WAVE-DOMAIN ADAPTIVE FILTERING: ACOUSTIC ECHO CANCELLATION FOR FULL-DUPLEX SYSTEMS BASED ON WAVE-FIELD SYNTHESIS

Herbert Buchner, Sascha Spors, Walter Kellermann

Multimedia Communications and Signal Processing
University of Erlangen-Nuremberg
Cauerstr. 7, 91058 Erlangen, Germany
{buchner,spors,wk}@LNT.de

ABSTRACT

For high-quality multimedia communication systems such as teleconferencing or virtual reality applications, multichannel sound reproduction is highly desirable. While progress has been made in stereo and multichannel acoustic echo cancellation (MC AEC) in recent years, the corresponding sound reproduction systems still imply a restrained listening area ('sweet spot'). A volume solution for spatial sound in a large listening area is offered by wave field synthesis (WFS) or by ambisonics, where arrays of loudspeakers generate a prespecified sound field. However, before this new technique can be utilized for full-duplex systems, an efficient solution to the MC AEC problem has to be found. This paper presents a novel approach that extends the current state of the art of MC AEC and transform-domain adaptive filtering by reconciling the flexibility of adaptive filtering and the underlying physics of acoustic waves in a systematic and efficient way. To achieve this, the new framework of wave-domain adaptive filtering (WDAF) exploits the spatial information provided by densely sampled contours for both, recording and reproduction. Experimental results with a 32-channel AEC verify the concept for both, simulated and actually measured room acoustics.

1. INTRODUCTION

To enhance sound realism in multimedia communication systems, such as teleconferencing or tele-teaching (especially of music), and to create a three-dimensional illusion of sound sources positioned in a virtual acoustical environment, multichannel sound reproduction is necessary. However, advanced loudspeaker-based approaches, like the 3/2-Surround format, still rely on a restrained listening area ('sweet spot'). A volume solution for a large listening space is offered by the Wave Field Synthesis (WFS) method [1], where arrays of a large number of individually driven loudspeakers generate a prespecified sound field. On the recording side of the two-way systems, the use of microphone arrays is an effective approach to separate desired and undesired sources in the listening environment, and to cope with reverberation of the recorded signal. Figure 1 shows the general setup for such a system.

However, before full-duplex communication can be deployed, the problem of acoustic feedback from the P loudspeakers to the Q microphones has to be addressed. Acoustic echo cancellation (AEC) for the resulting $P \cdot Q$ echo paths (where P may lie between 20 and several hundred, and Q may be on the order of 10 to 1000) poses the major problem in this MIMO (multi-input and multi-output) context [2, 3], and satisfactory general solutions for

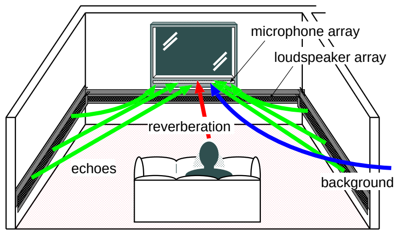

Fig. 1. General setup for multimedia communication.

AEC in conjunction with large loudspeaker arrays have not been presented yet.

In this paper, we propose and study a novel approach for spatio-temporal transform-domain adaptive filtering, called wave-domain adaptive filtering (WDAF) in the following. This concept is suitable for spatial audio reproduction systems like wave field synthesis with an arbitrarily high number of channels. Although we refer here to two-dimensional wave fields and WFS, the proposed technique can also be applied to ambisonics and extended to 3D fields.

2. WAVE FIELD SYNTHESIS

Sound reproduction by wave field synthesis (WFS) using loudspeaker arrays is based on Huygens principle. It states that any point of a wave front of a propagating sound pressure wave $p(\mathbf{r}, t)$ at any instant of time conforms to the envelope of spherical waves emanating from every point on the wave front at the prior instant. This principle can be used to synthesize acoustical wavefronts of arbitrary shape. A mathematical formulation of it is given by the Kirchhoff-Helmholtz integrals (e.g., [1, 4]) which can be derived from the acoustic wave equation (given here for lossless media)

$$\nabla^2 p(\mathbf{r}, t) - \frac{1}{c^2}\frac{\partial^2 p(\mathbf{r}, t)}{\partial t^2} = 0 \tag{1}$$

and Newton's second law

$$-\nabla p(\mathbf{r}, t) = \rho \frac{\partial \mathbf{v}(\mathbf{r}, t)}{\partial t}, \tag{2}$$

where c denotes the velocity of sound, ρ is the density of the medium, and $\mathbf{v}(\mathbf{r}, t)$ is the particle velocity. Since we assume two-dimensional wave fields, we choose polar coordinates (r, θ) throughout this paper. Using the second theorem of Green, applied to a curve C enclosing an region $\mathcal{S}$, we obtain from (1) and

