2014 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)

AUTOMATIC SPATIAL GAIN CONTROL FOR AN INFORMED SPATIAL FILTER

Sebastian Braun, Oliver Thiergart, and Emanuël A. P. Habets

International Audio Laboratories Erlangen*
Am Wolfsmantel 33, 91058 Erlangen, Germany
{sebastian.braun, oliver.thiergart, emanuel.habets}@audiolabs-erlangen.de

ABSTRACT

When capturing speech in a multi-talker telecommunication scenario, it is desirable to keep the enhanced signal at an equal loudness level for each speaker. Single-channel automatic gain control systems are not able to adjust the level of different talkers when they are simultaneously active. In this work, an automatic spatial gain control (ASGC) algorithm is proposed that adjusts the directional response of an existing informed spatial filter such that the direct sound of multiple sources can be kept at a constant desired loudness level at the output. The spatial filter additionally reduces diffuse sound and ambient noise. It is shown that the proposed AGSC works well within the tested scenario, and is able to adjust the levels of different speakers even during double talk scenarios.

Index Terms— automatic gain control, microphone arrays, informed spatial filtering

1. INTRODUCTION

Modern communication systems aim at acquiring the speech of multiple talkers being active at the same time in a reverberant environment. In such scenarios it is desirable to capture the speech of the different talkers with a high signal-to-noise ratio (SNR) and a high signal-to-reverberation ratio. For the sound acquisition, we usually employ a microphone array. There exists a huge variety of spatial filters to capture the speech that enable dereverberation and noise reduction.

A typical problem is that the speech of several talkers is often captured with different loudness levels. This problem arises, for instance, if the talkers are located at different distances to the microphone array or if the individual loudness levels vary among the talkers. The loudness differences result in a high dynamic range of the captured speech, which can be problematic for speech coding and transmission as well as sound reproduction with high speech intelligibility. Therefore, it would be desirable to capture the speech of the different talkers with equal loudness for all directions.

Algorithms that aim at adjusting the level of sound sources, such that a desired level is achieved, are referred to as automatic gain control (AGC). There exist several AGC algorithms for various applications [1–4] that can be applied, for example, after a speech enhancement algorithm. Typically, such AGC algorithms require a rather long and noticeable adaption time, which is problematic when multiple talkers with different levels alternate quickly. Furthermore, they are unable to adjust the level of individual talkers during multi-talk. An approach for directional AGC suitable for hands-free communication was proposed in [5] that uses collected spatial information to control an adaptive beamformer together with an acoustic echo control. However, this approach assumes that only one source is active at the same time, which is problematic in a scenario involving simultaneously active talkers.

In this paper, we propose an automatic spatial gain control (ASGC) system that allows to capture multiple speech sources with equal loudness from all directions. The proposed approach is based on the informed linearly constrained minimum variance (LCMV) filter introduced in [6]. It exploits instantaneous spatial information on the acoustic scene such as the direction-of-arrival (DOA) of the sound. In contrast to the directional AGC method proposed in [5], we do not need a beamformer that scans all directions to obtain the directions of the talkers. By incorporating the spatial information into the AGC algorithm, we can assign an individual gain to each direction and therefore each spatially separated source. By using individual gains, the adaption time is smaller or equal then the time required by conventional single-channel AGC algorithms. This way, level fluctuations as a result of alternating talkers can be reduced almost completely. Since the proposed system is based on a multi-wave signal model (i. e., multiple simultaneously arriving plane waves per time and frequency), the gain for each individual talker can be adapted independently, even during multi-talk.

The paper is organized as follows: Section 2 formulates the problem. In Sec. 3, the informed spatial filter is formulated and the ASGC system is derived in Sec. 4. The proposed system is evaluated in Sec. 5 based on simulations. The conclusions are drawn in Sec. 6.

