RE-PANNING OF DIRECTIONAL SIGNALS AND ITS IMPACT ON LOCALIZATION

Alexander Adami, Michael Schoeffler, Jürgen Herre

International Audio Laboratories Erlangen*
Am Wolfsmantel 33, 91058 Erlangen
alexander.adami@audiolabs-erlangen.de

ABSTRACT

For multichannel audio reproduction systems, it is crucial to set up the speakers correctly according to the multichannel format's specification. Especially, the predefined angle of every speaker with respect to the listening position must be strictly kept to avoid spatial distortions of virtual sources, the so called phantom sources. In a normal living room environment, a specification compliant setup is usually not possible. This means, the resulting audio scene may differ heavily from the originally intended scene, i.e., the phantom sources' positions change. To mitigate these spatial distortions, we propose a re-panning method of directional signals. The method groups pairs of adjacent loudspeakers into segments, analyses the direction of arrivals (DOAs) within each segment by means of a direct-ambience decomposition and re-renders the direct components with respect to the actual reproduction setup. The re-panning method was perceptually evaluated by means of a localization listening test.

Index Terms— Spatial audio, format conversion, localization

1. INTRODUCTION

Modern home-cinema high-fidelity systems provide a plurality of loudspeaker channels. Multichannel formats like 5.1, 7.1 or ones with even more and also elevated speakers are available [1–3]. To get the optimal listening experience as intended by the sound engineer, it is crucial that the loudspeakers are placed correctly according to the corresponding format's specification. For example, for a 5.1 audio system, the ITU recommends a setup with the speakers placed equidistantly from the listener and with speaker positions at $0°$, $\pm 30°$ and $\pm 110°$. Since in a normal living room environment it is often not possible to place the speakers in such a way, the speakers' actual positions deviate quite heavily from the ideal ones in distance as well as in angle. While the faulty distances can be quite easily compensated for by applying delays, the angular deviation still causes spatial distortions of the audio scene, i.e., a phantom source at a certain angle will not appear at the intended position.

To overcome this problem, systems were developed which allow to render a given audio scene to an arbitrary reproduction setup. This can be done, for instance, by exploiting physical properties of the audio scene. In [4], the sound propagation in the original sound field and in that of the actual reproduction setup is modeled which allows to derive a conversion matrix between both setups aiming at the physical properties of the sound field in the listening point remaining the same. Directional audio coding (DirAC) is an approach which uses a B-format representation of the input channels to extract spatial parameters like DOAs and diffuseness estimates [5]. The diffuseness estimates can then be used to separate the signals into their direct and diffuse parts, where the former can be re-positioned in accordance to their corresponding DOA and with respect to the reproduction setup [6]. In [7] a system is described which uses principal component analysis (PCA) to separate the input signals into primary and ambient signal parts. The signal parts are spatially analyzed and encoded. At the decoder, the primary and ambient signals are used to render the audio scene according to the reproduction setup. Some additional methods can also be found in [8, 9].

Since the directional information of a phantom source is contained within the direct parts of two or more signals, it is often desirable to decompose the input signals into their direct and ambient signal parts to extract such information. One way to do this is based on pairwise correlations, e.g., [10–12] but also PCA can be used as in [7, 11]. In [13], an analysis of a direct-ambience decomposition based DOA estimator was carried out.

We propose a segment-based re-panning method which uses a correlation-based pair-wise direct-ambience decomposition to extract the directional information of phantom sources within each segment of a multichannel signal produced for a certain loudspeaker setup. The source directions and the extracted direct signals are processed by a re-panner which positions the phantom sources at their estimated position with respect to the actual reproduction setup. In previous methods, only one dominant source per time and frequency instant is allowed which is often too restricted for good sound quality. In the proposed method, this restriction is extended to allow one dominant source per segment and time and frequency instant. The method is perceptually evaluated with respect to localization.

2. PROBLEM FORMULATION

Let us assume a loudspeaker setup as given in Figure 1, where $\mathcal{P}_0 ... \mathcal{P}_i$ denote the ideal loudspeaker positions and $\tilde{\mathcal{P}_0} ... \tilde{\mathcal{P}_i}$ denote the loudspeaker positions in the actual reproduction setup with $i = 0...I-1$ and I denoting the number of available loudspeakers. To each loudspeaker at the ideal position $\mathcal{P}_i$ belongs a driving signal $\mathcal{L}_i(k,m)$, where k and m denote the discrete frequency and time indices of a signal in short-time Fourier transform (STFT) domain. The objective is to determine the loudspeaker driving signal $\tilde{\mathcal{L}_i}(k,m)$ corresponding to the loudspeaker position $\tilde{\mathcal{P}_i}$ in the actual reproduction setup which is compensated with respect to a potential displacement.

