Smoothing individual head-related transfer functions in the frequency and spatial domains

Eugen Rasumow,[a] Matthias Blau, and Martin Hansen
Jade Hochschule, Institut für Hörtechnik und Audiologie, Ofener Straße 16-19, 26121 Oldenburg, Germany

Steven van de Par, Simon Doclo, and Volker Mellert
Universität Oldenburg, Department für Medizinische Physik und Akustik, Exzellenzcluster Hearing4All, Carl-von-Ossietzky-Straße 11, 26129 Oldenburg, Germany

Dirk Püschel
Akustik Technologie Göttingen, Bunsenstraße 9c, 37073 Göttingen, Germany

(Received 2 January 2013; revised 9 February 2014; accepted 13 February 2014)

When re-synthesizing individual head related transfer functions (HRTFs) with a microphone array, smoothing HRTFs spectrally and/or spatially prior to the computation of appropriate microphone filters may improve the synthesis accuracy. In this study, the limits of the associated HRTF modifications, until which no perceptual degradations occur, are explored. First, complex spectral smoothing of HRTFs into constant relative bandwidths was considered. As a prerequisite to complex smoothing, the HRTF phase spectra were substituted by linear phases, either for the whole frequency range or above a certain cut-off frequency only. The results indicate that a broadband phase linearization of HRTFs can be perceived for certain directions/subjects and that the thresholds can be predicted by a simple model. HRTF phase spectra can be linearized above 1 kHz without being detectable. After substituting the original phase by a linear phase above 5 kHz, HRTFs may be smoothed complexly into constant relative bandwidths of 1/5 octave, without introducing noticeable artifacts. Second, spatially smoother HRTF directivity patterns were obtained by levelling out spatial notches. It turned out that spatial notches do not have to be retained if they are less than 29 dB below the maximum level in the directivity pattern.
 [http://dx.doi.org/10.1121/1.4867372]

PACS number(s): 43.66.Qp, 43.66.Pn, 43.64.Ha, 43.66.Ba [JFC] Pages: 2012–2025

I. INTRODUCTION

Spatial information is a major factor in the perception and appraisal of sounds. A typical way to include spatial information into recordings and measurements is to use so-called artificial heads, which are reproductions of real human heads with microphones placed in the ear canals [cf. Paul (2009) for a recent overview]. Alternatively, the direction- and frequency-dependent head-related transfer functions (HRTFs) can be approximately re-synthesized using a set of spatially distributed microphones with appropriate digital filtering [cf. Chen *et al.* (1992), Tohtuyeva and Mellert (1999), Kahana *et al.* (1999), Atkins (2011), and Rasumow *et al.* (2011, 2013)]. Such a device is referred to as a virtual artificial head (VAH).

The performance of such a VAH not only depends on the microphone array (topology, number of microphones, calibration, mechanical stability) and the filter design procedure but also on the desired directivity patterns, i.e., the HRTFs to be re-synthesized. In general, more microphones are needed when the directivity patterns exhibit more spatial detail, or equivalently, the accuracy decreases given a fixed number of microphones.

This is illustrated in Fig. 1 where a sample HRTF set (IRCAM database[1] subject #1002, right ear) is re-synthesized using different numbers of microphones (the steering vectors of which were approximated by pure delays) and the methodology described in Rasumow *et al.* (2013). At $f = 1$ kHz (left diagram) the HRTF directivity is spatially smooth and can accurately be re-synthesized with $N = 8$ microphones, whereas at $f = 11$ kHz it becomes spatially more detailed and the re-synthesis with $N = 8$ microphones fails completely (center diagram). If the same directivity is re-synthesized with $N = 24$ microphones, an accurate fit is obtained again (right diagram).

For the VAH, one seeks to minimize the number of microphones because this not only reduces the cost of the system but also its sensitivity to, e.g., microphone gain, phase and position errors [cf. Rasumow *et al.* (2011)].

Hence, given the observation made above, a preprocessing step involving spatial smoothing of the HRTFs could improve the concept of a VAH in terms of the number of microphones needed. It needs to be assured then of course that the preprocessing does not cause perceptible degradations of the HRTFs. Therefore, we will investigate in this paper what type and amount of smoothing is still imperceptible.