2. PROBLEM FORMULATION

In the following, we employ the multi-wave signal model introduced in [6]. Let us consider an array of M omnidirectional microphones located at $\mathbf{d}_{1\dots M}$. For each time-frequency instant we assume a sound field composed of $L < M$ plane waves propagating in an isotropic and spatially homogenous diffuse sound field. The L plane waves represent the direct sound of multiple sound sources which are located in a reverberant environment while the diffuse sound models the reverberation. In the short-time Fourier transform (STFT) domain, the microphone signals $\mathbf{x}(k,n) = [X(k,n,\mathbf{d}_1)\dots X(k,n,\mathbf{d}_M)]^{\mathrm{T}}$ at frequency index k and time index n can be written as

$$\mathbf{x}(k,n) = \sum_{l=1}^{L} \mathbf{x}_l(k,n) + \mathbf{x}_{\mathrm{d}}(k,n) + \mathbf{x}_{\mathrm{n}}(k,n), \qquad (1)$$

where $\mathbf{x}_l(k,n)$ are the microphone signals proportional to the sound pressure of the l-th plane wave, $\mathbf{x}_{\mathrm{d}}(k,n)$ is the diffuse sound component, and $\mathbf{x}_{\mathrm{n}}(k,n)$ is a stationary noise component (e. g., stationary background noise and/or microphone self-noise). The vector

*A joint institution of the Friedrich-Alexander-University Erlangen-Nürnberg (FAU) and Fraunhofer IIS, Germany.

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

$\mathbf{x}_l(k,n)$ can be written as

$$\mathbf{x}_l(k,n) = \mathbf{a}_l(k,n)\, X_l(k,n,\mathbf{d}_1), \tag{2}$$

where $X_l(k,n,\mathbf{d}_1)$ is the sound pressure of the l-th plane wave at the first microphone. The m-th element of the propagation vector $\mathbf{a}_l(k,n)$ with $m \in \{1,2,\ldots,M\}$ is the transfer function for a plane wave from the first to the m-th microphone, which depends on the DOA $\theta_l(k,n)$ of the l-th plane wave. The power spectral density (PSD) of the l-th plane wave given by

$$\phi_l(k,n) = \mathrm{E}\left\{|X_l(k,n,\mathbf{d}_1)|^2\right\} \tag{3}$$

is assumed independent of $\mathbf{d}_m$, which is a reasonable assumption if the microphones are positioned sufficiently close to each other.

Assuming that all components in (1) are mutually uncorrelated, we can express the PSD matrix of the microphone signals as

$$\mathbf{\Phi}_x(k,n) = \mathrm{E}\left\{\mathbf{x}(k,n)\,\mathbf{x}^{\mathrm{H}}(k,n)\right\} \tag{4a}$$

$$= \sum_{l=1}^{L} \mathbf{\Phi}_l(k,n) + \mathbf{\Phi}_{\mathrm{u}}(k,n), \tag{4b}$$

where $\mathbf{\Phi}_l(k,n) = \phi_l(k,n)\;\mathbf{a}_l(k,n)\,\mathbf{a}_l^{\mathrm{H}}(k,n)$ is the PSD matrix resulting from the l-th plane wave and $\mathbf{\Phi}_{\mathrm{u}}(k,n) = \mathbf{\Phi}_{\mathrm{d}}(k,n) + \mathbf{\Phi}_{\mathrm{n}}(k,n)$ is the PSD matrix of the diffuse sound and stationary noise.

In this paper, we aim at capturing the direct sound of the sound sources without distortion and with the desired loudness from each direction while suppressing diffuse sound and stationary noise. The desired output signal can be expressed as a weighted sum of the L plane waves at the first microphone, i. e.,

$$Y(k,n) = \sum_{l=1}^{L} G(\theta_l,n)\, X_l(k,n,\mathbf{d}_1), \tag{5}$$

where the spatial gain function $G(\theta,n) \in \mathbb{R}^+$ is adjusted such that the L direct sound components in $Y(k,n)$ yield the desired loudness.

The estimation of $Y(k,n)$ given $G(\theta,n)$ is shown in Sec. 3. The computation of $G(\theta,n)$ is derived in Sec. 4.

3. INFORMED SPATIAL FILTER

An estimate of desired signal $Y(k,n)$ in (5) is obtained by a linear combination of the microphone signals $\mathbf{x}(k,n)$, i. e.,

$$\widehat{Y}(k,n) = \mathbf{h}_{\mathrm{iLCMV}}^{\mathrm{H}}(k,n)\,\mathbf{x}(k,n), \tag{6}$$

where $\mathbf{h}_{\mathrm{iLCMV}}(k,n)$ is a complex weight vector of length M. To find the weights $\mathbf{h}_{\mathrm{iLCMV}}(k,n)$, we employ the informed LCMV filter introduced in [6] that minimizes the diffuse sound and noise while capturing the L plane waves distortionless and with the desired spatial gain $G(\theta_l,n)$. The filter is formulated as

