Simultaneous Channel Estimation and Joint Time-Frequency Domain Crosstalk Cancellation in Multichannel Personal Audio Systems

Harsh Tataria*, Paul D. Teal*, Mark Poletti†, Terence Betlehem†
*School of Engineering and Computer Science, Victoria University of Wellington, Wellington, New Zealand.
†Callaghan Innovation, Lower Hutt, New Zealand.

15 February 2014

Abstract

In this paper, we present two important contributions. The first is demonstration of the use of subliminal levels of pseudo-random noise to enhance channel estimation, and the second is a joint time and frequency domain algorithm for multichannel inversion. An adaptive system is presented where the acoustic channel is accurately estimated and utilized. In this implementation, maximum length sequences in the form of pseudo-random noise are superimposed on the source signals to aid the estimation quality of the acoustic channels. Upon estimation, crosstalk cancellation filters are designed using a time and frequency domain technique which uses a window to achieve more efficient and effective cancellation of crosstalk. For a 3×2 crosstalk system, the presented results show the improvement in channel estimation quality when low levels of maximum length sequences are superimposed on the source signals. On average -25dB of crosstalk cancellation is achieved.

This document is similar to: H. Tataria, P.D. Teal, M. Poletti, and T. Betlehem. Simultaneous channel estimation and joint time-frequency domain crosstalk cancellation in multichannel personal audio systems. In IEEE Workshop on Statistical Signal Processing, Gold Coast, June 2014.

1 Introduction

Crosstalk cancellation is applicable to many audio situations. In state-of-the-art personal audio systems, crosstalk cancellation systems aim to provide independent transmission of speech or music to multiple listeners in the same listening space [1]. Crosstalk cancellation systems are designed to accurately reproduce sound at particular points in the acoustic space of interest, whilst suppressing the acoustical leakage from other sound sources. To design crosstalk cancellation filters, the impulse response (or the transfer function) from the loudspeakers to the listening position(s) must be known [12, 8, 13]. In this paper, we propose an adaptive system which allows the channel impulse responses to be implicitly characterized (estimated) by adding low levels of maximum length sequences (MLS) to source signals, while the channels are in use. Upon estimating the acoustic channels, we perform crosstalk cancellation via the proposed time and frequency domain technique, inverting the estimated impulse responses. In doing so, we make use of a window function to permit low levels of echo in the estimated impulse responses.

Due to the low cross-correlation and near to impulsive auto-correlation properties of MLS, impulse response measurement techniques with MLS have been extensively studied and utilized [3, 15, 16]. These properties have enabled their use in multisource multireceiver systems which must accomplish simultaneous measurements in a limited time manner [18]. MLS have also been used in multichannel acoustic echo cancellation systems to solve the non-uniqueness problem for perfect system identification. There the fundamental problem is that the multiple channels may carry linearly related signals which in turn may make the multichannel equations unsolvable. It was shown in [2] that a solution to this problem was to reduce the cross-correlation between different loudspeaker signals by adding incoherent noise, in order to get well behaved estimates of the echo paths. Other methods have been proposed to do this including different preprocessing for each loudspeaker channel, such as nonlinear preprocessing [10], time-varying prefiltering [4] and resampling [17]. However, the challenge here is to reduce the coherence sufficiently without affecting the audio perception and quality. In addition to the above, techniques have also been proposed to estimate the room impulse responses in the presence of an audience with music signals by using complex modulation transfer functions with anechoic and reverberant envelopes [14]. However, results from these methods have shown that sufficient estimation accuracy cannot be achieved to gain information regarding the room parameters. The approach that we are proposing allows the channels to be simultaneously characterized with the addition of subliminal levels of MLS to source signals. This enhances channel estimation quality whilst keeping the audio perception and quality unchanged.

Many algorithms which obtain accurate crosstalk cancellation with two or more loudspeakers are also described in the current literature. Most of the algorithms formulate the problem as an inverse filter with the least-squares optimization criteria minimising the resultant crosstalk [12, 11, 7]. Generalized crosstalk cancellation and equalization using multiple loudspeakers for multiple listeners is proposed in [5], where, as well as formulating the crosstalk problem in a least-squares sense, the exact solution and the minimum-norm solutions are derived. However, the performance of these techniques rapidly degrades in the presence of errors in the estimated channel and noise. A more robust approach was proposed by [6] which opts for multichannel inversion if the channel estimation error is low and beamforming if the channel estimation error is high. It describes a tradeoff between the quality of channel inversion and beamforming. In this paper we combine adaptivity with a windowed time and frequency domain approach in performing multichannel inversion. The rest of the paper is structured as follows. Section 2 details the proposed joint time and frequency domain crosstalk cancellation technique. In section 3, we describe the proposed channel estimation technique with MLS. Section 4 presents results of both the proposed techniques for real channels and Section 5 concludes the paper.

