A PSYCHOACOUSTIC-BASED ANALYSIS OF THE IMPACT OF PRE-ECHOES AND POST-ECHOES IN SOUNDFIELD RENDERING APPLICATIONS

L. Bianchi, F. Antonacci, A. Canclini, A. Sarti, S. Tubaro

Dipartimento di Elettronica ed Informazione, Politecnico di Milano
Piazza Leonardo da Vinci, 32, 20133 Milano, Italy

ABSTRACT

In this paper we propose two metrics for the evaluation of the impact of pre-echoes and post-echoes on the perceived quality in soundfield rendering applications. These metrics are derived from psychoacoustic-based considerations, in particular the masking effect, well known in the literature of perceptual coding. The measurement is accomplished through a virtual microphone array that samples the soundfield on a circumference. The soundfield within the circle is estimated by means of the circular harmonic decomposition. As a result, space-time impulse responses of the rendering system are obtained, which are then analyzed through the masking curve to extract the pre- and post-echoes metrics. A comparison between experimental and simulative results, conducted using the same setup, allows to discriminate the impact of the adopted rendering engine and of the non-idealities of the real system (environment and loudspeakers) on pre- and post-echoes.

Index Terms— Loudspeaker array, Circular Harmonic Decomposition, Pre-echoes, Post-echoes

1. INTRODUCTION

Soundfield rendering techniques, such as Wave Field Synthesis [1], suffer from non-idealities. As an example, in [2] the authors have shown that the spatial sampling causes artifacts in the rendered wavefield, due to the use of point-like loudspeakers. Authors show that even the analysis on simulative results reveals the presence of artifacts. When working in a real environment, other non-idealities appear, such as a frequency-dependent response of the emission system (including D/A converter and amplification stages); the non-omnidirectional nature of the loudspeakers; and the reflective behavior of the walls of the room in which the rendering system is operating.

In this paper we aim at characterizing the artifacts introduced by these non-idealities adopting a space-time representation of the soundfield, in which propagating wavefronts at different time instants are analyzed. It has been shown in [2] that the above mentioned non-idealities produce acoustic wavefronts that impinge on the listening area before or after the desired wavefront; they are named *pre-echoes* and *post-echoes*, respectively. An evaluation of these artifacts that does not account for psychoacoustic effects could be not informative to evaluate the impact on the human perception. In [3] the authors show that these artifacts affect the perception of quality of a rendered soundfield, but a methodology to quantify the effect is not introduced. The definition of a metric that incorporates psychoacoustic-based criteria for the analysis of the impact of pre- and post-echoes is therefore in order.

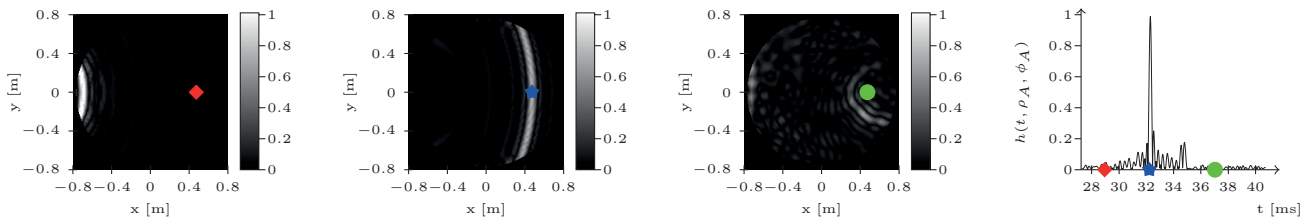
It can be noticed from simulations presented in [2] that for typical rendering systems installed in small auditoria, pre- and post-echoes are mainly concentrated in a short time window (about 10 ms) around the main desired peak, so that they are not perceived as separated echoes. Hence an analysis based on the well known *precedence effect* does not provide useful information. Contrarily, we have based our analysis on the *masking effect*, according to which we are able to explain the different impact of pre- and post-echoes. More specifically, we adopt the masking curve presented in [4] and widely adopted in the literature of perceptual coding.