$$\mathbf{h}_{\mathrm{iLCMV}}(k,n) = \arg\min_{\mathbf{h}}\; \mathbf{h}^{\mathrm{H}}\,\mathbf{\Phi}_{\mathrm{u}}(k,n)\,\mathbf{h} \tag{7a}$$

$$\text{s.t.}\quad \mathbf{h}^{\mathrm{H}}\mathbf{a}_l(k,n) = G(\theta_l,n) \quad \forall\, l \in \{1,2,\ldots,L\}, \tag{7b}$$

for which a closed-form solution is given in [7]. The PSD matrix $\mathbf{\Phi}_{\mathrm{u}}(k,n)$ is computed using an estimate of the diffuse sound PSD matrix $\mathbf{\Phi}_{\mathrm{d}}(k,n)$ that can be obtained using an auxiliary spatial filter as shown in [6,8], and an estimate of the noise PSD matrix $\mathbf{\Phi}_{\mathrm{v}}(k,n)$ that for example can be estimated during speech pauses. Since the stationary noise statistics can also change at some point in time, $\mathbf{\Phi}_{\mathrm{v}}(k,n)$ is denoted time-dependent. Such changes can be tracked for example using the method proposed in [9]. The elements of the propagation vectors $\mathbf{a}_l(k,n)$ can be estimated with narrow-band DOA estimators such as ESPRIT [10] or root MUSIC [11]. Note that the constraints in (7b) as well as $\mathbf{\Phi}_{\mathrm{u}}(k,n)$ typically fluctuate highly across time and frequency. Thus, the filter weights $\mathbf{h}_{\mathrm{iLCMV}}(k,n)$ are recomputed for each k and n to adapt quickly to the current situation. If the spatial gain function is set to $G(\theta,n) = 1\ \forall(\theta,n)$, an uniform spatial power response is obtained for all arriving waves. Consequently, all direct sound components are preserved with their original power while reducing background noise and reverberation.

4. AUTOMATIC SPATIAL GAIN CONTROL

4.1. Spatial Loudness Equalization

As discussed in Sec. 2, the spatial gain function $G(\theta,n)$ used in (5) and (7), respectively, has to be adjusted such that the direct sound components in $Y(k,n)$ yield the desired loudness. To find the correct $G(\theta,n)$, we introduce the spatial loudness distribution (SLD), which characterizes the loudness of the arriving direct sound for each direction of the angle θ. The SLD $\Psi(\theta,n)$ is computed by summing across frequency the loudness-weighted power of the plane waves arriving from θ_l, i. e.,

$$\Psi(\theta,n) = \sum_{k=1}^{K} \beta^2(k) \sum_{l=1}^{L} \delta_{\theta,\theta_l}\, \phi_l(k,n), \tag{8}$$

where K is the highest frequency band of interest, $\beta(k)$ is a loudness weighting, and δ_{θ,θ_l} is the Kronecker delta. An estimate of the PSDs $\phi_l(k,n)$ can be obtained by computing the PSD of the output of an additional spatial filter that estimates $X_l(k,n,\mathbf{d}_1)$. For instance, we can employ the informed LCMV filter (7) by setting $G(\theta,n) = \delta_{\theta,\theta_l}$ in (7b) to extract the l-th plane wave. Alternatively, one can also use the informed minimum mean square error (MMSE) filter proposed in [8].

In this paper, we aim at keeping the long-term spatial loudness distribution (LT-SLD) at a desired level. The LT-SLD is obtained by

$$\Psi_{\mathrm{LT}}(\theta,n) = \alpha_{\mathrm{LT}}\,\Psi_{\mathrm{LT}}(\theta,n-1) + (1-\alpha_{\mathrm{LT}})\,\Psi(\theta,n), \tag{9}$$

where $\alpha_{\mathrm{LT}} \in [0,1]$ is the recursive long-term averaging factor.