2 Joint Time-Frequency Domain Crosstalk Cancellation

The goal of crosstalk cancellation in personal audio systems can be simplified to independent delivery of S source signals to $M = S$ microphones or equivalent pressure matching points. Independent sound streams are delivered via the use of L loudspeakers. Typically, $L \geq S$ to ensure at least one extra degree of freedom is achieved and the probability of deep nulls in the acoustic channels between each loudspeaker and microphone is minimized. In order to successfully cancel crosstalk, crosstalk cancellation filters are required. Crosstalk cancellation filters are designed considering the effects of reverberation and are traditionally based on inversion of the minimum-phase component of the estimated channel impulse responses. However, here we propose an adaptive approach to crosstalk cancellation where the crosstalk filters are designed using alternation between the time and frequency domains.

Let N_h denote the length of the crosstalk cancellation filter impulse response from the sth source to the lth loudspeaker $\mathbf{h}_{ls}$. Let N_c denote the length of the sampled acoustic channel impulse response $\mathbf{c}_{ml}$ from the lth loudspeaker to the mth microphone. The overall

impulse response $\mathbf{r}_{ms}$ from the sth source to the mth microphone is given by

$$\mathbf{r}_{ms} = \sum_{l=1}^{L} \mathbf{c}_{ml} * \mathbf{h}_{ls} \tag{1}$$

or in the frequency domain as

$$\mathbf{R}_{ms}(\omega) = \sum_{l=1}^{L} \mathbf{C}_{ml}(\omega)\mathbf{H}_{ls}(\omega). \tag{2}$$

The length of the overall impulse response is given by $N_\mathrm{r} = N_\mathrm{c} + N_\mathrm{h} - 1$. We design $\mathbf{h}_{ls}$ such that the crosstalk signal paths are attenuated as much as possible. Ideally, they should be given by $\mathbf{R}_{ms}(\omega) = 0$ when $m \neq s$. The ideal delivered transfer functions for the direct signal paths when $m = s$ must have $|\mathbf{R}_{ss}(\omega)| = 1$. By appropriately stacking the values for the M microphones and L loudspeakers, we obtain

$$\mathbf{R}(\omega) = \mathbf{C}(\omega)\mathbf{H}(\omega). \tag{3}$$

The equations in (1) and (2) are solved by alternation between the time and frequency domains. The goal in the frequency domain is to achieve in $\mathbf{R}_{ms}$ low crosstalk responses (i.e., for $m \neq s$) and flat direct responses (i.e., for $m = s$). This is successively achieved by scaling the frequency domain responses towards the desired amplitudes (0 or 1) without altering the phase of the signals. The goal in the time-domain is to achieve filters in $\mathbf{h}_{ls}$ of length no longer than N_h (where the DFT length is $N_\mathrm{r} \geq N_\mathrm{h}$), and as close to minimum-phase as possible. The requirement that the filters contain $N_\mathrm{r} - N_\mathrm{h} = N_\mathrm{c} - 1$ consecutive zero entries is often ignored in the frequency domain filter design literature. This is gradually achieved by multiplying the estimated impulse responses with the designed window. The window function ensures that the beginning and the end of the impulse responses are zero. A short period of early reverberation is permitted by assigning a magnitude of 1 to a section of the window function as it contributes beneficially to the delivery of acoustic energy without detracting from the listening experience. Setting the magnitude of this weight to 1 also ensures that the early attack of impulses in the estimated impulse responses are retained. Thus, the weight vector of N_r2 samples is denoted by $\mathbf{w}_\mathrm{r2}(n) = 1$. The weight vector of N_r3 samples is given by $\mathbf{w}_\mathrm{r3} = e^{-\beta(n-N_\mathrm{r3})/C}$. The parameters β and C control the exponential envelope of the designed vector. The choice of this vector replicates the exponential decay of the late reverberant tails seen in real room responses. The algorithm alternates between the time and frequency domains using the DFT and per-frequency matrix multiplication. Fig. 1 depicts the overall shape of the window. The algorithm proceeds by repetition of the steps presented in Algorithm 1.