In order to separate the impact of artifacts introduced by the spatial sampling from those introduced by non-idealities due to the the real-world acoustic setup, we proceed in two steps. First, the analysis is performed on simulative data in order to highlight the impact of the finite size and discrete nature of the loudspeaker array, and then the analysis is repeated on real data, to highlight the impact of the real acoustic scene. In order to acquire the space-time impulse response of the rendering system, we adopt the analysis methodology presented in [5, 6], where an extrapolation on the soundfield measured over discrete positions in space is used to estimate the soundfield within a limited region. For our purposes, the rendering system produces an excitation signal that is white in a limited bandwidth. The response of the environment to the rendered signal is recorded over a set of points disposed on a circumference and the soundfield in the bounded circle is extrapolated through the technique presented in [5], to obtain a space-frequency representation of the soundfield. By performing an inverse Fast Fourier Transform we obtain the space-time impulse response of the rendering system. An analysis based on the masking curve is then accomplished. We remark that this methodology is general enough to be employed with most of the rendering techniques. We also provide some examples of the proposed metric for the geometric rendering technique [7]. We are currently working on the validation of the proposed metric by comparing experimental results with subjective evaluation.

The remainder of the paper is structured as follows: Section 2 illustrates the procedure to obtain a space-time impulse response from measurements of the rendered wave field on a circumference. Section 3 introduces a background on masking in time-domain and describes the proposed psychoacoustic-based metric. Section 4 provides some experimental results. Finally, Section 5 draws some conclusions.

2. MEASUREMENT OF THE SPACE-TIME IMPULSE RESPONSE

In this section we describe the methodology to obtain a space-time impulse response, mutuated from [5]. A device controlled by a stepper motor allows to place a cardioid microphone in n_{mic} positions, sampling the soundfield in an uniform fashion over a circumference of radius ρ_0, at angles ϕ_i, for $i = 1, \ldots, n_{\text{mic}}$; the sampled wave-



(a) $|h(t, \rho, \phi)|, t = 28.9$ ms. (b) $|h(t, \rho, \phi)|, t = 32.3$ ms. (c) $|h(t, \rho, \phi)|, t = 37$ ms. (d) $|h(t, \rho, \phi)|$ *at the marked point.*

Fig. 1: Space-time impulse response of the loudspeaker system described in Section 4, controlled to reproduce a virtual omnidirectional source at point $(-5\ \mathrm{m}, 0\ \mathrm{m})$. The absolute value of the soundfield pressure is showed for three time instants (1a, 1b, 1c) and at a given point (1d).

field is $p(t, \rho_0, \phi_i)$. For each position of the microphone a reproducible soundfield is emitted by the loudspeaker array. This soundfield is generated by a rendering system, like *Wave Field Synthesis* [1], *Geometric Rendering* [7] or others, which computes filters $h_m(t)$ for the M loudspeakers in the array.

The evaluation methodology proposed in this paper is based on the analysis of the space-time impulse response $h(t, \rho, \phi)$ of the rendering system in the circular region enclosed by the circumference of radius ρ_0. In order to go from $p(t, \rho_0, \phi_i)$, $i = 1, \ldots, n_{\text{mic}}$ to $h(t, \rho, \phi)$, first we extract the impulse responses $h(t, \rho_0, \phi_i)$ from $p(t, \rho_0, \phi_i)$. This is done using a *Maximum Length Sequence (MLS)* as input signal to the rendering system. The individual signals $p(t, \rho_0, \phi_i)$, obtained through the acquisition of the soundfield over the circumference, are then converted into the impulse responses $h(t, \rho_0, \phi_i)$ using the *Hadamard Transform*, as described in [8].

The estimation of the space-time impulse response $h(t, \rho, \phi)$ from $h(t, \rho_0, \phi_i)$, $i = 1, \ldots, n_{\text{mic}}$ is then accomplished by performing the interpolation and extrapolation procedure already presented in [5, 6] and here summarized for convenience. The space-time impulse response $h(t, \rho_0, \phi_i)$ is transformed in frequency domain to obtain $H(\omega, \rho_0, \phi_i)$, $i = 1, \ldots, n_{\text{mic}}$. From this, we compute the μ-th circular harmonic as [5]:

$$C_\mu(\omega) = \frac{H(\omega, \rho_0, \phi_\mu)}{\alpha J_\mu(k\rho_0) - \text{sgn}(\omega)(1-\alpha) J'_\mu(k\rho_0)}. \quad (1)$$

