation. Department of Computer Science, Dartmouth College, Hanover, NH, 2005. [Online]. Available: www.cs.dartmouth.edu/farid/publications/apthesis05.html

[19] Z. Qu, W. Luo, and J. Huang, "A convolutive mixing model for shifted double jpeg compression with application to passive image authentication," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Las Vegas, USA, April 2008, pp. 4244–1483.