(2) the 2D forward Kirchhoff-Helmholtz integral

$$\underline{p}^{(2)}(\mathbf{r},\omega) = \frac{-jk}{4}\oint_C \Big\{\underline{p}(\mathbf{r}',\omega)\cos\varphi H_1^{(2)}(k\Delta r) + j\rho c \underline{v}_{\mathrm{n}}(\mathbf{r}',\omega)H_0^{(2)}(k\Delta r)\Big\}\, d\ell \quad (3)$$

and the 2D inverse Kirchhoff-Helmholtz integral

$$\underline{p}^{(1)}(\mathbf{r},\omega) = \frac{-jk}{4}\oint_C \Big\{\underline{p}(\mathbf{r}',\omega)\cos\varphi H_1^{(1)}(k\Delta r) + j\rho c \underline{v}_{\mathrm{n}}(\mathbf{r}',\omega)H_0^{(1)}(k\Delta r)\Big\}\, d\ell, \quad (4)$$

where $\Delta r = ||\mathbf{r}-\mathbf{r}'||$ and $k = \omega/c$. The total wave field is given by the sum of the incoming and outgoing contributions w.r.t. $\mathcal{S}$:

$$\underline{p}(\mathbf{r},\omega) = \underline{p}^{(1)}(\mathbf{r},\omega) + \underline{p}^{(2)}(\mathbf{r},\omega). \quad (5)$$

$H_n^{(1)}$ and $H_n^{(2)}$ are the Hankel functions of the first and second kind, respectively, which are the fundamental solutions of the wave equation in polar coordinates. v_{n} denotes the radial component of $\mathbf{v}$. All quantities in the temporal frequency domain are underlined.

The 2D Kirchhoff-Helmholtz integrals (3) and (4) state that at any listening point within the source-free listening area the sound pressure can be calculated if both the sound pressure and its gradient are known on the curve C enclosing this area. For practical implementations in 2D sound fields the acoustic sources on the closed contour are realized by loudspeakers on discrete positions. Note that (3) and (4) can analogously be applied for wave field analysis using a microphone array consisting of pressure and pressure gradient microphones. The spatial sampling on the contour defines the aliasing frequencies. While microphone spacings are usually designed for a wide frequency range, lower aliasing frequencies may be tolerated on the reproduction side as the human auditory system seems not to be very sensitive to these aliasing artifacts above approximately 1kHz.

3. MC AEC - STATE OF THE ART AND CONVENTIONAL MULTICHANNEL ADAPTIVE ALGORITHMS

Classical AEC applications are hands-free telephony or teleconference systems, where most of them are still based on monaural sound reproduction. Only recently, first stereophonic prototypes appeared [5], [6], and lately, it has become possible to extend the system to the multichannel case (for 5-channel surround sound see, e.g., [7]). The concept of this frequency-domain framework will here be extended for WFS in Sect. 4.

3.1. Multichannel Acoustic Echo Cancellation

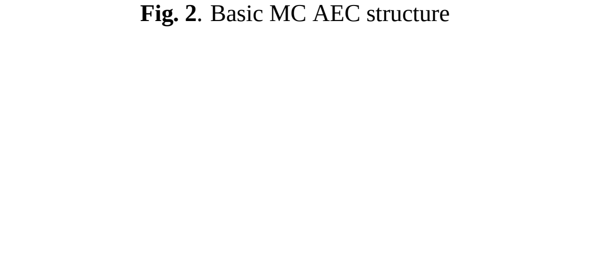

Fig. 2. Basic MC AEC structure

The fundamental idea of any P-channel AEC structure (Fig. 2) is to use adaptive FIR filters of length L with impulse response vectors $\hat{\mathbf{h}}_i(n)$, $i = 1,\ldots,P$ that identify the truncated (generally time-varying) echo path impulse responses $\mathbf{h}_i(n)$. The filters $\hat{\mathbf{h}}_i(n)$ are stimulated by the loudspeaker signals $x_i(n)$ and, then, the resulting echo estimates $\hat{y}_i(n)$ are subtracted from the microphone signal $y(n)$ to cancel the echoes. For multiple microphones, each of them is considered separately in this way. The filter length L may be on the order of several thousand.