where $C_\mu(\omega)$ represents the μ-th circular harmonic as function of frequency ω; $J_\mu(\cdot)$ is the Bessel function of first kind and order μ, and $J'_\mu(\cdot)$ its derivative; $\alpha = 0.5$ for a cardioid microphone [5]; $k = \omega/c$ is the wave number, c being the sound speed. Since circular harmonic coefficients are independent from the radius ρ, we can obtain the soundfield $H(\omega, \rho, \phi)$ in the whole region of interest and thus its space-time impulse response $h(t, \rho, \phi)$:

$$\mathcal{FT}\{h(t, \rho, \phi)\} = H(\omega, \rho, \phi) = \sum_{\mu=-\infty}^{+\infty} C_\mu(\omega) J_\mu(k\rho) \mathrm{e}^{j\mu\phi}. \quad (2)$$

As a result, with the methodology reported here, we are able to obtain the space-time impulse response $h(t, \rho, \phi)$ of the rendering system starting from the filters $h_m(t)$, $m = 1, \ldots, M$.

An illustrative example of a measured space-time impulse response is depicted in Figure 1. The rendering system is composed of $M = 32$ loudspeakers disposed on a linear array of length 2.035 m. For this experiment, the system was installed in a rectangular dry room, the reverberation time T60 being approximately equal to 50 ms. The rendering system was simulating the presence of an omnidirectional virtual source located 2.5 m behind the array. In particular, Figures 1a-1c show two snapshots of the normalized absolute value of the propagating wavefronts for time instants t in the range $[28.9\ \text{ms} \sim 37\ \text{ms}]$, the time $t = 0$ s being the time at which the wavefront is emitted by the loudspeaker array. Notice the presence of secondary wavefronts immediately after and before the main one in Figure 1a and 1b, which produce the pre- and post-echoes effects, while Figure 1c clearly shows reflected wavefronts due to the presence of the walls in front and behind the array. Figure 1d shows the relationship between the impulse response $|h(t, \rho, \phi)|$, taken at time instant and the impulse response $|h(t, \rho_x, \phi_x)|$ considered in the marked point at coordinates $\rho_x = 0.5$ m, $\phi_x = 0$ rad.

3. PROPOSED METRICS

In this section we introduce a psychoacoustic-based analysis of the space-time impulse response $h(t, \rho, \phi)$ obtained in the previous section, to assess the impact of the pre- and post-echoes artifacts. First, we introduce a quantitative view of the *masking* effect in time domain; then we propose a metric that relies on the masking effect in evaluating the importance of pre- and post-echoes on the perception of quality of a rendered soundfield.

3.1. Masking in time domain

Masking is a perceptive effect by which a sound stimulus, the *maskee*, becomes inaudible due to the presence of a louder stimulus, the *masker*. Masking can occur when the two stimuli are simultaneous (which is the case *masking in frequency domain*, but it also occurs when the two stimuli are shifted in time. The presence of a louder stimulus (such as the main peak of our temporal impulse response) makes inaudible secondary stimuli coming after it (*post-masking*) or before it (*pre-masking*). While the effect of post-masking is more or less expected (it corresponds to a decay of the masker [9]), the effect of pre-masking appears to be non-causal. This effect is interpreted by psychoacoustics: the perception of a sound does not occur instantaneously, but requires a build-up time to be perceived. Thus the weaker stimulus cannot be perceived because of the arrival of a louder stimulus which is processed "faster" by the hearing system [9]; however, only events occurring in a time window shorter with respect to post-echoes are masked. Moreover, the pre-masking effect presents a relevant variation among different subjects. Nonetheless, pre-masking is exploited in perceptual audio coding systems to hide the presence of pre-echoes generated by blocking artifacts. To quantitatively describe pre-masking and post-masking, we rely on experimental data presented in [4] and reported in Figure 2. The figure shows the masking threshold (in dB) as a function of relative time position of the masker and the maskee.

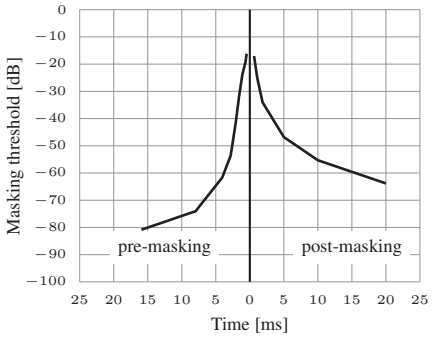


Fig. 2: Masking experiment, reprinted from [4]. The figure shows masking threshold in dB due to a Gaussian-shaped impulse occurring at time $t = 0$ ms as a function of the relative time position between the masker and a noisy maskee.