To model the loudspeaker signals, the ideal loudspeaker setup is subdivided into segments, where a pair of adjacent loudspeakers form a segment. This leads to the segments $[\{\mathcal{P}_0, \mathcal{P}_1\}, \{\mathcal{P}_1, \mathcal{P}_2\}, \cdots, \{\mathcal{P}_i, \mathcal{P}_j\}]$ with the loudspeaker signals $[\{\mathcal{L}_0(k,m), \mathcal{L}_1(k,m)\}, \{\mathcal{L}_1(k,m), \mathcal{L}_2(k,m)\}, \cdots, \{\mathcal{L}_i(k,m), \mathcal{L}_j(k,m)\}]$, where $j = (i+1)\%I$ and % is the modulo operator. Each loudspeaker driving signal $\mathcal{L}_i(k,m)$ contributes to two segments where each segment signal $S(k,m)$ is assumed to consist of a superposition of a di-

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

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

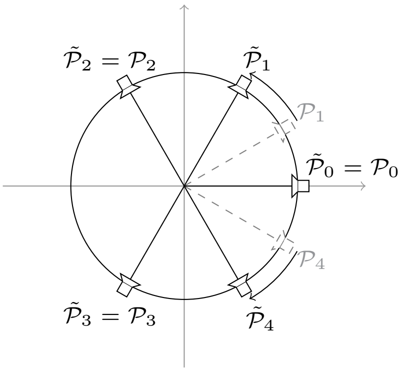

$\tilde{\mathcal{P}}_2 = \mathcal{P}_2$
$\tilde{\mathcal{P}}_1$
$\mathcal{P}_1$
$\tilde{\mathcal{P}}_0 = \mathcal{P}_0$
$\mathcal{P}_4$
$\tilde{\mathcal{P}}_3 = \mathcal{P}_3$
$\tilde{\mathcal{P}}_4$

Fig. 1. Loudspeaker setup with ideal ($\mathcal{P}_i$) and actual ($\tilde{\mathcal{P}}_i$) loudspeaker positions with $i = 0...I-1$ and $I = 5$. The arrows indicate a shift in position.

rect and an ambient signal component denoted by $D(k,m)$ and $A(k,m)$, respectively. With the focus on only one active source, we can model the driving signal as

$$\mathcal{L}_j(k,m) = \underbrace{D^i_j(k,m) + A^i_j(k,m)}_{S^i_j(k,m)} + \underbrace{D^j_j(k,m) + A^j_j(k,m)}_{S^j_j(k,m)}, \quad (1)$$

with superscripts indicating the corresponding segments and where $S^i_j(k,m)$ and $S^j_j(k,m)$ denote the corresponding segment signals. In the following, we assume the speaker driving signals to be equally distributed over the corresponding segment signals:

$$S^i_j(k,m) = S^j_j(k,m) = \frac{1}{2}\mathcal{L}_j(k,m). \quad (2)$$

In Figure 2, the signals corresponding to a segment are depicted.

The corrected loudspeaker signals in the actual reproduction setup can then be modeled as

$$\begin{aligned}\tilde{\mathcal{L}}_j(k,m) &= \tilde{D}^i_j(k,m) + \tilde{A}^i_j(k,m) + \tilde{D}^j_j(k,m) + \tilde{A}^j_j(k,m)\\ &= \eta^i_j(k,m)D'^i_j(k,m) + \zeta^i_j(k,m)A'^i_j(k,m) \quad (3)\\ &\quad + \eta^j_j(k,m)D'^j_j(k,m) + \zeta^j_j(k,m)A'^j_j(k,m),\end{aligned}$$

where $\eta^i_j(k,m)$ and $\eta^j_j(k,m)$ denote re-panning gains, i.e., scaling factors of the respective direct signals corresponding to the jth loudspeaker signal and originating from segments i and j, respectively. Furthermore, $\zeta^i_j(k,m)$ and $\zeta^j_j(k,m)$ denote the scaling factors of the respective ambient signals and $\tilde{(\cdot)}$ denotes entities of the actual reproduction setup. The signals $D'^i_j(k,m)$, $D'^j_j(k,m)$ and $A'^i_j(k,m)$, $A'^j_j(k,m)$ denote the estimated direct and ambient signals actually used for the re-panning.

In this paper, we focus on re-panning of directional signals. The estimated ambient signals will be close to zero and we can set $A'^i_j(k,m) = \widehat{A}^i_j(k,m)$, $A'^j_j(k,m) = \widehat{A}^j_j(k,m)$ and $\zeta^i_j(k,m) = \zeta^j_j(k,m) = 1$. In the remainder the frequency and time indices will be omitted for brevity.