The SLD $\Psi(\theta,n)$ is used as a control parameter as explained in detail in Sec. 4.2, whereas the LT-SLD $\Psi_{\mathrm{LT}}(\theta,n)$ is used to compute the spatial gain function $G(\theta,n)$. More precisely, the desired target loudness $\Psi_{\mathrm{target}}(\theta)$ at the output can be obtained for all sources by computing $G(\theta,n)$ as

$$G(\theta,n) = \sqrt{\frac{\Psi_{\mathrm{target}}(\theta)}{\Psi_{\mathrm{LT}}(\theta,n)}}. \tag{10}$$

When the target SLD $\Psi_{\mathrm{target}}(\theta)$ is independent of θ, the desired signal $Y(k,n)$ will be characterized by an equal loudness of the direct sound for all directions. As the spatial gain function $G(\theta,n)$ is defined frequency independent, we avoid undesired spectral coloration of broadband sound sources.

The global system is depicted in Fig. 1. The parameter estimation block includes the estimation of the diffuse sound plus stationary noise PSD matrix, DOA estimation, and the plane wave PSD estimation. In the ASGC block, the SLD and LT-SLD are computed, which are used to compute the spatial gain function. The spatial gain

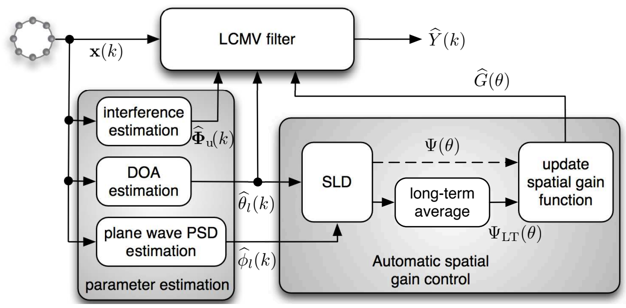

Fig. 1. ASGC algorithm for an informed LCMV filter

function and the other parameters are then used to compute the informed LCMV filter. In the next subsection, an adaptive procedure to update the spatial gain function is described.

4.2. Computation of the Spatial Gain Function

While the spatial gain function $G(\theta, n)$ in (10) equalizes the long-term directional loudness, its direct use in (7) would not yield a satisfactory result. Firstly, directly using (10) would lead to a large gain for directions where no source is located and the LT-SLD $\Psi_{\mathrm{LT}}(\theta, n)$ is consequently low. This would potentially amplify the noise and diffuse sound and lead to an unstable system. Secondly, the spatial and temporal variance of the LT-SLD would yield a fluctuating gain function $G(\theta, n)$, which might cause audible artifacts in practice.

To mitigate these problems, we first apply an iterative approach to compute a spatially smooth $G'(\theta, n)$ that possesses unit gain for directions where the SLD $\Psi(\theta, n)$ is below a specific threshold $\Psi_{\mathrm{min}}(n)$. The computation of $G'(\theta, n)$ is outlined in Algorithm 1. In this paper, the time-varying threshold $\Psi_{\mathrm{min}}(n)$ is obtained by the mean of the SLD $\Psi(\theta, n)$ limited by a fixed lower bound Ψ_0, i. e.,

$$\Psi_{\mathrm{min}}(n) = \max\left(\frac{1}{2\pi}\int_0^{2\pi} \Psi(\theta, n)\ \mathrm{d}\theta,\ \Psi_0\right). \tag{11}$$

Secondly, to reduce temporal fluctuations of the spatial gain function between subsequent frames, we compute a function $\widehat{G}(\theta, n)$ which approaches $G'(\theta, n)$ by an adaptive update. Finally, the function $\widehat{G}(\theta, n)$ is used in (7). The logarithmic gain difference to the adaptive gain function of the previous frame is given by

$$\Delta_G(\theta, n) = \exp\left[\log G'(\theta, n) - \log \widehat{G}(\theta, n-1)\right]. \tag{12}$$

The adaptive spatial gain function is then obtained by the update rule

$$\widehat{G}(\theta, n) = \widehat{G}(\theta, n-1)\left[1 + \nu(\theta, n)\right], \tag{13}$$

Algorithm 1 computing the spatially smooth gain function $G'(\theta, n)$

1. Initialize $G'(\theta, n)$ with unit gain, i. e., $G'(\theta, n) = 1\ \forall\theta$
2. Find the global maximum θ_{max} of the LT-SLD $\Psi_{\mathrm{LT}}(\theta, n)$ and compute the spatial gain $G(\theta_{\mathrm{max}}, n)$ using (10)
3. Generate a spatial window of predefined width Θ which has unit gain at the edges and gain $G(\theta_{\mathrm{max}}, n)$ at the center
4. Place the spatial window in $G'(\theta, n)$ at θ_{max} and set $\Psi_{\mathrm{LT}}(\theta, n)$ to zero at directions that are covered by the spatial window
5. Continue with Step 2 until all points in $\Psi_{\mathrm{LT}}(\theta, n)$ are below the threshold $\Psi_{\mathrm{min}}(n)$.

