prime. If the proposed algorithm is combined with radix-2 fast algorithms, the overall computational complexity can be minimised, e.g. we can assume $N = p.q$, where p is an odd integer and $q = 2^m$. The radix-2 algorithm is used for the computation of length-q GDHTs. For $N = p.2^m$, where p = 3, 5 and 15, Fig. 2 compares the number of arithmetic operations needed by the proposed algorithm and the algorithm in [5] which appears to use the smallest number of operations reported in the literature. Comparison with other prime factor algorithms for the GDHT is not made due to the lack of such algorithms in the published literature. Fig. 2 shows that the proposed algorithm requires the lowest computational complexity for $p = 3$ and highest computational complexity for $p = 5$. The computational complexity required by the proposed algorithm for all cases is much less than that achieved by the algorithm in [5] for $p = 1$, due mainly to the elimination of twiddle factors in the prime-factor decomposition.

Conclusion: A fast prime-factor algorithm for the GDHT has been presented. Compared to other reported algorithms, the proposed algorithm achieves substantial savings in terms of the number of additions and multiplications, and requires a simple and general index mapping process. In-place computation was also achieved which enables easy implementation.

Appendix A: If p and q are relatively prime, eqns. 3, 7 and 8 can be transformed into a length-p type-II DCT. For simplicity, we ignore $(-1)^n$ in eqn. 8. According to the Chinese remainder theorem, we can always find a set of integers j so that for each n

$$j = (n + iq) \bmod p \qquad 0 \le i, j \le p-1 \quad 0 \le n < q \quad (14)$$

Based on eqn. 14, eqns. 7 and 8 can be expressed as

$$u(n,m) = \sum_{j=0}^{p-1} x[l(n,j)] \cos \frac{\pi(2j+1)(2m)}{2p} \qquad 0 \le m \le (p-1)/2 \quad (15)$$

$$v(n,m) = \sum_{l=0}^{p-1} x[l(n,j)] \cos \frac{\pi(2j+1)(p-2m)}{2p} \qquad 1 \le m \le (p-1)/2 \quad (16)$$

where $l(n, j)$ is the index of the original input sequence associated with the values of n and j. Eqn. 15 corresponds to the computation associated with even indexed DCTs and eqn. 16 corresponds to the computation associated with odd indexed DCTs. Fig. 1 shows the relationship between j and $l(n, j)$.

14 July 1999
Electronics Letters Online No: 19991208
DOI: 10.1049/el: 1991208

Guoan Bi and Chao Lu (*School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, Republic of Singapore*)

References

1 WANG, Z.: 'Harmonic analysis with a real frequency function, I aperiodic, case, II periodic and bounded cases, and III data sequence', *Appl. Math. Comput.*, 1981, **9**, pp. 53–255

2 WANG, Z.: 'Fast algorithms for the discrete W transform and for the discrete Fourier transform', *IEEE Trans. Acoust. Speech Signal Process.*, 1984, **ASSP-32**, pp. 803–816

3 XI, J., and CHICHARO, J.F.: 'Computing running discrete Hartley transform and running discrete W transforms based on the adaptive LMS algorithm', *IEEE Trans. Circuits Syst. II*, 1997, **44**, (3), pp. 257–260

4 WANG, Z.: 'Comments on 'Generalized discrete Hartley transform'', *IEEE Trans. Signal Process.*, 1995, **43**, (7), pp. 1711–1712

5 NENG-CHUNG HU, *et al.*: 'Generalized discrete Hartley transforms', *IEEE Trans. Signal Process.*, 1992, **40**, (12), pp. 2931–2940

6 HU, , and LIU, F.F.: 'Fast computation of the two-dimensional generalized Hartley transforms', *IEE Proc. Vis. Image Signal Process.*, 1995, **142**, (1), pp. 35–39

7 SOO-CHANG PEI, and TZYY-LIANG LUO : 'Split-radix generalized fast Fourier transform', *Signal Process.*, 1996, **54**, pp. 137–151

Speech enhancement based on hybrid algorithm

Zhong-Xuan Yuan, Soo Ngee Koh and Ing Yann Soon

A hybrid algorithm that incorporates the minimum mean square error (MMSE) algorithm and the auditory masking algorithm for speech enhancement is proposed. The proposed algorithm results in a significant reduction in residual noise compared with that in the enhanced speech produced by the MMSE algorithm alone.