Several types of smoothing of HRTFs will be considered in this study. One reasonable and well investigated way to smooth HRTFs is to reduce their spectral resolution, which has the indirect effect of obtaining spatially smoother

[a]Author to whom correspondence should be addressed. Electronic mail: eugen.rasumow@jade-hs.de

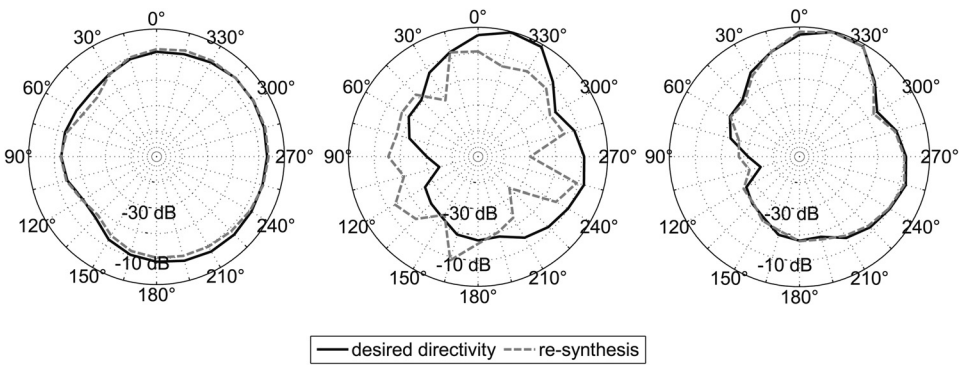

FIG. 1. Illustration of the interrelationship between the spatial complexity of HRTFs and the number of microphones needed to accurately re-synthesize them. The example HRTF set was taken from the IRCAM database (right ear of subject #1002), the original directivity is shown as solid line, the re-synthesis as dashed line. Left: $f = 1$ kHz $N = 8$ microphones, center: $f = 11$ kHz $N = 8$ microphones, right: $f = 11$ kHz $N = 24$ microphones.

directivity patterns. Spectral smoothing of HRTFs is often done by truncating the length of the corresponding head related impulse responses (HRIRs). Typically, filter lengths of 512 taps (corresponding to about 11.6 ms at a sampling frequency of $f_s = 44.1$ kHz) are considered to represent the individual cues of the HRTFs sufficiently well in order to get an externalized virtual directional perception [cf. Kulkarni *et al.* (1999), Huopaniemi *et al.* (1999), and Algazi *et al.* (2001)]. Truncating the length of the HRIRs corresponds to smoothing the HRTFs into constant absolute band-widths. From a psychoacoustic point of view, smoothing into constant relative bandwidths is preferable [cf. Breebaart and Kohlrausch (2001)] since it is a well known phenomenon that the human ear groups incoming sounds into frequency bands that broaden with increasing center frequencies ["critical bands," cf. Fletcher (1940), Patterson and Nimmo-Smith (1980), and Moore (2003)]. Also, compared to smoothing into constant absolute bandwidths, smoothing into constant relative bandwidths automatically results in more smoothing at higher frequencies, which in turn will result in smoother directivity patterns per (constant bandwidth) frequency bin at higher frequencies and is therefore beneficial to the VAH. The concept of smoothing HRTFs into constant relative bandwidths was, for instance, applied by Breebaart and Kohlrausch (2001) and Breebaart *et al.* (2010), who found that smoothing into approximately one critical band was acoustically transparent for spatial audio coding applications using non-individual HRTFs, or by Xie and Zhang (2010) who even proposed to smooth into up to 3.5 equivalent rectangular bandwidths (ERBs) at higher frequencies ($f > 5000$ Hz) for the contralateral ear. However, smoothing into such broad bandwidths is assumed to lead to discriminable artifacts for the ipsilateral ear and generally at lower frequencies ($f < 5000$ Hz).

Although smoothing into constant relative bandwidths is psychoacoustically appealing, it may also produce inconsistencies between magnitude and phase spectra. For instance, when a complex spectral smoothing is applied at frequencies above about 5 kHz, the noisy and/or steep phase characteristics of measured HRTFs result in notches in the magnitude of the complexly smoothed HRTF, which are not plausible, cf. Fig. 2 (dashed line). Most often this problem is circumvented by smoothing the magnitude and the phase of measured HRTFs separately [cf. Kulkarni and Colburn (1998), Breebaart and Kohlrausch (2001), and Breebaart *et al.* (2010)] or by smoothing the magnitude only and supplementing it with a minimum-phase reconstruction [cf. Huopaniemi and Karjalainen (1996)]. However, a separate manipulation of magnitude and phase may result in unwanted interaural cues. As an alternative, we propose to first simplify the phase spectrum in a perceptually transparent way and to subsequently smooth magnitude and phase spectra *simultaneously* (complex smoothing) into constant relative bandwidths.

