SPECTRAL AMPLITUDE ESTIMATION-BASED X-RAY IMAGE RESTORATION: AN EXTENSION OF A SPEECH ENHANCEMENT APPROACH

Til Aach
Medical University of Lübeck, Ratzeburger Allee 160, D-23538 Lübeck, GERMANY,
e-mail: aach@informatik.mu-luebeck.de
Dietmar Kunz
Philips Research Laboratories, Weisshausstr. 2, D-52066 Aachen, GERMANY
e-mail: kunz@pfa.research.philips.com

ABSTRACT

This paper describes a class of spectral amplitude estimation-based algorithms for the restoration of low dose X-ray images. Since estimation of spectral amplitude from noisy observations is a widely reported approach to restore noisy speech signals, we discuss our algorithms from the viewpoint of extending a well established speech processing technique to images. We include an analysis of residual noise. Moreover, we provide qualitative and quantitative processing results.

1 INTRODUCTION

In spectral amplitude estimation algorithms, noise reduction is achieved by attenuating observed frequency coefficients G_k depending on their *instantaneous* signal-to-noise ratios (SNR) r_k^2 according to

$$\hat{F}_k = G_k \cdot h(r_k) \ , \quad \text{where } r_k^2 = |G_k|^2 / \Phi_n(k) \ , \qquad (1)$$

where $\Phi_n(k)$ denotes an (estimate of) the noise power spectrum (NPS), and $\hat{F}(k)$ the estimate for the noise-free spectral coefficient F_k. The *attenuation function* $h(r)$ takes values between zero and one, and increases monotonically over r_k. The total effect is that each coefficient is attenuated the more, the more likely it is to represent mainly noise. Example curves – based on an MMSE-approach (cf. [1]) – are shown in Fig. 1. When applied to speech, the observed coefficients are obtained by e.g. the Discrete Fourier Transform (DFT) or Discrete Cosine Transform (DCT) from short, overlapping time intervals in order to adapt to the short time stationarity of speech.

The reason why we chose this concept to process clinical low dose X-ray images is that in these images, noise has a lowpass-shaped and potentially anisotropic NPS [1, 2]. Through the NPS $\Phi_n(k)$ in (1), a spatial frequency-domain approach can straightforwardly take this property into account. To adapt this concept to short space stationarity (lines, edges, etc.) of images, processing is now carried out using small, overlapping blocks.

With respect to the human observer, this processing pursues a twofold goal: a first target is to enhance the perceived quality of low dose X-ray images, for instance to reduce observer fatigue in clinical routine. Secondly, one wants to improve task-oriented diagnostic image quality, which is often measured by determining the receiver operating characteristic of human observers asked to detect details hidden in the images [3].

Original applications of this concept to speech target corresponding objectives, viz. to enhance subjective image quality (also in order to reduce potential listener fatigue), and to increase speech intelligibility in noisy conditions [4, 5, 6], measured by e.g. the diagnostic rhyme test. Spectral amplitude estimation methods usually achieve only the former objective when applied to speech corrupted by broadband noise [6], and both objectives for speech degraded by narrow band background noise with high intensity [4, p.24].

2 MODELS FOR SIGNAL AND NOISE

The exact shape of the attenuation function $h(r)$ in (1) depends on noise and signal models as well as on the objective function [1, 7, 8]. Noise in the spectral domain is adequately modelled as being complex Gaussian. Models for speech distinguish between periods when speech is actually present, and silent intervals. Presence of speech can in turn be regarded as a time series of the alternating states "voiced speech" and "unvoiced speech". Production of voiced speech is generally modelled by a time-varying linear filter corresponding to the vocal tract, which is excited by a train of pulses [5]. Restoration can then be formulated as estimation of amplitudes of deterministic spectral lines in Gaussian noise [8, sec.IIC]. For voiced speech we can furthermore assume that signal energy is well compressed by the transform used in (1) [7, p.1114].

In unvoiced speech, the vocal tract is excited by a noise-like signal rather than a deterministic pulse train. Hence, a stochastic signal model is more appropriate. Also, compaction of signal energy is less pronounced than in voiced speech. In practice, however, speech restoration algorithms use only a single model for both types of speech, as otherwise the hidden states would have to be estimated.