The specific problems of MC AEC include all those known for mono AEC, but in addition to that, MC AEC often has to cope with high correlation of the different loudspeaker signals [2, 3]. The correlation results from the fact that the signals are almost always derived from common sound sources in the transmission room, as shown in Fig. 2. The optimization problem therefore often leads to a severely ill-conditioned normal equation to be solved for the $P \cdot L$ filter coefficients. Therefore, sophisticated adaptation algorithms taking the cross-correlation into account are necessary for MC AEC [3].

3.2. Multichannel Adaptive Filtering

For various ill-conditioned optimization problems in adaptive signal processing, such as MC AEC, the recursive least-squares (RLS) algorithm is known to be the optimum choice in terms of convergence speed as it exhibits properties that are independent of the eigenvalue spread [8]. The update equation of the multichannel RLS (MC RLS) algorithm reads

$$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \mathbf{R}_{\mathrm{xx}}^{-1}(n)\mathbf{x}(n)e(n), \quad (6)$$

where $\hat{\mathbf{h}}(n)$ is the multichannel coefficient vector obtained by concatenating the impulse response vectors $\hat{\mathbf{h}}_i(n)$ of all input channels, $e(n) = y(n) - \hat{y}(n)$ is the current residual error vector between the echoes and the echo replicas. The length-PL vector $\mathbf{x}(n)$ is a concatenation of the input signal vectors containing the L most recent input samples of each channel. The correlation matrix $\mathbf{R_{xx}}$ takes all auto-correlations within, and - most importantly for MC AEC - all cross-correlations between the input channels into account (see upper left corner of Fig. 3). However, the major problems of RLS algorithms are the very high computational complexity (mainly due to the large matrix inversion) and potential numerical instabilities which often limit the actual performance in practice.

An efficient and popular alternative to time-domain algorithms are frequency-domain adaptive filtering (FDAF) algorithms [9]. In FDAF, the adaptive filters are updated in a block-by-block fashion, using the fast Fourier transform (FFT) as a powerful vehicle. Recently, the FDAF approach has been extended to the multichannel case (MC FDAF) by a mathematically rigorous derivation based on a weighted least-squares criterion [10, 7]. It has been shown that there is a generic wideband frequency-domain algorithm which is equivalent to the RLS algorithm. As a result of this approach, the arithmetic complexity of multichannel algorithms can be significantly reduced compared to time-domain adaptive algorithms while the desirable RLS-like properties and the basic structure of (6) are maintained by an inherent block-diagonalization of the correlation matrix as shown in the second column of Fig. 3. This allows to perform the matrix inversion in (6) in a frequency-bin selective way using only small and better conditioned $P \times P$ matrices $\mathbf{S}_{\mathrm{xx}}^{(\nu)}$ in the bins $\nu = 0,\ldots,2L-1$. Note that all cross-correlations are still fully taken into account by this approach.

4. THE NOVEL APPROACH: WAVE-DOMAIN ADAPTIVE FILTERING

With the dramatically increased number of highly correlated loudspeaker channels in WFS-based systems, even the matrices $\mathbf{S}_{\mathrm{xx}}^{(\nu)}$ become large and ill-conditioned so that current algorithms cannot be used. In this section we extend the conventional concept of MC FDAF by a more detailed consideration of the spatial dimensions and by exploitation of wave physics as shown in Sect. 2.