3. PAIRWISE DIRECT-AMBIENCE-DECOMPOSITION

To extract the direct and ambient signal parts, the signals of each segment undergo a pairwise direct-ambience decomposition which results in four signals per segment: an estimate of the direct and an estimate of the ambient components per input signal. For the ambience energy extraction, we chose the method proposed in [11] which leads to the ambience energy estimate

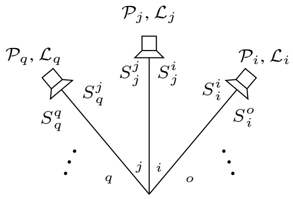

$\mathcal{P}_j, \mathcal{L}_j$
$\mathcal{P}_q, \mathcal{L}_q$
$\mathcal{P}_i, \mathcal{L}_i$
S^j_j
S^i_j
S^j_q
S^q_q
S^i_i
S^o_i
q
j
i
o

Fig. 2. Section of the loudspeaker setup depicting loudspeaker positions $\mathcal{P}$, loudspeaker driving signals $\mathcal{L}$, segment signals S and segments o, i, j, q with $o = (i-1)\%I$ and $q = (i+2)\%I$

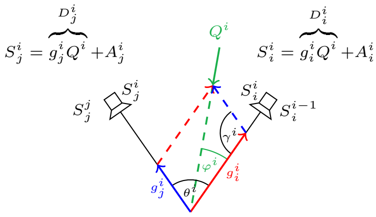

$S^i_j = \overbrace{g^i_j Q^i}^{D^i_j} + A^i_j$
Q^i
$S^i_i = \overbrace{g^i_i Q^i}^{D^i_i} + A^i_i$
S^j_j
S^i_j
S^i_i
S^{i-1}_i
γ^i
φ^i
g^i_i
g^i_j
θ^i

Fig. 3. Entities used for DOA estimation within a segment [13].

$$\Phi_{\widehat{A}^i} = \frac{1}{2}\left(\Phi_{S^i_i} + \Phi_{S^i_j} - \sqrt{(\Phi_{S^i_i} - \Phi_{S^i_j})^2 + 4\left|r_{S^i_i S^i_j}\right|^2}\right), \quad (4)$$

assuming equal ambience energies in both input signals, where $\Phi_X = \mathrm{E}\left\{|X|^2\right\}$ denotes the power spectral density (PSD) of a signal X, r_{XY} denotes the covariance of the signals X and Y, $\mathrm{E}\left\{\cdot\right\}$ denotes the mathematical expectation operator, and $|\cdot|$ is the magnitude operator. For detailed information on this method, the reader is referred to [11]. With (4), we can define the ambient- and direct-to-total power ratios of the input signals:

$$\Omega^i_i := \frac{\Phi_{\widehat{A}^i}}{\Phi_{S^i_i}}, \qquad \Omega^i_j := \frac{\Phi_{\widehat{A}^i}}{\Phi_{S^i_j}} \quad (5)$$

$$\Psi^i_i := \frac{\Phi_{\widehat{D}_i}}{\Phi_{S^i_i}} = 1 - \Omega^i_i, \qquad \Psi^i_j := \frac{\Phi_{\widehat{D}_j}}{\Phi_{S^i_j}} = 1 - \Omega^i_j. \quad (6)$$

The direct- and ambient signal parts can then be calculated according to

$$\widehat{A}^i_i = \sqrt{\Omega^i_i} \cdot S^i_i, \qquad \widehat{A}^i_j = \sqrt{\Omega^i_j} \cdot S^i_j \quad (7)$$

$$\widehat{D}^i_i = (1 - \sqrt{\Omega^i_i}) \cdot S^i_i, \qquad \widehat{D}^i_j = (1 - \sqrt{\Omega^i_j}) \cdot S^i_j, \quad (8)$$

which assures $\tilde{\mathcal{L}}_i = \mathcal{L}_i$ if no setup modification has taken place.