Algorithm 1 Joint time-frequency domain crosstalk cancellation

Step 1: $\mathbf{R}(\omega) \leftarrow \mathbf{C}(\omega)\mathbf{H}(\omega)$,
Step 2: $|\mathbf{R}_{ms}(\omega)| \leftarrow \gamma|\mathbf{R}_{ms}(\omega)|$ for $m \neq s$,
$|\mathbf{R}_{mm}(\omega)| \leftarrow |\mathbf{R}_{mm}(\omega)|^{\delta}$,
Step 3: $\mathbf{H}(\omega) \leftarrow (\mathbf{C}(\omega)^H\mathbf{C}(\omega) + \lambda\mathbf{I})^{-1}\mathbf{C}(\omega)\mathbf{R}(\omega)$,
Step 4: $\mathbf{h} \leftarrow$ inverse DFT($\mathbf{H}$),
Step 5: $\mathbf{h}_\mathrm{ls} \leftarrow \mathbf{h}_{ls}(n)\mathbf{w}(n)$,
Step 6: $\mathbf{H} \leftarrow$ DFT($\mathbf{h}$).
Here $0 < \gamma < 1$, $0 < \delta < 1$ and $\lambda > 0$.

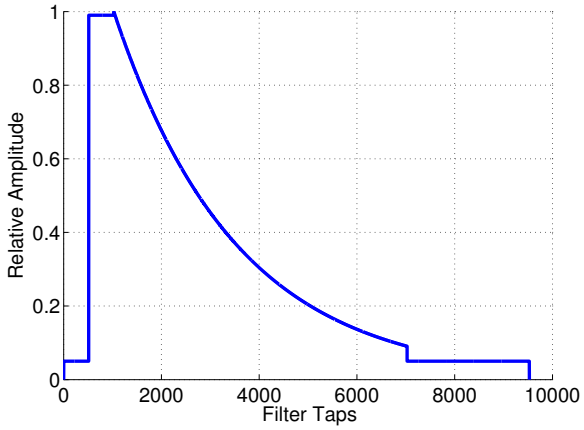


Figure 1: Window function to achieve filter length no longer than N_{h}. Estimated room impulse responses are multiplied with the window to gradually achieve this over the total number of iterations of the algorithm.

3 Simultaneous Channel Estimation with Low levels of MLS

The signals sent to the loudspeakers of a multichannel audio system are highly correlated and do not necessarily excite all the frequencies that may be of interest in the near future. Moreover, the spectral content of speech or music signals changes abruptly and thus channel estimates obtained using earlier signals may not be adequate. The approach which we propose here overcomes this problem by adding subliminal quantities of independent noise to each source signal to sound each channel simultaneously. In order to perform accurate simultaneous channel estimation, signals which are used to estimate the channels need to possess near to zero cross-correlation properties. MLS have near to zero cross-correlation and thus aid in decorrelating these signals. Upon sounding the channels, we can build the crosstalk cancelling filters which are based on the method described in Section 2. Fig. 2 describes the proposed technique to carry out simultaneous channel estimation for a 3×2 crosstalk cancellation system (3 loudspeakers and 2 microphones). The source signals are combined with the MLS to sound each channel. The channel estimates are based on a least mean squares (LMS) type update. Each update is governed by

$$\hat{\mathbf{C}}_{ml}^{(k+1)}(\omega) = \alpha \hat{\mathbf{C}}_{ml}^{(k)}(\omega) + (1-\alpha)\frac{M_m^{(k)}(\omega)}{(\psi)S_l^{(k)}(\omega) + (1-\psi)L_l^{(k)}(\omega)}. \tag{4}$$

Here $M_m^{(k)}(\omega)$ is the windowed signal received at the microphone m, $S_l^{(k)}(\omega)$ is the windowed source signal emitted by loudspeaker l and $L_l^{(k)}(\omega)$ is the MLS signal emitted by loudspeaker l. The update permits the tradeoff between the level of signal and the level of MLS. Setting $\psi = 0$ yields the channels being estimated only by MLS while setting $\psi = 1$ yields the channels being estimated by the windowed source signals. In the latter we expect the channel estimation error to be higher as the channel sounding signals are unable excite all frequencies due to its rapidly varying spectral content. The value of ψ can be adjusted to obtain the optimal performance in terms of the estimation error without affecting the audio perception and quality.