4.1. Basic Concept

From a physical point of view, the nice properties of FDAF result from the orthogonality property of the DFT basis functions, i.e., the complex exponentials which also separate the temporal dimension of the wave equation (1). Therefore, it is desirable to find a suitable spatio-temporal transform domain based on orthogonal basis functions that allow not only a decomposition among the temporal frequencies as in MC FDAF, but also a spatial decomposition as illustrated by the third column of Fig. 3. These basis functions must also fulfill (1). In the next subsection we will propose a suitable transform domain. Performing the adaptive filtering in

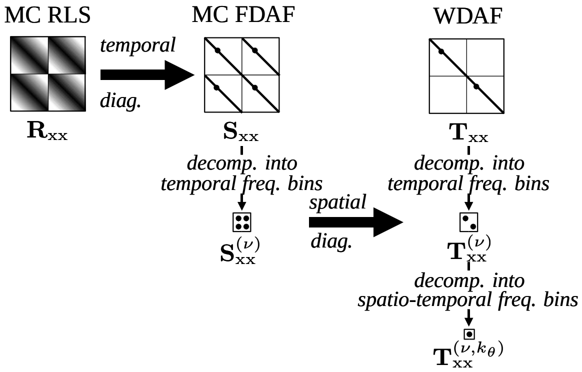

Fig. 3. Illustration of the WDAF concept and its relation to conventional algorithms.

a spatio-temporal transform domain requires spatial sampling on both, the input and the output of the system, i.e., in contrast to conventional MC FDAF, not only all loudspeaker signals, but also all microphone signals are simultaneously taken into account for the adaptive processing. Moreover, due to the orthogonality between the spatial components in the transform domain, there are no cross-channels to be adapted. This leads to the general setup of WDAF-based AEC shown in Fig. 4. Due to the decoupling of

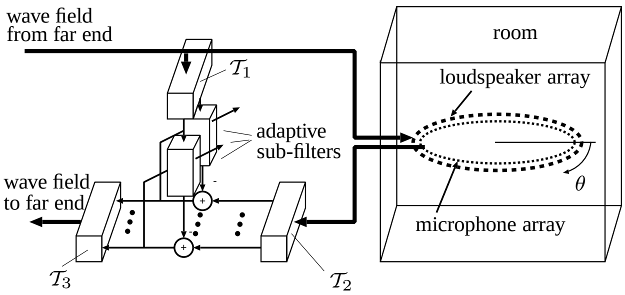

Fig. 4. Setup for proposed AEC in the wave domain.

the channels, not only the convergence properties are improved but also the computational complexity is reduced dramatically. Let us assume $Q = P$ microphone channels. Instead of P^2 filters in the conventional approach, we only need to adapt P filters in the transform domain. By additionally taking into account the symmetry property of spatial frequency components, the number is further reduced to $P/2$. Thus, for a typical system with $P = 32$, the number of channels is reduced from 1024 to 16.

4.2. Transformations and Adaptive Filtering

In this section we introduce suitable transformations $\mathcal{T}_1$, $\mathcal{T}_2$, $\mathcal{T}_3$ shown in Fig. 4. Note that in general there are many possible spatial transformations depending on the choice of the coordinate system. A first approach to obtain the desired decoupling would be to simply perform spatial Fourier transforms analogously to the temporal dimension. This corresponds to a decomposition into plane waves [4] which is known to be a flexible format for auralization purposes [11]. However, in this case we would need loudspeakers and microphones at each point of the listening area which is not suitable. Therefore, plane wave decompositions taking into account the Kirchhoff-Helmholtz integrals are desirable. These transformations depend on the array geometries. Circular arrays are known to show a particularly good performance in wave field analysis [11], and lead to an efficient WDAF solution. For clarity, we will give the transforms in their continuous formulation in the following as they follow from the physical equations (2)-(4) [11]. For the realization, temporal and spatial sampling are necessary according to the desired aliasing frequency. For transform $\mathcal{T}_1$ we obtain the following plane wave decomposition of the incident wave field, emitted by the loudspeaker array with radius R:

$$\begin{aligned}\underline{\tilde{x}}^{(1)}(k_\theta,\omega) &= \frac{j^{1-k_\theta}}{D_R(k_\theta,\omega)}\Big\{H^{(2)'}_{k_\theta}(kR)\underline{\tilde{p}}_{\mathrm{x}}(k_\theta,\omega) \\ &\quad -H^{(2)}_{k_\theta}(kR)j\rho c\underline{\tilde{v}}_{\mathrm{x,n}}(k_\theta,\omega)\Big\},\end{aligned} \tag{7}$$