4. DOA ESTIMATION AND RE-PANNING

4.1. Direction of Arrival Estimation

The direct-ambiance decomposition provides estimates Ψ^i_i and Ψ^i_j of the direct-to-total power ratios for segment i. These ratios can be used to determine a DOA estimate of a phantom source within the considered segment (see [13] for details). We consider a phantom source signal Q^i which is panned to the angle φ^i between loudspeaker positions corresponding to segment i as illustrated in Figure 3. The superscript indicating the considered segment will be omitted within this section to prevent confusion with exponentials. It is assumed that the phantom source had been panned using vector base amplitude panning (VBAP) [14]. The estimated direct signal components can be substituted by the scaled phantom source signal and the segment signal powers can be modeled as

$$\begin{aligned} \Phi_{L_i} &= \overbrace{g_i^2 \Phi_Q}^{\Phi_{\widehat{D}_i}} + \Phi_{\widehat{A}} \\ \Phi_{L_j} &= \underbrace{g_j^2 \Phi_Q}_{\Phi_{\widehat{D}_j}} + \Phi_{\widehat{A}}, \end{aligned} \tag{9}$$

where g_i and g_j denote the panning gains. Using (6) and (9), the ratio of the direct-to-total signal powers can be expressed as

$$\frac{\Psi_i}{\Psi_j} = \frac{g_i^2}{g_j^2} \cdot \frac{\Phi_{L_j}}{\Phi_{L_i}}. \tag{10}$$

With the relation $g_i^2 + g_j^2 = 1$, (10) can be solved for g_i and g_j, leading to $g_i = (\frac{\Psi_i \Phi_{L_i}}{\Psi_i \Phi_{L_i} + \Psi_j \Phi_{L_j}})^{0.5}$ and $g_j = (\frac{\Psi_j \Phi_{L_j}}{\Psi_i \Phi_{L_i} + \Psi_j \Phi_{L_j}})^{0.5}$. The corresponding DOA can be obtained using the law of cosines which leads to [13]

$$\varphi = \cos^{-1}\left(\frac{1 - 2g_i g_j \cos(\gamma) + g_i^2 - g_j^2}{2g_i\sqrt{1 - 2g_i g_j \cos(\gamma)}}\right), \tag{11}$$

where $\gamma = 180 - \theta$ and θ is the aperture angle of the corresponding segment.

4.2. Re-Panning

Knowing the DOA as well as the panning gains corresponding to the phantom source within each segment, it is possible to adjust the phantom sources' positions with respect to the actual reproduction setup. Let us consider the loudspeaker setup, as illustrated in Figure 4, where the left and right front speakers were displaced from their ideal positions $\mathcal{P}_1$ and $\mathcal{P}_4$ at ±30° azimuth to suboptimal positions $\tilde{\mathcal{P}}_1$ and $\tilde{\mathcal{P}}_4$ at ±45° azimuth, i.e., an enlargement of segment $i = 0$ in positive and segment $i = 4$ in negative angular direction. Please note for the re-panning, we consider only one active source in this paper. Dependent on the phantom source's position, we can distinguish three processing paradigms as indicated in Figure 4 by different colors and filling. The formal processing paradigms for a segment i can be found in Table 1 and are qualitatively described as follows.

◁ This processing paradigm applies to unaltered segments, to shrunk segments and to those positions of an enlarged segment which overlap with the original segment. The corresponding direct signals are re-panned, i.e., the direct signal in each loudspeaker is normalized according to their panning gains with respect to the ideal loudspeaker setup and afterwards scaled according to the panning gains with respect to the modified setup.

◀ This paradigm applies to added positions of an in positive direction enlarged segment. In addition to the re-panning, a reallocation of the phantom source needs to take place, e.g., from segment $\{\mathcal{P}_1, \mathcal{P}_2\}$ in the ideal setup to segment $\{\mathcal{P}_0, \tilde{\mathcal{P}}_1\}$ in the reproduction setup. This is done by setting the speaker signal corresponding to position $\tilde{\mathcal{P}}_2$ to zero and copying direct signal parts to the speaker signal at position $\tilde{\mathcal{P}}_0$ including a proper re-panning according to the modified speaker position.

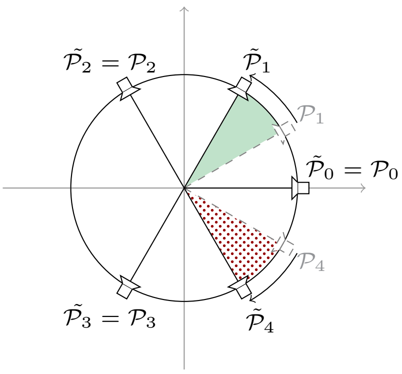

$\tilde{\mathcal{P}}_2 = \mathcal{P}_2$
$\tilde{\mathcal{P}}_1$
$\mathcal{P}_1$
$\tilde{\mathcal{P}}_0 = \mathcal{P}_0$
$\mathcal{P}_4$
$\tilde{\mathcal{P}}_3 = \mathcal{P}_3$
$\tilde{\mathcal{P}}_4$

Fig. 4. Processing paradigms for a suboptimal reproduction setup dependent on the analyzed position of a phantom source (speakers at ±30° were moved from their ideal positions to ±45°).