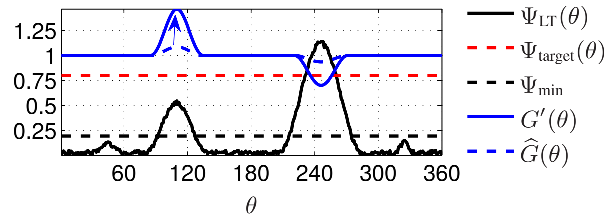

Fig. 2. Schematic computation of the spatial gain function

where $\nu(\theta, n)$ is the update term given by

$$\nu(\theta, n) = \begin{cases} \mu_G\ \Delta_G(\theta, n), & \Psi(\theta, n) > \Psi_{\mathrm{min}}(n) \\ \mu_{\mathrm{r}}\ \mathrm{sgn}\left[1 - \widehat{G}(\theta, n-1)\right], & \text{otherwise,} \end{cases} \tag{14}$$

where $\mu_G > 0$ is the stepsize, μ_{r} is the reset constant and $\mathrm{sgn}[\cdot]$ denotes the sign function. The reset constant μ_{r} should be chosen sufficiently small, depending on the application. The second case in (14) provides an automatic reset if a source becomes inactive or changes position: for directions where the SLD is below the threshold, the spatial gain function gradually tends towards a unit gain. The adaptive spatial gain function $\widehat{G}(\theta, n)$ can be spatially lowpass filtered before it is used in (7) to remove possible discontinuities along θ.

A typical LT-SLD is shown in Fig. 2 as black solid line. The desired target SLD $\Psi_{\mathrm{target}}(\theta)$ and the threshold Ψ_{min} are shown as red dashed line and black dashed line. The direct spatial equalizing function $G'(\theta)$ as obtained by Algorithm 1 is shown as blue solid line and the adaptive gain function $\widehat{G}(\theta)$ as obtained by (13) as blue dashed line, indicating that it is approaching $G'(\theta)$ over time.

The obtained spatial gain function $\widehat{G}(\theta, n)$ adapts the level of all sources to the desired spatial output loudness. For a specific time depending on μ_{r}, the gain function stays almost constant for directions θ where currently no source is active, i. e., where $\Psi(\theta, n)$ is low. Therefore, a pausing and resuming speaker has immediately the desired target loudness at the output.

5. EVALUATION

In this section, we evaluate the ASGC algorithm using simulated room impulse responses.

5.1. Setup and Parameters

A shoebox room of size $5 \times 6 \times 3.5$ ms with a $T_{60} = 300$ ms is simulated with the image method [12]. Three talkers simulated as isotropic point-like sources are positioned at $101°, 333°$ and $189°$ at a distance of 1.8 m from the array center. The first speaker is active continuously, the second and third speakers are active alternately. We used a circular array with $M = 12$ equally spaced microphones with an array radius of 5 cm. Spatially white noise was added with an SNR of 40 dB.

For the STFT processing, a sampling rate of 16 kHz, a Hann window of length 32 ms, and 50 % overlap is used. All parameters of the informed spatial filter are estimated online. Two DOAs per time-frequency, i. e. $L = 2$, are estimated via root MUSIC in the beamspace domain [13]. The noise PSD matrix is estimated as proposed in [14] and the diffuse sound PSD matrix is estimated as proposed in [6]. The PSD matrices were obtained by recursive averaging with a time constant of 70 ms. The plane wave PSDs are obtained by the MMSE spatial filter proposed in [8]. For the frequency loudness weighting function $\beta(k)$ we employed the K-frequency

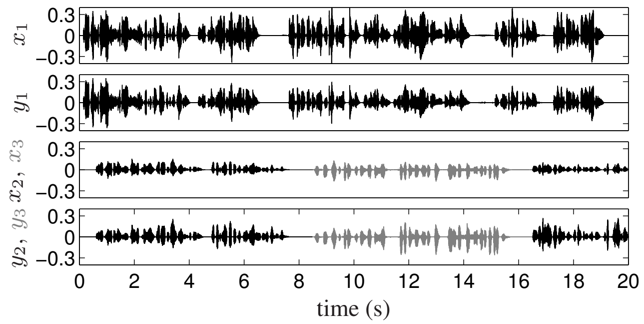

Fig. 3. Direct components x_l of the three speakers at the reference microphone and corresponding direct filtered output signals y_l.