3.2. Pre-echoes and post-echoes evaluation metrics

Consider a temporal impulse response $h(t, \rho_x, \phi_x)$ obtained by sampling $h(t, \rho, \phi)$ at point $x = (\rho_x, \phi_x)$. We model this time-domain impulse response as a main peak (representing the wavefront propagating in the environment) and a series of secondary peaks coming before (pre-echoes) and after it (post-echoes). This secondary peaks represent correlated undesired wavefronts in the soundfield. Our goal is to assess the impact of these secondary peaks on the overall perceived quality. An analysis based on the *precedence effect* is not useful for our purposes, due to the short time-interval between main and secondary peaks.

The key idea of the proposed metrics is to evaluate the average power ratio between the main peak of the temporal impulse response and the tails preceding and following the main peak. According to the previously described masking curve, we consider only those peaks which exceed the masking threshold at the considered time instant. The proposed metric is analogous to the *Direct-to-Reverberant Ratio* presented in [10] and used to evaluate the impulse response of a reverberant room based on psychoacoustic criteria. We introduce *Direct-to-(pre)-Echo Ratio* DER^- and *Direct-to-(post)-Echo Ratio* DER^+ as:

$$\mathrm{DER}^- = 10\log_{10}\left[\frac{\dfrac{1}{\Delta\tau^+ + \Delta\tau^-}\displaystyle\sum_{t=\tau-\Delta\tau^-}^{\tau+\Delta\tau^+} h^2(t,\rho_x,\phi_x)}{\dfrac{1}{\tau-\Delta\tau^-}\displaystyle\sum_{t=0}^{\tau-\Delta\tau^-} \overline{h}^2(t,\rho_x,\phi_x)}\right], \quad (3)$$

$$\mathrm{DER}^+ = 10\log_{10}\left[\frac{\dfrac{1}{\Delta\tau^+ + \Delta\tau^-}\displaystyle\sum_{t=\tau-\Delta\tau^-}^{\tau+\Delta\tau^+} h^2(t,\rho_x,\phi_x)}{\dfrac{1}{\tau+\Delta\tau^+}\displaystyle\sum_{t=\tau+\Delta\tau^+}^{N} \overline{h}^2(t,\rho_x,\phi_x)}\right], \quad (4)$$

where τ is the measured time of arrival of the main echo; $\Delta\tau^-$ and $\Delta\tau^+$ are the time instants (obtained from the results reported in Figure 2) at which the threshold of audibility of the maskee is at -20 dB; $h(t, \rho_x, \phi_x)$ is the impulse response at the considered analysis point $x = (\rho_x, \phi_x)$; N is the length of the impulse response and finally $\overline{h}(t, \rho_x, \phi_x)$ is the part of the impulse response whose values exceed the threshold of audibility. Figure 3 shows on a sample impulse response the contributions in the computation of DER^+ and of DER^-.

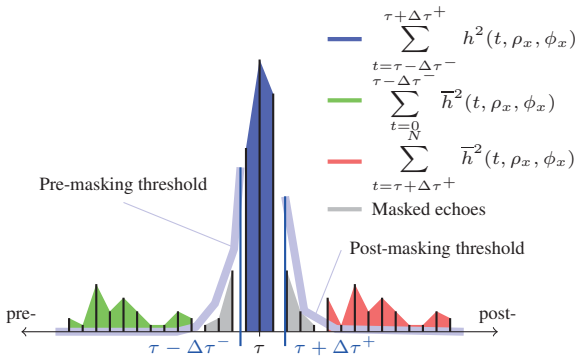


Fig. 3: Contributions in the computation of DER^+ and DER^-. The filled areas show the energy contribution of pre-echoes, post-echoes and main peak in the computation of the presented metrics.