◁	$\tilde{\mathcal{L}}_i$:	$\eta_i^o = \eta_i^i$	$\eta_i^i = \frac{\tilde{g}_i^i(\varphi^i)}{g_i^i(\varphi^i)}$	${D'}_i^o = \widehat{D}_i^o$	${D'}_i^i = \widehat{D}_i^i$
$i = 0..4$	$\tilde{\mathcal{L}}_j$:	$\eta_j^i = \frac{\tilde{g}_j^i(\varphi^i)}{g_j^i(\varphi^i)}$	$\eta_j^j = \eta_j^i$	${D'}_j^i = \widehat{D}_j^i$	${D'}_j^j = \widehat{D}_j^j$
◀	$\tilde{\mathcal{L}}_i$:	$\eta_i^o = \eta_i^i$	$\eta_i^i = \frac{\tilde{g}_i^i(\varphi^j)}{g_j^j(\varphi^j)}$	${D'}_i^o = {D'}_i^i$	${D'}_i^i = \widehat{D}_j^j$
$i = 0$	$\tilde{\mathcal{L}}_j$:	$\eta_j^i = \frac{\tilde{g}_j^i(\varphi^j)}{g_j^j(\varphi^j)}$	$\eta_j^j = \eta_j^i$	${D'}_j^i = \widehat{D}_j^j$	${D'}_j^j = {D'}_j^i$
	$\tilde{\mathcal{L}}_q$:	$\eta_q^j = 0$	$\eta_q^q = 0$	${D'}_q^j = 0$	${D'}_q^q = 0$
▨	$\tilde{\mathcal{L}}_o$:	$\eta_o^p = 0$	$\eta_o^o = 0$	${D'}_o^p = 0$	${D'}_o^o = 0$
	$\tilde{\mathcal{L}}_i$:	$\eta_i^o = \eta_i^i$	$\eta_i^i = \frac{\tilde{g}_i^i(\varphi^o)}{g_i^o(\varphi^o)}$	${D'}_i^o = {D'}_i^i$	${D'}_i^i = \widehat{D}_i^o$
$i = 4$	$\tilde{\mathcal{L}}_j$:	$\eta_j^i = \frac{\tilde{g}_j^i(\varphi^o)}{g_i^o(\varphi^o)}$	$\eta_j^j = \eta_j^i$	${D'}_j^i = \widehat{D}_j^o$	${D'}_j^j = {D'}_j^i$

Table 1. Formal re-panning processing paradigms if a phantom source is located at positions indicated by the respective filling in Figure 4, where $p = (i - 2)\%I$.

▨ This paradigm applies to added positions of an in negative direction enlarged segment. The phantom source needs to be reallocated, e.g., from segment $\{\mathcal{P}_3, \mathcal{P}_4\}$ to segment $\{\mathcal{P}_3, \tilde{\mathcal{P}}_4\}$. A similar processing paradigm as at the previous considered positions has to be applied but since the speaker is displaced in the opposite direction, the processing formally differs.

5. SUBJECTIVE LISTENING TEST

The proposed re-panning method was evaluated using an experiment with respect to changes in localization. Three different conditions were defined:

C30 5.1 loudspeaker arrangement with front left and front right speakers at ±30°.

C45 5.1 loudspeaker arrangement with front left and front right speakers at ±45°.

C45RP 5.1 loudspeaker arrangement with front left and front right speakers at ±45°. Before reproduction, the output signals were processed by our proposed re-panning method.

The hypothesis of the experiment is that the reproduction of stimuli results in a smaller localization error for C45RP than C45 in comparison to the reference position which is the perceived location of the stimuli reproduced by C30.

Participants: Twenty-one participants (16 males, 5 females) ranging in age from 22 to 38 ($M = 27.6$ years, $SD = 3.8$)[1], volunteered to participate in the experiment. Eighteen participants reported to be professionals in audio, where five reported to be also experts in spatial audio and five reported to be experts in timbre. Only one participant reported never having taken part in a listening test before.

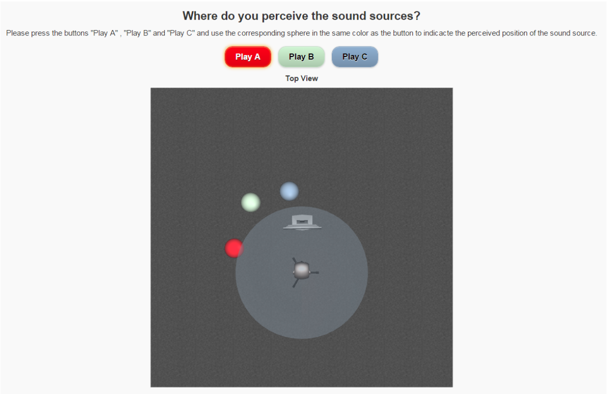
Stimuli: The stimuli consisted of a five-second pulsed pink noise signal (peak = -8.5 dB, crest factor = 14.7 dB) which was panned to twelve azimuth angles within a 5.1 loudspeaker setting: $\phi_{\text{ref}} = [0, -7, 15, -21, 30, -37, 45, -58, 71, -84, 97, -110]$. The pulses were 215 ms long, including a 5 ms and 10 ms long attack and decay time followed by 100 ms of silence.