weighting proposed in [15]. The recursive smoothing constant α_{LT} for the LT-SLD was chosen corresponding to 0.6 s and 2 s for increasing and decreasing values of the SLD, respectively. The desired target SLD was chosen $\Psi_{\text{target}} = 3$ dB, the lower bound threshold $\Psi_0 = -15$ dB, the adaptive stepsize was set to $\mu_G = 0.007$ and the reset constant $\mu_{\text{r}} = 0.2$ dB/s. The spatial window in Algorithm 1 was a square-root sine window of width $\Theta = 50^\circ$ as shown in Fig. 2, and $\widehat{G}(\theta, n)$ was spatially low-pass filtered.

5.2. Results

Since an evaluation of the diffuse sound and noise reduction performance can be already found in [6], we focus on the performance of the ASGC. The simulation setup is chosen such that different cases for the ASGC algorithm can be studied in one scenario: i) sources that have to be attenuated or amplified, ii) speech pause and resuming source, and iii) more source positions than simultaneously available DOA estimates. Figure 3 shows the direct signal components $x_l(t)$ at the reference microphone and the corresponding processed direct signal components $y_l(t)$. It can be observed that the direct signal component in $y_1(t)$ is attenuated, while it is amplified in $y_2(t)$ and $y_3(t)$.

Figure 4 shows the SLD on top and the binary mask of $\Psi(\theta, n) > \Psi_{\text{min}}(n)$ below. The gain function $\widehat{G}(\theta, n)$ adapts only for black time-DOA points in this plot. The middle plot in Fig. 4 shows the LT-SLD. To match the chosen target SLD $\Psi_{\text{target}} = 9$ dB, the first source at 101° has to be slightly attenuated, the second source at 333° to be amplified, and the third source at 189° has also to be slightly amplified. The finally computed spatial gain function $\widehat{G}(\theta, n)$ in Fig. 4 shows exactly this desired behavior. Due to the temporal averaging and the adaptive update, it takes a short time period until the desired output LT-SLD is reached. At $t = 9$ s, where source $x_3(t)$ becomes active, its presence is tracked and the gain for the corresponding directions is adjusted. The bottom plot in Fig. 4 shows the effective output LT-SLD, i. e. the multiplication of the LT-SLD with the finally applied spatial gain function. The desired target SLD is matched for the three directions with a fluctuation of less than ± 1 dB.

The perceived loudness calculated after [15] is shown in Fig. 5. It shows that the perceived loudness of the direct sound of the single sources adapts towards an equal level. Informal listening tests[1] confirm these results.

[1] Audio examples are available at `http://www.audiolabs-erlangen.de/resources/2014-ICASSP-ASGC`

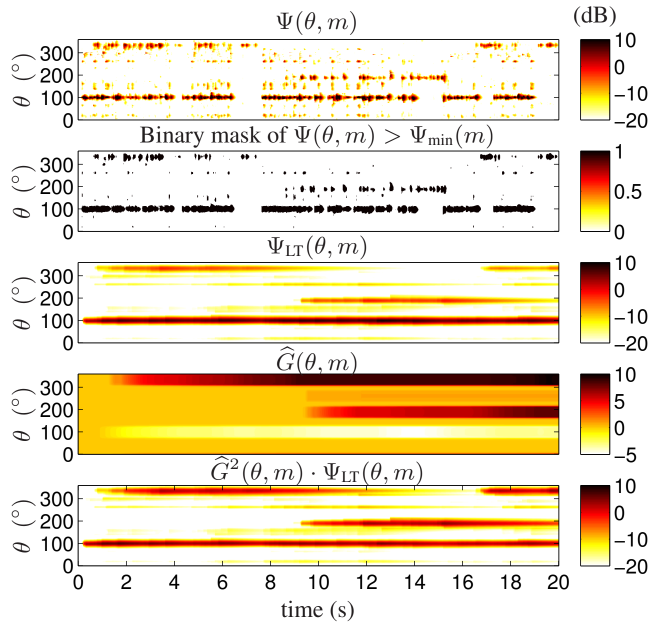

Fig. 4. From top to bottom: SLD, binary mask of the time-DOA points where the gain function adapts, LT-SLD, spatial gain function, effective output LT-SLD.