$$\begin{aligned}\underline{\tilde{x}}^{(2)}(k_\theta,\omega) &= \frac{-j^{1+k_\theta}}{D_R(k_\theta,\omega)}\Big\{H^{(1)'}_{k_\theta}(kR)\underline{\tilde{p}}_{\mathrm{x}}(k_\theta,\omega) \\ &\quad -H^{(1)}_{k_\theta}(kR)j\rho c\underline{\tilde{v}}_{\mathrm{x,n}}(k_\theta,\omega)\Big\},\end{aligned} \tag{8}$$

$$D_R(k_\theta,\omega) = H^{(1)}_{k_\theta}(kR)H^{(2)'}_{k_\theta}(kR) - H^{(2)}_{k_\theta}(kR)H^{(1)'}_{k_\theta}(kR). \tag{9}$$

$H^{(\cdot)'}_{k_\theta}$ denotes the derivative of the respective Hankel function with the angular wave number k_θ. Underlined quantities with a tilde denote spatio-temporal frequency components, e.g.,

$$\underline{\tilde{p}}_{\mathrm{x}}(k_\theta,\omega) = \frac{1}{2\pi}\int_0^{2\pi} \underline{p}_{\mathrm{x}}(\theta,\omega)e^{-jk_\theta\theta}d\theta. \tag{10}$$

Analogously to (7) and (8) the reflected plane wave components $\underline{\tilde{y}}^{(1)}(k_\theta,\omega)$ and $\underline{\tilde{y}}^{(2)}(k_\theta,\omega)$ in the receiving room are obtained by $\mathcal{T}_2$ from $\underline{\tilde{p}}_{\mathrm{y}}(k_\theta,\omega)$ and $\underline{\tilde{v}}_{\mathrm{y,n}}(k_\theta,\omega)$ using the pressure and pressure gradient microphone elements. On the loudspeaker side an additional extrapolation assuming free field propagation of each loudspeaker signal to the microphone positions is necessary within $\mathcal{T}_1$ prior to using (7) and (8) in order to obtain p_{x} and $v_{\mathrm{x,n}}$ of the incident waves on the microphone positions.

Adaptive filtering is then carried out for each spatio-temporal frequency bin. This corresponds to FIR filters for each spatial frequency bin. Note that efficient conventional single-channel FDAF algorithms can directly be applied to each sub-filter in Fig. 4. Since the plane wave representation after the AEC is independent of the array geometries, the clean plane wave components can either be sent to the far end directly, or they can be used to synthesize the total spatio-temporal wave field using an extrapolation $\mathcal{T}_3$ of the wave field [4].

5. SIMULATIONS

To verify the proposed concept, we study two different scenarios.

5.1. Case Study 1: Simulated Acoustic Environment

In the first simulation, we consider an idealized 2D scenario with two circular 32-element arrays. The radius of the loudspeaker array is 1m and the radius of the microphone array (which is located concentric inside the loudspeaker array) is 0.98m. Using wave field synthesis a virtual point source (music signal) has been positioned at 2m distance from the array center at an angle of $\theta = 90^o$. After 5sec this angle has been changed to $\theta = 101.3^o$ to verify the quality of the estimated room parameters [2, 3]. All signals were downsampled according to the aliasing frequency of $f_{\rm al} \approx 900$Hz. The wall at one side of the array at $\theta = 270^o$ is assumed to be reflective (calculation using the mirror-image method), the other walls of the room are ideal absorbers. For the adaptation of the parameters, FDAF algorithms (filter length 1024 each) with an overlap factor 32 after [7] were used. Figure 5 shows the so-called echo return loss enhancement (ERLE), i.e., the attenuation of the echoes. Considering the low number of data samples due to the

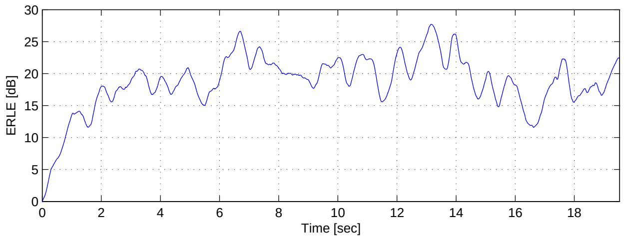

Fig. 5. Simulation result for case 1.

low sampling rate, the convergence is extremely fast. Due to the change of the source position after 5sec there is only a slight temporary decrease of ERLE.