Material and Apparatus: The experiment took place in a soundproof listening room with measurements (H x W x D) 256 x 455 x 452 cm and an average reverberation time (RT60) of 0.13 s. A 5.1 loudspeaker setup (Focal cms 40) as indicated in Figure 4 with $\mathcal{P}_0 = 0°$, $\mathcal{P}_{1,4} = \pm 30°$, $\mathcal{P}_{2,3} = \pm 110°$ and radius of 1.9 m was used during the experiment. To realize the displaced speakers, another speaker pair at $\tilde{\mathcal{P}}_{1,4} = \pm 45°$ was added. The C30 condition used the loudspeakers $\mathcal{P}_0$ to $\mathcal{P}_4$, whereas conditions C45 and C45RP used the loudspeakers $\mathcal{P}_0, \tilde{\mathcal{P}}_1, \mathcal{P}_2, \mathcal{P}_3, \tilde{\mathcal{P}}_4$.

The listening position, in the middle of the room, provided a chair for the participants with a small table in front of it on which a 24″ widescreen LCD monitor was mounted. A black-colored, 360° masking curtain made of deco-molton was fixed to an aluminum ring with a diameter of 2 m and attached to three truss stands at a height of 212 cm to veil the loudspeakers. The curtain attenuates frequencies above 300 Hz by about 2 dB. The lighting in the room was adjusted such that participants could not spot the loudspeakers beyond the curtain. The loudness of the stimuli was calibrated with a measurement microphone (Brüel&Kjær Type 4189-A-021) and pink noise (peak = -0.7 dB, crest factor = 12.8 dB) to 65 dBA SPL for each loudspeaker at the listening position.

Participants reported the location of the stimuli using a revised version of a 2D-based graphical user interface (GUI) which was evaluated in [15]. It showed a single orthographic view of a virtual scene representing the room the participants were sitting in. The virtual scene was true to scale and contained the participant's head, a monitor, the masking curtain and three colored spheres which the participants used to indicate the perceived locations of the stimuli. Figure 5 depicts a screenshot of the 2D-based GUI.

Procedure: The experiment had a subject-within design where every participant localized twelve stimuli for each condition. Thus, each participant had to give 36 responses. All participants were guided by an experimenter to the chair in the middle of the room in a way that they could not spot the loudspeakers while entering the room. Then, the experimenter left the room and all subsequent instructions were given by the experiment software. Starting with a questionnaire, the participants were asked whether they took part in a listening test before, whether they are an audio professional, whether they are expert listeners in timbre, whether they where expert listeners in spatial audio and their age. To get familiar with the GUI and the localization task, participants had to undergo a training. The training consisted of two trials where in each trial three stimuli were presented which had to be localized by placing the three colored spheres at the corresponding positions in the virtual scene. Head movements where allowed during localizing the stimuli. A trial was accomplished if each stimulus was played back and each sphere was placed within a range around the stimulus' actual location. After the training, participants proceeded with the actual localization task which started by presenting the instructions, shown before the training, again. Subsequently, the participants had to localize 36 stimuli in twelve trials, where the sequence of the trials was randomly chosen. At the end of the experiment, the participants filled out another questionnaire where feedback could be given to the experimenters.



Where do you perceive the sound sources?

Please press the buttons "Play A" , "Play B" and "Play C" and use the corresponding sphere in the same color as the button to indicacte the perceived position of the sound source.

Fig. 5. Screenshot of the GUI used for reporting the location of the stimuli.