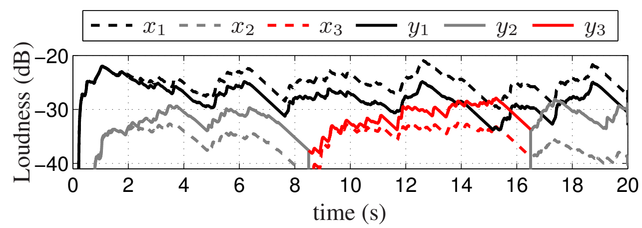

Fig. 5. Loudness (ITU BS.1770-3) before and after processing.

In general, the results in Fig. 4 show that the ASGC algorithm is able to adapt to changing sound scenarios, such as switching positions. The fact that $\widehat{G}(\theta, n)$ remains almost constant for directions where no speaker is active ensures a spatial memory. This effect can be observed when speaker two is talking again at $t = 16.5$ s. His output loudness is immediately adjusted to the desired level. A faster reset constant μ_{r} reduces this spatial memory effect.

6. CONCLUSION

We have proposed an automatic spatial gain control (ASGC) for a multi-source scenario where the long-term loudness of each source is adjusted to a desired level. The ASGC is realized by controlling the directional response of a recently proposed informed spatial filter. This spatial filter is able to capture the direct sound of the sources with an arbitrary response while attenuating diffuse sound and noise. Simulation results show that the ASGC can adjust the loudness of multiple simultaneously active sources to an equal level as desired. Moreover, the system is able to adapt quickly to changing room/speaker scenarios and stores the spatial gain for each direction for a certain time period. For the studied scenario, the system performed robust without producing unpleasant temporal fluctuations in the output signal.

7. REFERENCES

[1] U. Zoelzer, *DAFX: Digital Audio Effects*, John Wiley & Sons, Inc., New York, NY, USA, 2002.

[2] P. L. Chu, "Voice-activated AGC for teleconferencing," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 1996, vol. 2, pp. 929–932 vol. 2.

[3] F. J. Archibald, "Software Implementation of Automatic Gain Controller for Speech Signal," *Texas Instruments SPRAAL1 White Paper*, 2008.

[4] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control: A pracical Approach*, Wiley, New Jersey, USA, 2004.

[5] K. Kobayashi, Y. Haneda, K. Furuya, and A. Kataoka, "A hands-free unit with adaptive microphone array for directional AGC," in *Proc. Intl. Workshop Acoust. Signal Enhancement (IWAENC)*, Sept. 2005.

[6] O. Thiergart and E. A. P. Habets, "An informed LCMV filter based on multiple instantaneous direction-of-arrival estimates," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, May 2013.

[7] H. L. van Trees, *Detection, Estimation, and Modulation Theory*, vol. IV, Optimum Array Processing, Wiley, New York, USA, Apr. 2002.

[8] O. Thiergart, M. Taseska, and E. A. P. Habets, "An informed MMSE filter based on multiple instantaneous direction-of-arrival estimates," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Marrakesh, Morocco, Sept. 2013.

[9] M. Taseska and E. A. P. Habets, "MMSE-based blind source extraction in diffuse noise fields using a complex coherence-based a priori SAP estimator," in *Proc. Intl. Workshop Acoust. Signal Enhancement (IWAENC)*, Sept. 2012.

[10] R. Roy and T. Kailath, "ESPRIT - estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 984–995, 1989.

[11] B. D. Rao and K. V. S. Hari, "Performance analysis of root-music," in *Signals, Systems and Computers, 1988. Twenty-Second Asilomar Conference on*, 1988, vol. 2, pp. 578–582.

[12] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[13] M. D. Zoltowski and C. P. Mathews, "Direction finding with uniform circular arrays via phase mode excitation and beamspace root-MUSIC," vol. 5, pp. 245–248, 1992.

[14] M. Souden, J. Chen, J. Benesty, and S. Affes, "An integrated solution for online multichannel noise tracking and reduction," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 7, pp. 2159–2169, 2011.

[15] ITU-R BS.1770, "Algorithms to measure audio programme loudness and true-peak audio level," 2012.