Similarly as in transform image coding, our image restoration approach assumes that within each block, the transform compresses the signal to be recovered into only a few spectral lines G_k. In all other coefficients, signal is not present. For each coefficient, the probability density function for the signal is hence a composite of a Dirac impulse at zero amplitude weighted by the probability P_0 of signal absence, and a complex Gaussian density weighted by the probability P_1 of signal presence. Assuming furthermore that in case of signal presence its variance is much larger than the noise variance (but otherwise unknown) [1], the following attenuation function was derived based on an MMSE-approach in [1]:

$$h(r) = (1 + \lambda \exp(-\alpha r^2))^{-1} \quad , \qquad (2)$$

with α being a weighting factor similar to the one used in generalized Wiener filters. The parameter λ is proportional to the ratio P_0/P_1. For the present, it is regarded as a free parameter controlling the trade-off between noise reduction and signal preservation. This attenuation function is plotted in Fig. 1 for $\alpha = 1$ and $\lambda = 1.5$.

2.1 Noise Estimation

A crucial issue is the estimation of the NPS $\Phi_n(k)$ used in (1). In speech processing, the NPS can conveniently be estimated during silent periods. Changes of noise properties during intervals with speech can, however, result in mismatches of observed and estimated noise properties [10]. These mismatches are one cause of a type of residual noise known as *musical noise*. In contrast to this, the NPS is generally well known when processing X-ray images, since it can be predicted with sufficient accuracy from acquisition parameter settings of the X-ray system and measurements of the system's transfer function. Residual noise in processed images is then mainly caused by the fact that the instantaneous SNR in (1) compares an observed *realization* of a random variable to (an estimate of) the NPS, which is an expected value.

2.2 Residual Noise Mechanisms

The mentioned musical noise in processed speech is formed by "switched" spectral spikes which change randomly from interval to interval. The corresponding phenomenon in processed images is the random occurrence of oriented sine patterns. In both cases, this type of residual noise can be kept low by retaining a low wide band noise "floor" masking these spikes [1, 4]. Moreover, using smooth attenuation functions – like the ones in Fig. 1 – can help to keep residual noise low. To see this, consider several realizations of the coefficient G_k containing a constant signal contribution F_k. Then, r_k fluctuates around the operating point $R = |F_k|/\sqrt{\Phi_n(k)}$. As these fluctuations are caused by noise, their influence on the output coefficient $\hat{F}_k$ should be low if not eliminated. From (1), the differential variation of the estimated coefficient is

$$\frac{d|\hat{F}_k|}{dr_k}|_R = \sqrt{\Phi_n(k)} \cdot \left[\frac{dh(r_k)}{dr_k}|_R \cdot R + h(R) \right] \quad . \qquad (3)$$

This relation shows that "differential noise" can be kept low if a smooth suppression curve free of steep slopes is chosen.

Some speech restoration versions of spectral amplitude estimation actually use knowledge of operating points by means of attenuation functions which depend not only on the instantaneous SNR r_k^2, but additionally on an *a priori* SNR s_k^2 [7, 8]. In [8, sec.IID], this a priori SNR is regarded as a free "suppression factor" to be specified externally. In contrast to this, the algorithm of Ephraim and Malah [7] predicts the a priori SNR *recursively* over successively processed time intervals (see also [11]). In this case, the a priori SNR is even the major factor influencing the attenuation function $h_{EM}(r_k, s_k)$. It can be shown that in noise-dominated intervals, where the instantaneous SNR r_k^2 fluctuates strongly, the variance of the recursively filtered a priori SNR s_k^2 is much lower than that one of r_k^2 [11]. Still, the a priori SNR is sensitive to abrupt signal transients, to which it reacts with a delay of one time frame. The overall effect is an almost complete elimination of musical noise with only minor signal distortion [7, p.1119],[11].

2.3 Treatment of the Phase

Eq. (1) is a zero-phase operation. In speech processing, leaving the noisy phase can be justified by the insensitivity of the human ear to minor phase distortions, which cannot be perceived as long as the SNR is about 6dB or better [10]. In images, phase noise results in displacements of the corresponding planar (co)sine waves. Although we are not aware of a similar perception threshold in the human visual system, it is known that phase bears the dominant part of the information in an image [12, p. 59]. Retaining the noisy phase in the filtered images can therefore be justified by arguing that inevitable phase estimation errors could lead to highly visible artifacts. More significant, however, is the fact that for both speech and images, keeping the noisy phase can be shown to be optimal given the absence of prior knowledge about the phase signal, and the fact that only a single coefficient is observed at a time [1, 7].