In order to obtain perceptually transparent phase simplifications of original HRTFs, the minimum-phase-plus-delay approach has been proposed [cf. Mehrgardt and Mellert (1977) and Kulkarni *et al.* (1999)]. This approach is now widely used, most often (if not always) motivated by the listening tests carried out by Kulkarni *et al.* (1999). Interestingly though, the results obtained by Kulkarni *et al.* (1999) already indicate that a linear phase might be a perceptually better choice than a minimum phase and that *at low frequencies* it is worthwhile "making the overall low-frequency ITD in the model HRTFs to be the same as the overall low-frequency ITD in the empirical HRTFs," which in fact questions the validity of a broadband phase substitution. Therefore, it appears to be necessary to re-investigate the perceptual limits of a broadband phase substitution with regard to the VAH.

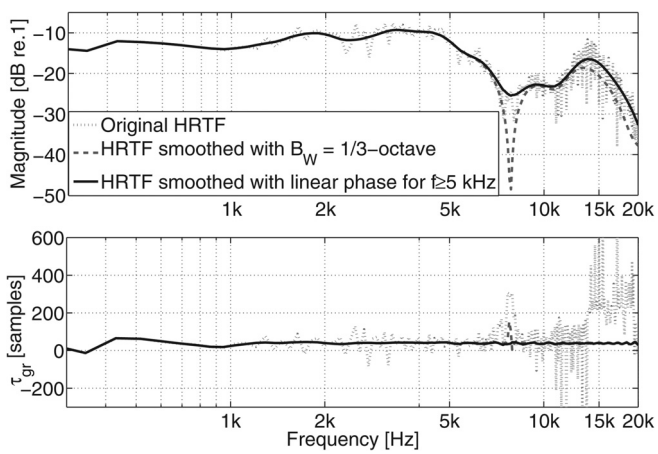

FIG. 2. Magnitude in dB (top) and group delay in samples (bottom) for an exemplary HRTF of subject S2 for the left ear from azimuth $\theta = 210°$. The dot-dashed lines show the original HRTF (NFFT = 512), the dashed lines show the resulting HRTF after complex smoothing into third octave bands, the solid black lines show the smoothed HRTF (third octave bands) when the phase of the HRTF is substituted by a linear phase above 5 kHz before smoothing.

As stated above, we hypothesize that primarily the *spatial* smoothness of HRTFs determines the effort needed to re-synthesize HRTFs with a VAH. Hence, a direct smoothing in the spatial domain may be even more profitable than a smoothing in the frequency domain, which would lower the spatial dynamic range only indirectly.

One well-known method to obtain spatially smoother HRTFs is to model the HRTFs through a spherical or cylindrical harmonic decomposition of finite order, cf. Duraiswami *et al.* (2004), Zotkin *et al.* (2009), and Atkins (2011). This, however, has been shown to require an extremely large number of measurement directions and microphones for appropriate model orders [cf. Zotkin *et al.* (2009) and Castaneda *et al.* (2013)]. Even if this additional effort is disregarded, the question of perceptual relevance remains and is more fundamental than the question of an appropriate spatial model structure. Therefore, we address the perceptual relevance without assuming any spatial model structure in terms of spherical or cylindrical harmonics.

Instead, we assume that the (spatial) dynamic range of narrowband directivity patterns can be limited by appropriately levelling out spatial notches. This in turn is motivated by assuming that spatial notches in the directivity patterns are of less perceptual relevance than spatial peaks, which seems a reasonable starting point given that spatial notches in measured HRTFs may become as low as −80 dB compared to the dominant direction (cf. Fig. 5). More specifically, spatial notches mainly occur at contralateral directions and may therefore be masked by stronger ipsilateral components. Hence, a selective limitation of the spatial dynamic range for directions associated with lower levels in the directivity pattern may introduce only slight artifacts while decreasing the number of microphones needed in the VAH. We will therefore investigate to what extent the spatial notches can be reduced in depth without perceptual effects.