Introduction: Many successful single channel speech enhancement algorithms are based on the estimation of short time spectral amplitude (STSA) with the assumption that the phase of the noisy speech is not perceived by the human ear. The existing major speech enhancement algorithms can be grouped into two classes: one is based on a statistical approach, such as the maximum likelihood estimator and optimal minimum mean square error (MMSE) estimator [1, 2] and the other is based on the perceptual point of view. The latter enhancement approach is based on the lateral inhibition principles of auditory nerves [3] or the psycho-acoustical auditory masking phenomenon [4, 5].

Unvoiced speech is noise-like and in general has a low energy level. According to [6], if the unvoiced speech segments are replaced by a different signal with noise-like fine structure and a similar spectral envelope, then the perceived quality does not drop significantly. These characteristics make the spectral components of unvoiced speech more easily satisfy the assumption of the MMSE algorithm. Since the energy of the spectral components of unvoiced speech is comparable to or much lower than that of the noise signal, it is very difficult to effectively mask the residual noise in the enhanced speech. Alternatively, voiced speech can be interpreted as a sequence of pitch-cycle waveforms. The general shape of these pitch-cycle waveforms evolves slowly as functions of time and adjacent cycles are often highly interdependent. Since the spectral components between adjacent analysis frames are correlated to a certain degree, the statistical independence assumption in the MMSE algorithm is not applicable. Since most of the energy of a voiced speech signal is concentrated at the harmonics of the fundamental frequency, these harmonic spectral components can be used to effectively mask residual noise. Based on the analysis just outlined, we propose a hybrid algorithm that incorporates the MMSE algorithm for low energy critical bands (CBs) and the auditory masking algorithm for high energy CBs.

Hybrid algorithm: Suppose that the noisy speech model is $y(n) = x(n) + d(n)$, where $y(n)$, $x(n)$ and $d(n)$ denote noisy speech, clean speech and noise, respectively. We let $X_k = A_k \exp(j\alpha_k)$, $Y_k = R_k \exp(j\theta_k)$ and D_k denote the kth spectral component of the signal $x(n)$, the noisy observations $y(n)$ and the noise $d(n)$, respectively. Assuming that both the speech and noise are uncorrelated, we have $R_k^2 = A_k^2 + D_k^2$. The proposed algorithm is performed on each CB. It consists of the following steps:
(i) determining speech/noise
(ii) windowing of noisy speech/noise segments with a Hanning window
(iii) performing spectrum decomposition using an FFT, keeping the phase of the noisy speech for synthesising the final enhanced speech and making the estimation of the properties of noise
(iv) performing CB analysis and calculate the masking threshold
(v) choosing a suitable algorithm for each CB depending on the masking threshold of the CB; if the masking threshold is lower than or equal to the noise energy in a CB, then the MMSE spectral estimator is used; otherwise, the auditory masking algorithm is used
(vi) merging of the modified spectral amplitude components
(vii) performing IFFT on the merged amplitude components using the phase of the noisy speech
(viii) performing the overlap and add operation to obtain the enhanced speech.

In the low energy CB, the MMSE estimate, $\hat{A}_k$, of A_k used in the investigation is given as follows:

$$\hat{A}_k = \frac{\Lambda(\xi_k, \gamma_k, q_k)}{1 + \Lambda(\xi_k, \gamma_k, q_k)} G_{MMSE}(\xi_k, \gamma_k) R_k \Big|_{\xi_k = \eta_k/(1-q_k)} \quad (1)$$

where

$$G_{MMSE}(\xi_k,\gamma_k) = \Gamma(1.5)\frac{\sqrt{v_k}}{\gamma_k}M(-0.5;1;-v_k) \qquad (2)$$

$$\Lambda(\xi_k,\gamma_k,q_k) = \mu_k \frac{\exp(v_k)}{1+\xi_k}\bigg|_{\mu_k=(1-q_k)/q_k} \qquad (3)$$

$$v_k = \frac{\eta_k}{1+\eta_k}\gamma_k \qquad (4)$$

$$\eta_k(n) = \alpha\frac{\hat{A}_k^2(n-1)}{\lambda_d(k,n-1)} + (1-\alpha)\max\{\gamma_k(n)-1,0\} \quad (5)$$

$$\gamma_k = \frac{R_k^2}{\lambda_d(k)} \qquad (6)$$

and q_k is the probability of signal absence in the kth spectral component. In the experiment, $q_k = 0.2$ and $\alpha = 0.99$. ξ_k and γ_k are the *a priori* and *a posteriori* signal-to-noise ratios, respectively. $\eta_k(n)$ is the estimator of η_k in the nth analysis frame. M is the confluent hyper-geometric function. $\lambda_d(k) = E\{|D_k|^2\}$. A more detailed description can be found in [2].