6. RESULTS

The reported azimuth location of a stimulus is defined as $\phi_R(c, s, p)$ where $c \in \{\text{C30}, \text{C45}, \text{C45RP}\}$ denotes the condition, s denotes the stimulus index and p denotes the participant index. The total number of stimuli is defined as S and the total number of participants is defined as P. The vector containing all absolute localization errors with respect to the reference position of a phantom source for condition c is defined as

$$\epsilon_{\text{ref}}(c) = \begin{bmatrix} |\phi_R(c,1,1) - \phi_{\text{ref}}(1)| \\ \cdots \\ |\phi_R(c,1,P) - \phi_{\text{ref}}(1)| \\ \cdots \\ |\phi_R(c,S,P) - \phi_{\text{ref}}(S)| \end{bmatrix}. \tag{12}$$

The absolute localization error between two conditions c_1 and c_2 is defined as

$$\epsilon_C(c_1, c_2) = \begin{bmatrix} |\phi_R(c_1,1,1) - \phi_R(c_2,1,1)| \\ \cdots \\ |\phi_R(c_1,1,P) - \phi_R(c_2,1,P)| \\ \cdots \\ |\phi_R(c_1,S,P) - \phi_R(c_2,S,P)| \end{bmatrix}. \tag{13}$$

Additionally, two subsets, indicated by $(\cdot)^{\text{Fr}}$ ('front') and $(\cdot)^{\text{Ba}}$ ('back'), are defined. The former only contains responses corresponding to stimuli where $|\phi_{\text{ref}}| \leq 45°$ and the latter corresponds to stimuli where $|\phi_{\text{ref}}| > 45°$.

In average, the individual experiment duration was 11.2 min ($SD = 5.4$). The mean absolute localization error of $\epsilon_{\text{ref}}(\text{C30})$ was 13.1 ° ($SD = 7.7$) and the mean of the two subsets $\epsilon_{\text{ref}}^{\text{Fr}}(\text{C30})$ and $\epsilon_{\text{ref}}^{\text{Ba}}(\text{C30})$ was 10.6 ° ($SD = 4.4$) and 16.6 ° ($SD = 9.7$), respectively. To answer the experiment hypothesis, the reported

[1] M = mean, SD = standard deviation.

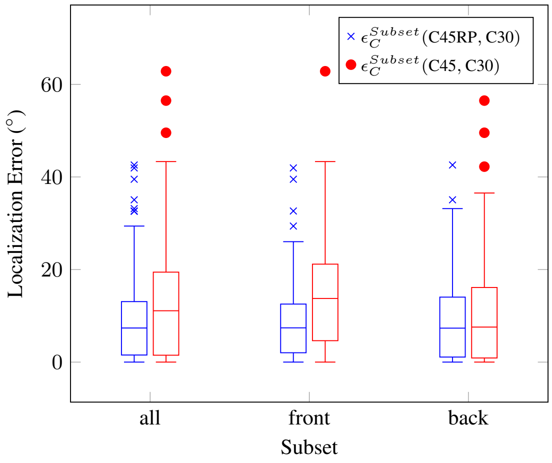


Fig. 6. A boxplot of the localization errors.

locations using conditions C45RP and C45 are compared to the reported locations of condition C30. The mean absolute localization errors between conditions C45RP and C30, i.e, $\epsilon_C(\text{C45RP}, \text{C30})$, was 8.7 ° ($SD = 8.4$), the mean for $\epsilon_C(\text{C45}, \text{C30})$ was 12.5 ° ($SD = 11.4$). Comparing both means reveals the proposed method to improve the localization by 3.8 °.

For the subsets, the mean of $\epsilon_C^{\text{Fr}}(\text{C45RP}, \text{C30})$ was 8.7 ° ($SD = 8.0$) and $\epsilon_C^{\text{Fr}}(\text{C45}, \text{C30})$ resulted in a mean of 13.8 ° ($SD = 11.0$). Especially for phantom sources positioned to the front, the proposed re-panning method improves the localization on average by 5.1 °. As expected, the improvements become smaller for phantom sources positioned at rear, since the localization blur dominates the responses: the mean of $\epsilon_C^{\text{Ba}}(\text{C45RP}, \text{C30})$ was 8.7 ° ($SD = 9.0$) and the mean of $\epsilon_C^{\text{Ba}}(\text{C45}, \text{C30})$ was 10.7 ° ($SD = 11.8$). Figure 6 depicts a boxplot showing the localization errors for all responses and responses for the two subsets. One might wonder why the localization errors of $(\cdot)^{\text{Fr}}$ are in the same range as the localization errors of $(\cdot)^{\text{Ba}}$ since the human localization blur is more present towards the rear. A major reason for increased location errors of frontal phantom sources is that they were mainly reproduced by the left and right speakers which were moved by 15°. E.g., a frontal phantom source placed at +30° was reproduced by almost only the right speaker resulting in an additional localization error of 15° compared to a phantom source placed at +110°.