In summary, a successful preprocessing of HRTFs for a VAH should reduce the spatial complexity of narrowband directivity patterns associated with the HRTFs, without being perceptually distinguishable from the original HRTFs. This can be obtained indirectly by smoothing the HRTFs in the frequency domain or directly by limiting the spatial dynamic range of narrowband directivity patterns (which in turn is facilitated by a limited spectral resolution). In the following four experiments, the associated HRTF manipulations are evaluated with the aim of finding limits until which the original HRTFs could be altered without perceptible effects. In order to guarantee the highest flexibility in future VAH applications, all possible cues that the auditory system is able to access, including coloration, loudness, timbre, localization, etc., are considered. It should be noted that this is the most stringent criterion possible to discrimination tasks, which in some other applications might be relaxed [cf. Kulkarni and Colburn (1998) and Romigh *et al.* (2013)].

Section II introduces and explains the performed experiments investigating the reduction of the spatial dynamic range and the complex smoothing as well as two experiments examining the discriminability of phase linearizations. The applied stimuli and methods as to investigate these smoothing methods and phase manipulations are presented in Sec. III. The results of these experiments are presented in Sec. IV and discussed in Sec. V, including the consequences of the examined types of smoothing for the re-syntheses using the VAH. The main findings and conclusions for an acoustically transparent preprocessing of HRTFs according to the VAH are summarized in Sec. VI.

II. HRTF SMOOTHING OPERATIONS CONSIDERED IN THIS STUDY

First, complex smoothing of HRTFs in the frequency domain into constant relative bandwidths was investigated. As discussed above, complex smoothing into constant relative bandwidths requires a prior treatment of phase spectra. Inspired by the results from Kulkarni *et al.* (1999), we chose a linear phase model as a substitute for the original HRTF phase spectra. The linear phase ϕ_{lin} was in this case calculated from the delay of the maximum of the Hilbert envelope of the respective HRIR, cf. Appendix A. Informal listening tests indicated that a broadband substitution of the original phase spectra by the such computed linear phase can indeed be perceptually distinguished from the original HRTFs. Therefore, in a first experiment, we investigated the sensitivity of listeners to a broadband substitution of the original phase by a linear phase (cf. Sec. III C 1). More specifically, we tested to which extent listeners are sensitive to a broadband change of the original phase spectrum when fading the (unwrapped) original phase ϕ_{orig} into the linear phase ϕ_{lin} using a variable mixing gain L_ϕ between 0 and 1 such that

$$\phi_{\mathrm{test}}(f) = L_\phi \phi_{\mathrm{lin}}(f) + (1 - L_\phi)\phi_{\mathrm{orig}}(f). \tag{1}$$

The threshold value of the mixing gain L_ϕ then provides a means of characterizing the sensitivity of listeners to a broadband phase linearization of HRTFs: If listeners were insensitive to a broadband phase linearization, then L_ϕ should approach 1 whereas a low value of L_ϕ would indicate that listeners are sensitive to a broadband phase linearization.

Assuming that the results of the informal listening tests were confirmed in the first experiment (i.e., that listeners are sensitive to a broadband phase linearization), one may go a step further and, in line with the arguments by Kulkarni *et al.* (1999), apply the phase substitution at higher frequencies above a certain cutoff-frequency f_c only. Therefore, in the second experiment, the original phase ϕ_{orig} was maintained below a variable cutoff-frequency f_c and substituted by a linear phase ϕ_{lin} above f_c, resulting in

$$\phi_{\mathrm{test}}(f) = \begin{cases} \phi_{\mathrm{orig}}(f), & \text{for } f \le f_c \\ \phi_{\mathrm{lin}}(f), & \text{for } f > f_c. \end{cases} \tag{2}$$

If the cut-off frequency f_c is chosen too low, it may lead to audible changes. The threshold value of the cut-off frequency f_c as determined in this experiment then provides information about the range of cut-off frequencies that are allowable in the third experiment where complex spectral smoothing was applied.

In the third experiment, the noticeability of complex smoothing into constant relative bandwidths with a variable

bandwidth B_W was investigated (cf. Sec. III C 3). As a prerequisite the original phase was linearized for frequencies $f \geq f_c$. Then, broader bandwidths B_W yield smoother HRTFs but also result in more audible artifacts and vice versa. The threshold value of the bandwidth B_W can then be used to formulate a smoothing rule for the VAH. The complex smoothing into constant relative bandwidths was implemented according to a method proposed by Hatziantoniou and Mourjopoulos (2000). This manipulation is comparable to smoothing into ERBs for frequencies above approximately 1 kHz [Glasberg and Moore (1990)], which are the frequencies of primary interest in VAH applications. A detailed description of the applied method for a complex smoothing of HRTFs is given in Appendix B.