For the high energy CB, the estimate, $\hat{A}_k$, of A_k is given as follows:

$$\hat{A}_k = G_{MASK}(k,\alpha,\gamma)R_k \qquad (7)$$

where

$$G_{MASK}(k,\alpha,\gamma) = \left[1-\alpha\left(\frac{D_k}{R_k}\right)^{\gamma}\right]^{1/\gamma} \qquad (8)$$

where $\gamma = 2$. α is called the over-subtraction factor and is defined as follows:

$$\alpha = 1 + 10\left(1 - \frac{1}{1+\exp(-amt)}\right) \qquad (9)$$

where *amt* is the auditory masking threshold in the CB. The algorithm for the estimation of *amt* can be found in [7]. Since *amt* has to be estimated from noisy speech, we have made some modifications to the original algorithm given by [7]. The main change is that the noise energy is subtracted from the noisy speech energy after performing the CB analysis on noisy speech.

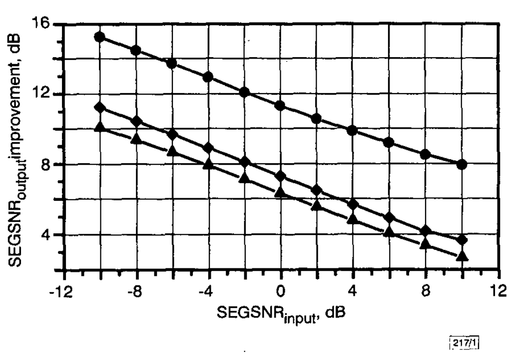

Fig. 1 *Segmental SNR improvement at various Gaussian white noise levels*

—●— theoretical limit
—◆— proposed algorithm
—▲— MMSE algorithm

Experimental results: The performance evaluation of the proposed algorithm relies mainly on objective measurements. The segmental signal-to-noise ratio (SNR) and bark spectral distortion (BSD) [8] are used. The test utterance, 'Jane may earn more money by working hard', is from the TIMIT database. It is sampled at 8kHz and quantised to 16 bit. The noisy speech is obtained by adding noise data to the clean speech data. For comparison, the same performance evaluations are performed on the enhanced speech produced solely by the MMSE algorithm. Fig. 1 shows the segmental SNR improvements obtained at various Gaussian white noise levels. Fig. 2 shows the BSD under the same testing conditions. The theoretical limits are also plotted in Figs. 1 and 2 as a benchmark.

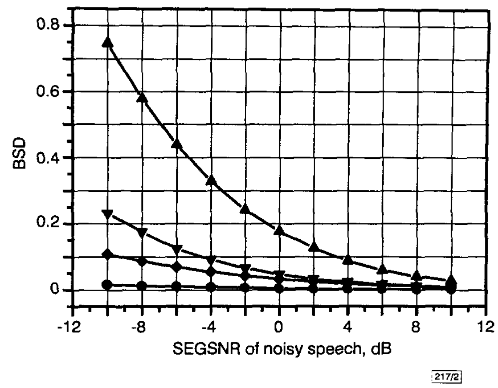

Fig. 2 *Bark spectral distortions for various Gaussian white noise levels*

—●— theoretical limit
—◆— proposed algorithm
—▼— MMSE algorithm
—▲— unprocessed

Table 1: SEGSNR improvement and BSD for various noises

Noise	SEGSNR improvement [dB]		BSD	
	Proposed	MMSE	Proposed	MMSE
Fan	17.2	14.4	0.0103	0.0145
White	9.67	8.63	0.0681	0.1247
F16	6.66	6.01	0.0628	0.0738

SEGSNR of input noisy speech is –6dB

Table 1 presents a comparison of segmental SNR and BSD at an input SNR = –6dB for three kinds of additive noise, i.e. fan noise, Gaussian white noise and F16 airplane noise. From Figs. 1 and 2 and Table 1, it can be seen that the proposed algorithm achieves a significant improvement over the MMSE enhancement algorithm. This objective improvement is also confirmed by an informal subjective evaluation. Eight clean speech sentences, four female and four male, from the TIMIT database are used for three kinds of noise at different noise levels. We found that the residual noise in the enhanced speech by the proposed algorithm has similar characteristics to, but a much smaller energy level than the classical MMSE algorithm.