A further analysis is applied to verify whether the differences between C45RP and C45 are statistically significant (the significance level α is defined as 0.05 in this paper). A Q-Q plot analysis showed, the responses, including the subsets, are not normally distributed. As the large differences in standard deviation between the conditions indicated, Levene's test for equal variances was found to be violated for all responses ($F(1, 502) = 22.5, p = .000$) as well as for the two subsets Fr ($F(1, 292) = 15.2, p = .000$) and Ba ($F(1, 208) = 4.9, p = .028$). To verify the differences of the means being statistically significant, a not-equal variance assuming paired t-test was applied. The t-test results show significant differences for the whole data set ($t(251) = 7.0, p = .000$) as well as for the two subsets 'front' ($t(146) = 7.3, p = .000$) and 'back' ($t(104) = 2.4, p = .019$). As the data is not normal distributed, a non-parametric Wilcoxon signed-rank test was applied to confirm the t-test results. It also indicated significant differences between the whole data set ($Z = 6.92, p = .000, r = 0.31$) as well as for the two subsets 'front' ($Z = 6.683, p = .000, r = 0.39$) and 'back' ($Z = 2.48, p = .013, r = 0.17$).

7. CONCLUSION

A segment-based re-panning method was proposed and evaluated for directional signals. The re-panning method estimates the DOAs within each segment utilizing direct-ambience decomposition and re-renders direct signal parts with respect to the actual reproduction loudspeaker setup. In a localization listening test, participants were asked to locate stimuli presented over a 5.1 surround setup, a modified surround setup and a modified surround setup with active re-panning processing. In the modified setups, the front speaker positions were altered to ±45°. A comparison of the responses showed that the proposed re-panning method improved the overall localization on average by 3.8°. If only positions in the front of the setup are considered, the improvement increases on average by 5.1°.

8. ACKNOWLEDGMENT

The authors would like to thank Emanuël A. P. Habets for constructive discussions during the writing of this paper.

REFERENCES

[1] ITU-R BS.775-2, "Multichannel Stereophonic Sound System With And Without Accompanying Picture," 07/2006.

[2] K. Hamasaki, K. Hiyama, and R. Okumura, "The 22.2 Multichannel Sound System and Its Application," in *118th Convention of the AES*, Bacelona, Spain, 2005.

[3] C. Eggers. (2014) Dolby Atmos for the Home. [Online]. Available: http://cdn-blog.dolby.com/wp-content/uploads/2014/08/Dolby-Atmos-for-the-Home-Theater.pdf

[4] A. Ando, "Conversion of Multichannel Sound Signal Maintaining Physical Properties of Sound in Reproduced Sound Field," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 6, pp. 1467–1475, 2011.

[5] V. Pulkki, "Spatial Sound Reproduction with Directional Audio Coding," *J. Audio Eng. Soc*, vol. 55, no. 6, pp. 503–516, 2007.

[6] V. Pulkki and J. Herre, "Method and Apparatus for Conversion Between Multi-Channel Audio Formats," US Patent US 2008/0 232 616 A1, 2008.

[7] M. Goodwin and J.-M. Jot, "Primary-Ambient Signal Decomposition and Vector-Based Localization for Spatial Audio Coding and Enhancement," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 1, 2007, pp. I–9 –I–12.

[8] J.-M. Jot, V. Larcher, and J.-M. Pernaux, "A Comparative Study of 3-D Audio Encoding and Rendering Techniques," in *16th International Conference of the AES*, Rovanieme, Finland, 1999.

[9] M. M. Goodwin and J.-M. Jot, "Multichannel Surround Format Conversion and Generalized Upmix," in *30th International Conference of the AES*, Saariselkä, Finland, 2007.

[10] C. Faller, "Multiple-Loudspeaker Playback of Stereo Signals," *J. Audio Eng. Soc*, vol. 54, no. 11, pp. 1051–1064, 2006.

[11] J. Merimaa, M. M. Goodwin, and J.-M. Jot, "Correlation-Based Ambience Extraction from Stereo Recordings," in *123rd Convention of the AES*, New York, NY, 2007.

[12] C. Avendano and J.-M. Jot, "Ambience extraction and synthesis from stereo signals for multi-channel audio up-mix," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 2, 2002, pp. II–1957 –II–1960.

[13] A. Adami and J. Herre, "Evaluation of a Frequency-Domain Source Position Estimator for VBAP-panned Recordings," in *138th Convention of the AES*, Warsaw, Poland, 2015.

[14] V. Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning," *J. Audio Eng. Soc*, vol. 45, no. 6, pp. 456–466, 1997.

[15] M. Schoeffler, S. Westphal, A. Adami, H. Bayerlein, and J. Herre, "Comparison of a 2D- and 3D-Based Graphical User Interface for Localization Listening Tests," in *Proceedings of the EAA Joint Symposium on Auralization and Ambisonics*, Berlin, Germany, 2014, pp. 107–112.