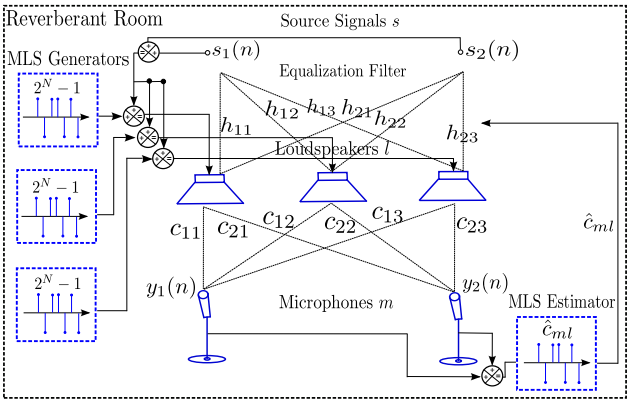


Figure 2: System model for simultaneous channel estimation for a 3×2 crosstalk cancellation system, showing the independently superimposed MLS with source signals resulting in channel estimates given by $\hat{\mathbf{c}}_{ml}$ from loudspeaker l to microphone m. The MLS channel estimator feeds back the channel estimates to design the crosstalk cancelling filters.

4 Results

The signal-to-noise-ratio (SNR) in the estimation process increases if multiple MLS repetitions are used [15]. For a 3×2 crosstalk cancellation system, the channel estimation quality was tested by examining the RMS channel estimation error in dB vs. the number of MLS repetitions needed to reduce the error. Fig. 3 shows the channel estimation quality for simultaneous and individual channel estimation with and without the presence of source signals. In the cases where the source signals were present, two music signals were used with the average signal strength being 65dB in sound pressure level (SPL). The average MLS strength was 26dB SPL. The channels were estimated in a real rectangular room of 3.5m×2.5m×4.5m with a reverberation time T_{60} of approximately 0.25s. The channel impulse responses were estimated using a sampling frequency of 44.1kHz. Order 16 MLS were chosen for all experiments. The RMS performance converges to -20dB with 40 MLS repetitions. The RMS error is the lowest at -23dB when individual channels are sounded with MLS and no source signals. Upon estimating the channels, crosstalk cancellation filters were designed with the window function presented in Fig. 1. We chose $\beta = 0.001$, as a value which is close to 0 suppresses majority of the late reverberations. With $\gamma = 0.01$, $\delta = 0.1$ and $\lambda = 0.001$, the resulting crosstalk and direct channel frequency responses are shown in Fig. 4. The average crosstalk magnitudes for both crosstalk channels is -25dB and the direct channel responses can be seen to have a flat frequency response with ± 0.5dB deviation from the average magnitude. MLS signal levels were altered to monitor the effect on the crosstalk magnitude and are presented in Fig. 5. The variations reflect how much MLS is used to estimate the channels. With MLS signal energy set to -26dB, low crosstalk magnitude is obtained. The MLS was empirically found to be subjectively audible when MLS signal energy is -30dB or above with no music signal present. The threshold of subjective audibility comes from the simultaneous masking limit [9]. With greater MLS signal energy, lower channel estimation error and lower crosstalk magnitude is achieved. However, increasing the MLS signal energy makes the MLS signals more dominant in terms of their audibility. It is only when the MLS injection becomes audible, a significant difference in the crosstalk levels are noticed. Thus, a tradeoff between the MLS signal energy vs. the source signal energy arises, as one aims for low crosstalk with enough MLS energy for low error channel estimation. Although not demonstrated here, the level of MLS added can be varied adaptively based on

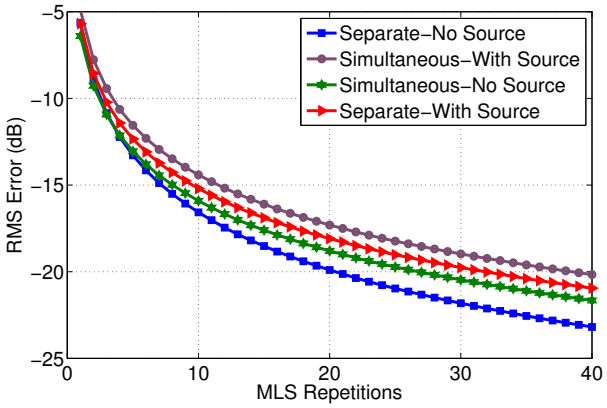


Figure 3: Channel estimation quality vs. MLS repetitions for a 3×2 crosstalk system. The figure shows RMS error (dB) for individual (separate) and simultaneous channel estimation with and without source signals.

the signal levels present.