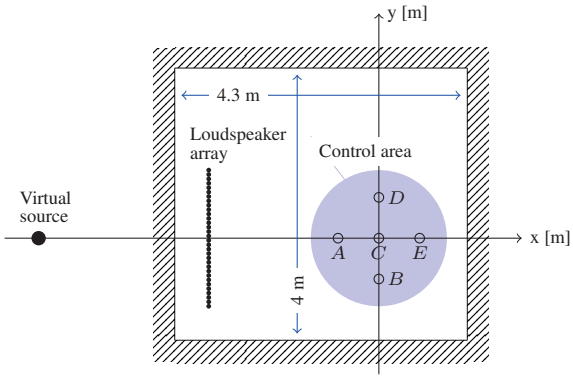


Fig. 4: Geometry of the room and installed sound reproduction system.

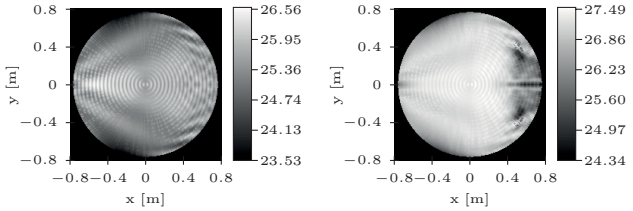
We expect that high values of DER^- and DER^+ are observed when pre-echoes and post-echoes are not relevant, respectively.

4. EXPERIMENTS

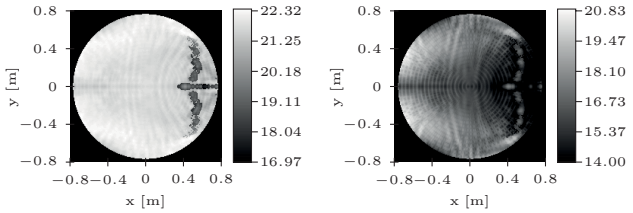
In this section we describe the experimental setup used to evaluate the proposed metrics. The sound reproduction setup is composed by a linear array of $M = 32$ equally spaced loudspeakers, whose aperture is $l = 2.035$ m; the mid-point of the array is placed 2.5 m far from the center of a circular listening area of radius 1 m; the geometry of the room and the installed sound reproduction system are depicted in Figure 4. This setup leads to a spatial Nyquist frequency $f_{\text{max}} = c/2d \approx 2.7$ kHz, where c is the speed of sound and d is the distance between adjacent loudspeakers. The virtual source, located at coordinates $x = -5$ m, $y = 0$ m, is omnidirectional.

Rendering engines usually define a solution $H_m(\omega)$ for loudspeaker filter coefficients in frequency domain. Hence, in order to use these filters in practical systems, we need to turn them into time-domain discrete filters $h_m(t)$. In order to implement *Geometric Rendering* technique, the frequency axis has been sampled at 135 points between 100 Hz and 3 kHz, which results in a frequency resolution $\Delta f \approx 21.5$ Hz at a sampling frequency of $f_s = 44.1$ kHz.

The excitation signal adopted in our procedure is a pseudo-random *Maximum Length Sequence (MLS)* of order 17, which results in an excitation signal $s(t)$ of length 131071 samples $= 2.9721$ s.



(a) DER^-, *simulated*. (b) DER^+, *simulated*.



(c) DER^-, *measured*. (d) DER^+, *measured*.

Fig. 5: Values of the metrics DER^- and DER^+ over the measurement area for both simulated (5a, 5b) and measured (5c, 5d) space-time impulse responses.

In order to sample the wavefield on a set of points, a sequential virtual circular array of radius $\rho_0 = 0.85$ m has been adopted, using a condenser cardioid microphone *AKG C1000s*. The angular resolution is fixed to 2° which leads to 180 acquisition points on the circumference. The extrapolation of the wavefield (i.e. obtaining $h(\omega, \rho, \phi)$ starting from $h(\omega, \rho_0, \phi_i)$) is accomplished by sampling the frequency axis at 4096 frequencies between 0 Hz and $f_s/2 = 22.05$ kHz and then applying the procedure described in Section 2.