Conclusion: Based on the properties of unvoiced and voiced speech, a hybrid algorithm based on MMSE and auditory masking for speech enhancement has been proposed. The proposed algorithm is performed on each critical band. The experimental results illustrate that the proposed algorithm can significantly reduce the residual noise in enhanced speech and outperform the speech enhancement scheme using only the MMSE approach.

14 July 1999
Electronics Letters Online No: 19991162
DOI: 10.1049/el:19991162

Zhong-Xuan Yuan, Soo Ngee Koh and Ing Yann Soon (*Communication Research Lab, School of Electrical and Electronic Engineering, Nanyang Technological University, Block S2, Nanyang Avenue, 639798, Singapore*)

References

1 MCAULAY, R.J., and MALPASS, M.L.: 'Speech enhancement using a soft-decision noise suppression filter', *IEEE Trans. Acoust. Speech Signal Process.*, 1980, **ASSP-28**, (2), pp. 137–145

2 EPHRAIM, Y., and MALAH, D.: 'Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator', *IEEE Trans. Acoust. Speech Signal Process.*, 1984, **ASSP-32**, (6), pp. 1109–1121

3 CHENG, YAN MING, and O'SHAUGHNESSY, D.: 'Speech enhancement based conceptually on auditory evidence', *IEEE Trans. Signal Process.*, 1991, **39**, (9), pp. 1943–1954

4 TSOUKALAS, D.E., MOURJOPOULOS, J.N., and KOKKINAKIS, G.: 'Speech enhancement based on audible noise suppression', *IEEE Trans. Speech Audio Process.*, 1997, **5**, (6), pp. 497–514

5 VIRAG, N.: 'Single channel speech enhancement based on masking properties of the human auditory system', *IEEE Trans. Speech Audio Process.*, 1999, **7**, (2), pp. 126–137

6 KUBIN, G., ATAL, B.S., and KLEIJN, W.B.: 'Performance of noise excitation for unvoiced speech'. Proc. IEEE Workshop on Speech Coding for Telecomm., 1993, pp. 35–36

7 JOHNSTON, J.D.: 'Transform coding of audio signals using perceptual noise criteria', *IEEE J. Sel. Areas Commun.*, 1988, **6**, (2), pp. 314–323

8 WANG, S., SEKEY, A., and GERSHO, A.: 'An objective measure for predicting subjective quality of speech coders', *IEEE J. Sel. Areas Commun.*, 1992, **10**, (5), pp. 819–829

Computation of resonant frequencies of dielectric loaded rectangular cavity using TLM method

J. Chuma, C.W. Sim and D. Mirshekar-Syahkal

The electromagnetic analysis of a dielectric loaded rectangular cavity using the 3D transmission line matrix method is presented. It is found that in order to excite all the modes within a frequency range, the field polarisation and location of the excitation brick in the TLM analysis must be properly selected. For the resonator of concern, the results of computation of the first 11 resonant frequencies show excellent agreement with the experimental results.

Introduction: Owing to their desirable properties and commercial availability at reasonable prices, dielectric resonators made from highly temperature stable low-loss ceramics are finding increasing application in microwave filters [1, 2]. Miniaturisation of microwave filters is the driving force behind the use of these ceramics.

Many elaborate numerical methods have been developed and reported for analysing dielectric loaded cavities. Examples are the mode matching technique [1 – 4], the finite difference method [5] and the finite element method [6]. Comparisons of these numerical methods can be found in [7, 8]. To the authors' knowledge, no analysis of dielectric loaded rectangular cavities has been carried out using the transmission line matrix (TLM) method. The TLM method is a highly accurate and efficient numerical method. However, if not used carefully, it can miss out some of the resonant frequencies.

In this Letter we describe the computation of resonant frequencies of a dielectric loaded rectangular cavity using a commercially available TLM method [9]. It is shown that, depending on the field polarisation and location of the excitation brick in the TLM analysis, only some of the modes can be excited. So far, this important point has not been properly elaborated in the literature and inexperienced users can find the technique unhelpful. We also present a comparison of the computed and measured frequencies for the first 11 modes.