5 Conclusions

An original simultaneous channel estimation and crosstalk cancellation approach is presented. Channel estimation is aided by the addition of subliminal levels of MLS to the loudspeaker signals. The crosstalk cancellation filters are designed using an original method involving alternation between the time and frequency domains to satisfy constraints in both domains. Estimation with multiple MLS shows a reduction in the estimation error and an improvement in the quality of the crosstalk magnitude. Decreasing the energy level of MLS results in degraded crosstalk performance, as the resulting estimate is a mixture of MLS energy and superimposed source signal energy. By controlling the late reverberations with an exponentially decaying weight, the designed crosstalk cancellation filters are found to require fewer taps than the traditional impulse response inversion techniques. This increased the efficiency of the crosstalk cancellation system. A real time system implementing this approach is currently under development.

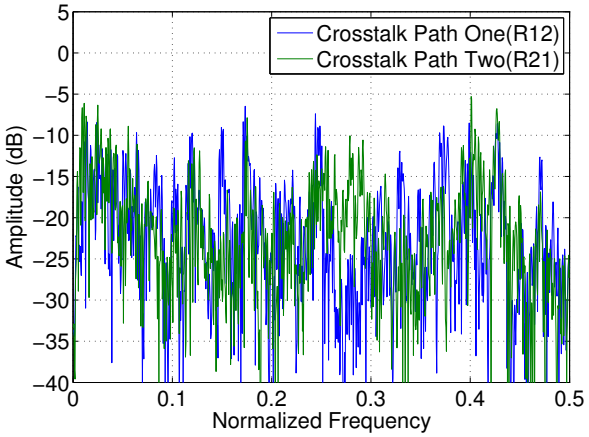
References

[1] B.S. Atal and M.R. Schroeder. Apparent sound source translator. In *U.S. Patent*, volume 3, pages 236–949, 1962.

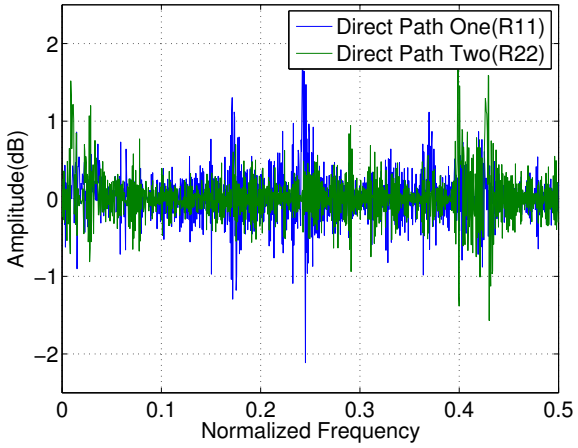
[2] J. Benesty, D.R. Morgan, and M.M. Sondhi. A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation. In *IEEE Transactions on Speech Audio Processing*, volume 6, pages 156–165. IEEE, 1998.

[3] D. Griesinger. Beyond mls - occupied hall measurement with fft techniques. In *Journal of Audio Engineering Society*, volume 44, page 1174. AES, 1996.

[4] J. Herre, H. Buchner, and W. Kellermann. Acoustic echo cancellation for surround sound using perceptually motivated convergence enhancement. In *IEEE International*



(a)



(b)

Figure 4: (a) With $\beta = 0.001$, crosstalk cancellation transfer functions (i.e., source 1 to microphone 2 and source 2 to microphone 1) in the frequency domain are presented. The average crosstalk magnitude in both crosstalk channels is -25dB. (b) The direct channel transfer functions (i.e., source 1 to microphone 1 and source 2 to microphone 2) are seen to be flat across all frequencies.