Figure 5 shows the values of the metrics DER^- and DER^+ over the whole measurement area for simulated and measured space-time impulse responses, obtained with the setup described in this section; a different scale for each picture has been adopted to emphasize the different spatial behavior of the values. Table 1 reports values of DER^- and DER^+ for the same impulse responses on the five points (from A to E) marked in Fig.4. From the analysis of simulative results notice that values of DER^- (Figure 5a) are approximately constant over the whole measurement area; hence, the impact of pre-echoes generated by the solely rendering technique is not dependent on the position of the listener. On the other hand, values of DER^+ (Figure 5b) exhibit a slight degradation at points far from the loudspeaker array; hence, the rendering technique introduces post-echoes artifacts whose perception is most impairing in a region far from the loudspeaker array. However, considering also data reported in Table 1, notice that, among the time-domain artifacts introduced by the rendering technique, pre-echoes are more impairing than post-echoes. Focusing on measured results, we notice that values of DER^- (Figure 5c) exhibit the same behavior of the corresponding simulative results (Figure 5a); hence pre-echoes artifacts due to non-idealities of the real environment impair the perception of quality, but independently on the listener position. Instead, observing values of DER^+ (Figure 5d) we notice again a behavior dependent on the listener position (the degradation is more relevant far from the loudspeaker array). Moreover, artifacts due to the real environment (above all the reflective behavior of walls) considerably affect the perception of quality, so that post-echoes are the most important time-domain artifacts to be considered in real-world soundfield rendering applications.

Point	**Simulated**		**Measured**	
	DER^- dB	DER^+ dB	DER^- dB	DER^+ dB
$A(-0.6, 0)$	25.87	27	19.73	18.3
$B(0, -0.6)$	24.99	27	21.25	18.64
$C(0, 0)$	25.7	27.3	20.7	19.16
$D(0, 0.6)$	25	27	21.41	18.11
$E(0.6, 0)$	25.71	26.12	20.63	16.82

Table 1: Pre-echoes and post-echoes average values at points A, B, C, D and E.

5. CONCLUSIONS

In this paper we have proposed an analysis methodology to assess the impact of pre-echoes and post-echoes on the perception of quality of a rendered soundfield. A well known methodology has been exploited in order to measure the space-time impulse response of the rendering system. Simulative and experimental results have been included, which show that a great influence on the quality of the rendered soundfield comes from the acoustic properties of the room in which the rendering system is operating. We are currently conducting a set of listening tests, in order to verify the matching between the proposed theoretical prediction and the perception of the listener.

6. REFERENCES

[1] A. Berkhout, D. D. Vries, and P. Vogel, "Acoustic control by wave field synthesis," *J. Acoust. Soc. Am.*, vol. 93, pp. 2764 – 2778, 1993.

[2] J. Ahrens, H. Wierstorf, and S. Spors, "Comparison of higher order ambisonics and wave field synthesis with respect to spatial discretization artifacts in time domain," in *Audio Engineering Society Conference: 40th International Conference: Spatial Audio: Sense the Sound of Space*, 10 2010.

[3] M. Geier, H. Wierstorf, J. Ahrens, I. Wechsung, A. Raake, and S. Spors, "Perceptual evaluation of focused sources in wave field synthesis," in *Audio Engineering Society Convention 128*, 5 2010.

[4] M. Kahrs and K. Brandenburg, Eds., *Applications of Digital Signal Processing To Audio and Acoustics*, Kluwer Academic Publishers, 2002.

[5] A. Kuntz and R. Rabenstein, "Cardioid pattern optimization for a virtual circular microphone array," in *EAA Symposium on Auralization*, Helsinki, June 2009, pp. 1–4.

[6] A. Canclini, P. Annibale, F. Antonacci, A. Sarti, R. Rabenstein, and S. Tubaro, "A methodology for evaluating the accuracy of wave field rendering techniques," in *ICASSP*, 2011, pp. 69–72.

[7] F. Antonacci, A. Calatroni, A. Canclini, A. Galbiati, A. Sarti, and S. Tubaro, "Soundfield rendering with loudspeaker arrays through multiple beam shaping," in *Applications of Signal Processing to Audio and Acoustics, 2009. WASPAA '09. IEEE Workshop on*, Oct. 2009, pp. 313–316.

[8] G. Stan, J. Embrechts, and D. Archambeau, "Comparison of different impulse response measurement techniques," *J. Aud. Eng. Soc.*, vol. vol. 50, no. 4, pp. 249–262, 2002.

[9] H. Zwicker and E. Fastl, *Psychoacoustics - Facts and Models*, Springer, 2007.

[10] M. Triki and D. T. M. Slock, "Iterated delay and predict equalization for blind speech dereverberation," in *Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, sept 2006.