3 USING LOCAL ORIENTATION

The multidimensional nature of images makes it possible to exploit oriented signal patterns, like lines and edges,

[1] Experiments show that the distributions of subband filter bank (i.e. spectral) coefficients exhibit a larger ratio of fourth to second central moment than the Normal distribution does, what indicates increased probabilities for very large and very small absolute values [9]. Our combination of a discrete probability at zero with a large-variance Gaussian pdf can be regarded as a reasonable and mathematically convenient approximation to these experimental results.

for filtration, what can considerably improve the performance of spectral magnitude estimation particularly with respect to perceptually important line and edge information. When the DFT is used in (1), the occurrence of an oriented structure in an image block results in the spectral domain in a concentration of energy along the line which is tilted 90 degrees to the spatial orientation and passes through the origin. To detect the presence of local orientation as well as its direction, we interpret the spectral energy as a distribution of mass across the two-dimensional discrete spatial frequency plane of each block. Orientation within each block can then be detected by means of a 2×2 *inertia* matrix, the eigenvectors of which determine in a least squares sense the axes along which energy concentration is strongest (local orientation) and least, respectively [13]. The degree of energy concentration can be derived from the corresponding eigenvalues.

Coefficients along the line of local orientation are highly likely to contribute to perceptually important detail information. These coefficients can now be subjected to reduced attenuation, or even be enhanced. The attenuation of other coefficients can be made dependent on their relative position to the local orientation axis. An example of such an angle-dependent family of attenuation curves is shown in Fig. 1.

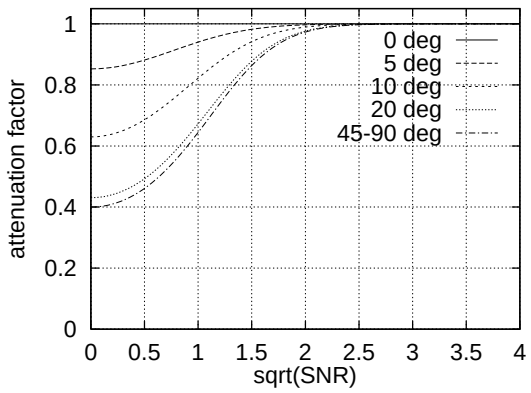

Figure 1: A family of attenuation curves plotted versus the square root r of the instantaneous SNR r^2. The 90-degrees curve corresponds to (2). For a given SNR, attenuation increases with the angular distance between coefficient G_k and the orientation.

4 RESULTS

Fig. 2 depicts a low dose X-ray image of a patient's intestines. Fig. 3 shows this image being processed by orientation-dependent noise reduction and enhancement [13]. A direct comparison confirms that on the one hand, noise could indeed be visibly reduced. Simultaneously, oriented structures, like fissures or transitions from intestines to background are clearly enhanced.

Quantitative measurements reveal that our intraframe filters typically reduce noise to less than one half of the original noise power (see [1]). Moreover, example measurements of the SNR before and after processing show that simultaneously image detail is well preserved, thus indeed resulting in an improved SNR. This is shown in Fig. 4 based on measurements from noisy realizations of a simulated, representative test image.

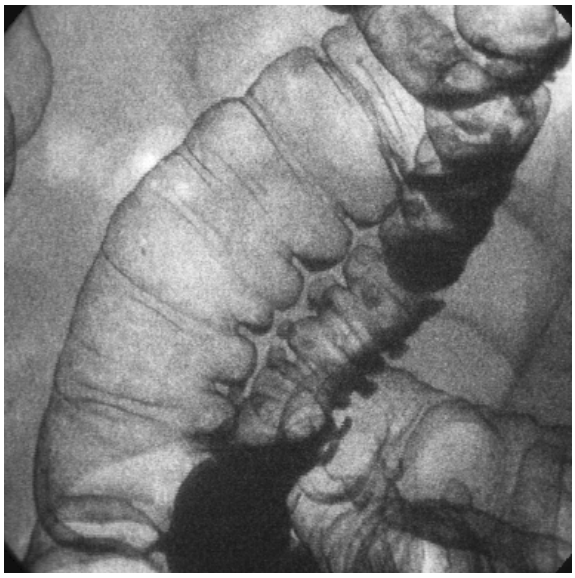

Figure 2: Portion of an original fluoroscopy image, depicting a patient's intestines.

Figure 3: Fig. 2 processed by orientation-dependent noise reduction and enhancement as described in Sec. 3.

5 CONCLUDING REMARKS

In this paper, we discussed an extension of spectral amplitude estimation from speech restoration to the processing of low-dose X-ray images. We showed that the goals one seeks to achieve by such processing are indeed

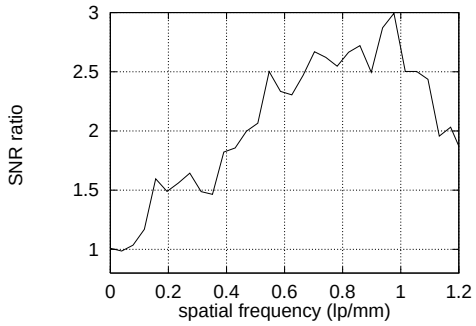

Figure 4: Ratio of SNR after processing to SNR before processing as a function of spatial frequency. Measurements based on 30 realisations of a simulated test image containing a thin guidewire and quantum noise corresponding to an acquisition at $3\mu R$.