Conference on Acoustics, Speech and Signal Processing (ICASSP), volume 1, pages I–17 – I–20. IEEE, 2007.

[5] Y. Huang, J. Benesty, and J. Chen. Generalized crosstalk cancellation and equalization using multiple loudspeakers for 3-d sound reproduction at the ears of multiple listeners. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 405–408. IEEE, 2008.

[6] F. Lim, M.R.P. Thomas, and P.A. Naylor. A spatially aware channel equalizer. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 1–4. IEEE, 2013.

[7] S. Miyabe, M. Shimada, T. Takatani, H. Saruwatari, and K. Shikano. Multi-channel inverse filtering with selection and enhancement of a loudspeaker for robust soundfield

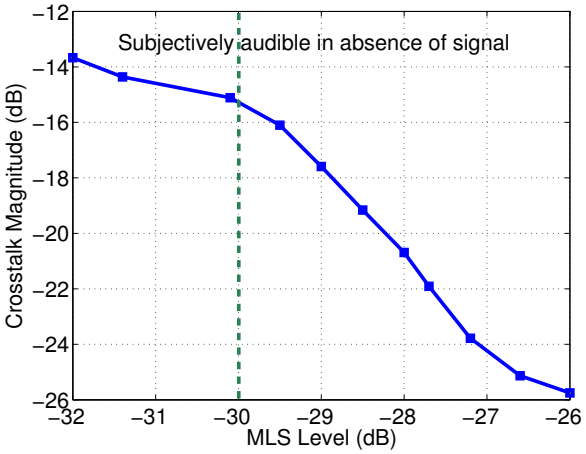


Figure 5: MLS level vs. Crosstalk Magnitude is shown. The subjectively audible boundary lies near −30dB MLS signal energy.

reproduction. In *International Workshop on Acoustic Signal Enhancement; Proceedings of IWAENC 2006*, pages 1–4. IWAENC, 2006.

[8] M. Miyoshi and K. Kaneda. Inverse filtering of room acoustics. In *IEEE Transactions on Acoustics, Speech, Signal Processing*, volume 36, pages 145–152. IEEE, 1988.

[9] B.C.J. Moore. *An Introduction to the Psychology of Hearing*. Elsevier Academic Press, London, UK, 2004.

[10] D.R. Morgan, J. L. Hall, and J. Benesty. Investigation of several types of nonlinearites for use in stereo acoustic echo cancellation. *IEEE Transactions on Speech Audio Processing*, 9(6):686–696, 2001.

[11] A. Mouchtaris, P. Reveliotis, and C. Kyriakakis. Inverse filter design for immersive audio rendering over loudspeakers. *IEEE Transactions on Multimedia Processing*, 2(2):77–87, 2000.

[12] P.A. Nelson, H. Hamada, and S.J. Elliott. Adaptive inverse filters for stereophonic sound reproduction. *IEEE Transactions on Signal Processing*, 40(7):1621–1633, 1992.

[13] P.A. Nelson, F. Orduna-Bustamante, and H. Hamada. Inverse filter design and equalization zones in multichannel sound reproduction. *IEEE Transactions on Speech and Audio Processing*, 3(3):185–192, 1995.

[14] J-D Polack, H. Alrutz, H., and M. Schroeder. The modulation transfer function of music signals and its applications to reverberation measurement. *Acta Acoustica*, 54(5):257–265, 1984.

[15] D.D. Rife and J. Vanderkooy. Transfer-function measurement with maximum length sequences. In *Journal of Audio Engineering Society*, volume 37, pages 419–444. AES, 1989.

[16] U.P. Svensson and J.L. Nielsen. Errors in mls measurements caused by time variance in acoustic systems. In *Journal of Audio Engineering Society*, volume 47, pages 907–927. AES, 1999.

[17] T.S. Wada and B.H. Juang. Multi-channel acoustic echo cancellation based on residual echo enhancement with effective channel decorrelation and resampling. In *International Workshop on Acoustic Signal Enhancement; Proceedings of IWAENC 2010*, volume 6, pages 1–4. IWAENC, 2010.

[18] D. Wilson, A. Ziemann, V.E. Ostashev, and A.G. Voronovich. An overview of acoustic travel-time tomography in the atmosphere and its potential applications. In *Acoustic Acta Acustic*, volume 87, pages 721–730. EAA, 2001.