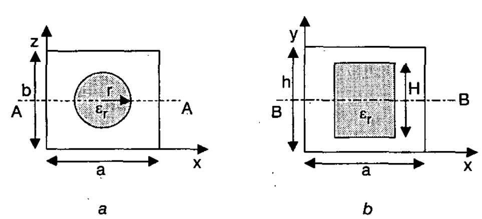

Fig. 1 *Configuration of dielectric loaded rectangular cavity*

a Equatorial plane (symmetry plane BB)
b Meridian plane (symmetry plane AA)

Analysis: The structure of the dielectric loaded rectangular cavity, which has been analysed, is shown in Fig. 1. The perfectly conducting metallic rectangular cavity of dimensions $a = b = 50$mm and $h = 30$mm is symmetrically loaded with a cylindrical dielectric (ceramic puck) of radius $r = 18$mm, height $H = 18$mm and dielectric constant $\varepsilon_r = 37$. The dielectric is assumed to be lossless.

To reduce the computation time and for the ease of resonant mode designation, symmetry conditions are employed. The modes are designated as $HEH_{nm\delta}$, $HEE_{nm\delta}$, $TEH_{0m\delta}$, $TEE_{0m\delta}$, $TMH_{0m\delta}$ and $TME_{0m\delta}$. The first two letters indicate whether the modes are hybrid (*HE*), transverse electric (*TE*) or transverse magnetic (*TM*). The third letter (*E* or *H*) indicates whether the symmetry plane BB in Fig. 1 is an electric wall or a magnetic wall. The first subscript *n* indicates the order of the angular variation of the fields. The second subscript *m* is the order of the resonant frequency and the third subscript δ indicates the axial field variations.

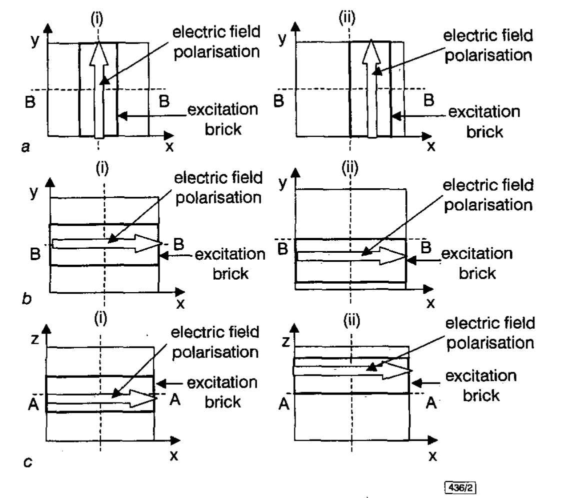

Fig. 2 *Different excitation locations for dielectric loaded rectangular cavity*

(i) symmetric excitation
(ii) asymmetric excitation
a Case A: Electrical field polarisation (y) and direction of propagation (z)
b Case B: Electrical field polarisation (x) and direction of propagation (z)
c Case C: Electrical field polarisation (x) and direction of propagation (y)

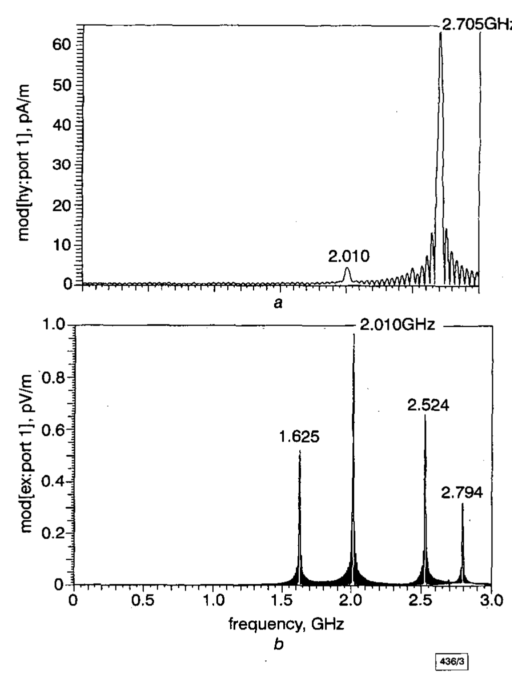

Fig. 3 *Resonant frequencies for two different excitation brick positions*

a Case A in Table 1
b Case C in Table 1

When using the TLM method, resonant modes in a dielectric loaded rectangular cavity can be excited either by placing the excitation brick symmetrically or asymmetrically with respect to one