comparable, viz. improvements in subjective as well as diagnostic image quality. The main reason why we based our X-ray image processing on spectral amplitude estimation is that the properties of noise in low-dose X-ray images are conveniently described in the spectral domain by a lowpass-shaped, potentially anisotropic (and signal-dependent, cf. [1]) noise power spectrum. The noise parameters can in our application be assumed as known from off-line system measurements, and settings of system parameters during image acquisition. In contrast to this, noise parameters in speech processing can – at least in single-microphone algorithms – only be estimated from the noisy signal itself, for instance during silent periods, or by minimum statistics in the presence of speech.

In both applications, the employed signal models rely mainly on the fact that the analyzing transform is more or less able to compress energy into only a limited number of spectral coefficients. In speech processing, a distinction is furthermore made between the states of speech being present, and silent intervals.

We then analyzed the differential nature of the noise which remains after processing, and showed how a smooth suppression curve can help to keep differential noise low. Note that some approaches to derive a noise suppression curve, like the ones starting from a Wiener filter or power spectrum equalization filter (see e.g. [5, 14]), do actually result in suppression curves with abrupt bends, which turned out to be quite unsuitable for our application [1].

In this context, we furthermore pointed out a methodology where noise suppression – and hence also residual noise – depend on a recursively estimated a priori SNR in addition to the observed a posteriori SNR as used in (1). This approach – reported in [7], and further examined in [11] – is, however, not applicable to our image processing problem since it assumes that signal varies more slowly over adjacent time frames than noise. Accordingly, only signal transients, but no impulses, are considered by Cappé in [11]. This may be certainly justified in speech processing, where, assuming a sampling rate of 10 kHz and time frames comprising 256 samples with 128 samples overlap, a new frame emerges every 12 ms, which is shorter than most speech sound components. Our image restoration algorithm is based on block sizes between 32×32 to 128×128 pixels, which overlap by between 8 to 32 pixels in each dimension, and are larger than image details of potentially diagnostic importance. As such details can hence not be predicted from adjacent blocks, our algorithms relies on only one block at a time. On the other hand, the multidimensional nature of image blocks makes it possible to use within each image block potentially occurring orientation, as discussed in section 3.

References

[1] T. Aach, D. Kunz: Spectral estimation filters for noise reduction in X-ray fluoroscopy imaging. Proc. EUSIPCO-96, Trieste, Sept. 10–13, 571–574, 1996.

[2] T. Aach, U. Schiebel, G. Spekowius: Digital image acquisition and processing for medical X-ray imaging applications. Proc. Intl. Symp. Electr. Photog. (ISEP), Cologne, Sept. 21–22, 82-90, 1996.

[3] A. L. Evans, *The Evaluation of Medical Images*. Bristol: Adam Hilger Ltd, 1981.

[4] D. van Compernolle, "DSP techniques for speech enhancement," in *Speech Processing in Adverse Conditions*, Cannes-Mandelieu, 10.-13. Nov., 21–30, 1992.

[5] J. S. Lim, A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proc. IEEE* 67(12), 1586–1604, 1979.

[6] J. S. Lim, "Evaluation of a correlation subtraction method for enhancing speech degraded by additive white noise," *IEEE Trans. Acoust. Speech Sig. Proc.* 26(5), 471–472, 1978.

[7] Y. Ephraim, D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust. Speech Sig. Proc.* 32(6), 1109–1121, 1984.

[8] R. J. McAulay, M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. Acoust. Speech Sig. Proc.* 28(2), 137–145, 1980.

[9] E. P. Simoncelli, E. H. Adelson, "Noise removal via Bayesian wavelet coring," Proc. ICIP, Lausanne, Sept. 16–19, 379–382, 1996.

[10] P. Vary, "Noise suppression by spectral magnitude estimation – mechanism and limits," *Signal Processing* 8(4), 387–400, 1985.

[11] O. Cappé, "Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor," *IEEE Trans. Speech Audio Proc.* 2(2), 345–349, 1994.

[12] B. Jähne, "Digitale Bildverarbeitung", Springer Verlag, Berlin, 1989.

[13] T. Aach, D. Kunz, "Anisotropic spectral magnitude estimation filters for noise reduction and image enhancement", Proc. ICIP, Lausanne, Sept. 16–19, 335–338, 1996.

[14] J. S. Lim, "Image restoration by short space spectral subtraction," *IEEE Trans. Acoust. Speech Sig. Proc.* 28(2), 191–197, 1980.