5.2. Case Study 2: Measurements in a Real Room

For the simulations using measured data from a real room, we used 24 loudspeakers and a 32-element circular microphone array (the recording was done by one rotating sound field microphone mounted on a step motor) as shown in Fig. 6. The ceiling is not damped. Again, a virtual point source was placed by WFS. Figure 7 shows the $ERLE$ convergence using the same overlap factor 32 as above. Stable adaptation and sufficient attenuation levels can be achieved. Although the convergence speed is somewhat slower than in the simplified case 1 (which may result from the reflections on the ceiling that are not taken explicitly into account in the 2D approach), it is well comparable to conventional single-channel AECs w.r.t. the number of available data samples. Further experiments have shown that this data sparseness can largely be compensated by an increased overlap factor so that a practical solution is obtained.

6. CONCLUSIONS

A highly efficient novel approach to adaptive MIMO filtering in the wave domain has been proposed which enables AEC for WFS-based systems. The new framework is also applicable to several other challenging adaptive filtering problems and applications.

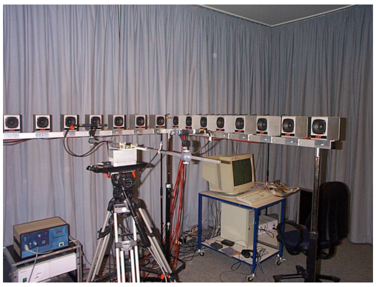

Fig. 6. Measurements for case study 2.

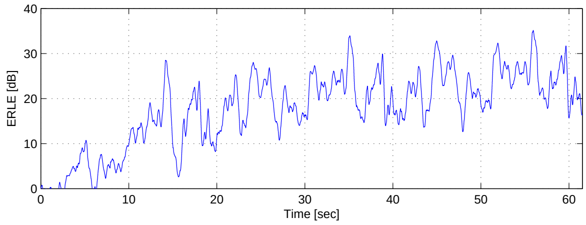

Fig. 7. Simulation result for case 2.

7. REFERENCES

[1] A.J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *Journal of the Acoustic Society of America*, vol. 93, no. 5, pp. 2764–2778, May 1993.

[2] M. M. Sondhi and D. R. Morgan, "Stereophonic Acoustic Echo Cancellation - An Overview of the Fundamental Problem," *IEEE SP Lett.*, Vol.2, No.8, Aug. 1995, pp. 148-151.

[3] J. Benesty, D.R. Morgan, and M.M. Sondhi, "A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation," *IEEE Trans. on Speech and Audio Processing*, vol. 6, no.2, March 1998.

[4] A.J. Berkhout, *Applied Seismic Wave Theory*, Elsevier, 1987.

[5] V. Fischer et al., "A Software Stereo Acoustic Echo Canceler for Microsoft Windows," in *Proc. IWAENC*, Darmstadt, Germany, pp. 87-90, Sept. 2001.

[6] H. Buchner, W. Herbordt, and W. Kellermann, "An Efficient Combination of Multichannel Acoustic Echo Cancellation With a Beamforming Microphone Array," in *Proc. HSC*, Kyoto, Japan, pp. 55-58, April 2001.

[7] H. Buchner, J. Benesty, and W. Kellermann, "Multichannel Frequency-Domain Adaptive Algorithms with Application to Acoustic Echo Cancellation," in J.Benesty and Y.Huang (eds.), *Adaptive signal processing*, Springer, Berlin, Jan. 2003.

[8] S. Haykin, *Adaptive Filter Theory*, 3rd ed., Prentice Hall Inc., Englewood Cliffs, NJ, 1996.

[9] J.J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE SP Magazine*, pp. 14-37, Jan. 1992

[10] J. Benesty, A. Gilloire, and Y. Grenier, "A frequency-domain stereophonic acoustic echo canceler exploiting the coherence between the channels," *J. Acoust. Soc. Am.*, vol. 106, pp. L30-L35, Sept. 1999.

[11] E. Hulsebos, D. de Vries, and E. Bourdillat, "Improved microphone array configurations for auralization of sound fields by Wave Field Synthesis," *110th Conv. of the AES*, May 2001.