Second, we investigated the audibility of limiting the spatial dynamic range ζ directly in the spatial domain, i.e., over the directions of incidence. The spatial dynamic range ζ is given by

$$\zeta(f) = 20 \log_{10} \left(\frac{\max\limits_{\theta} |\mathrm{HRTF}(f,\theta)|}{\min\limits_{\theta} |\mathrm{HRTF}(f,\theta)|} \right) \ \mathrm{dB}. \tag{3}$$

The directivity patterns of the measured HRTFs often exhibit large spatial dynamic ranges ζ of up to about 80 dB (cf. Fig. 5). As discussed above, it is unlikely that such a large dynamic range is really exploited by the human auditory system. Therefore, in the fourth experiment, spatial notches were, frequency bin by frequency bin, reduced in depth such that the resulting spatial dynamic range ζ' would become lower than that of the original HRTF, cf. Sec. III C 4. The smaller the resulting spatial dynamic range, the better a VAH configuration with a fixed number of microphones could re-synthesize the HRTF, but at the same time, the more likely it is for subjects to spot a difference in comparison to the original HRTF. Hence, the discriminability of the artifacts introduced by a reduction of the spatial dynamic range was investigated in this experiment. The resulting threshold value of the spatial dynamic range ζ' could then be used to preprocess HRTFs prior to the design of appropriate VAH filters. In order to evaluate the effect of the reduction of the spatial dynamic range, we exploited the fact that any modification of the spatial dynamic range will cause modifications in the time and frequency domain for the affected directions. For a chosen direction, one can then evaluate the difference between the modified and the original HRTF. The extent of the difference depends on the individual subject, frequency, and direction. Since only a fixed number of directions could be tested, the directions with the biggest subjective differences were identified individually for each subject by a preliminary listening test, see Sec. III C 4.

In the following, listening tests aimed at evaluating the modifications outlined above are described in more detail.

III. METHODS

All experiments were performed with a fixed set of subjects, stimuli and individual HRTFs. The various experiments only differed in the variables to be investigated as described in Secs. III C 1 to III C 4.

A. Subjects

A total of eight normal hearing subjects (four male, four female, aged between 21 and 46 years) participated in the experiments. Four of the subjects were members of the scientific staff of the Institut für Hörtechnik und Audiologie. They had extensive experience with psychoacoustical experiments and participated voluntarily. Three of them are among the authors of this study. The remaining four subjects were students who were paid for their participation. All subjects completed at least one run of each experiment as familiarization, succeeded by three to six subsequent runs which were used for the evaluation. The performance of all tests lasted approximately five hours for each subject, with each session not lasting longer than 90 min.

All experimental procedures were approved by the ethics committee at the Carl-von-Ossietzky-Universität Oldenburg.

B. HRTF measurements

Individual HRTFs were measured in an anechoic room. The subject was seated in the center of a circular loudspeaker array consisting of 24 uniformly distributed loudspeakers (equidistant 15° spacing) at a radius of 1.25 m in the horizontal plane. In order to limit the risk of reflections by the experimental apparatus, small loudspeakers with diameter of 7 cm and a very light support structure were used.

Binaural HRTFs were measured for each direction using the blocked ear method [cf. Hammershoi and Moller (1996)] with custom-made ear shells and Knowles FG-23329 miniature electret microphones. The transfer functions were estimated using white noise signals and standard FFT-based techniques [H_1 estimate with 8192-point Hann window, 50% overlap, 52 averages, cf. Mitchell (1982)]. The measured transfer functions were subsequently divided by the free field transfer function derived from a measurement with the same loudspeakers and a calibration microphone (G.R.A.S. Microphone Type 40AF pointed towards 90° elevation) at the position of the center of the head, in order to obtain the (free field related) HRTFs. The latter step was also important to ensure that small differences between the loudspeakers frequency responses at different positions were compensated for. The corresponding HRIRs were truncated to 512 samples (corresponding to about 11.6 ms at a sampling frequency of $f_s = 44.1$ kHz). Furthermore, the tails of the HRIRs were flanked with a one-sided tapered Hann window with a descending flank of 50 samples (≈1.1 ms). The delay for sound propagation between the loudspeaker and the head was removed so that the obtained impulse responses had minimal initial delays. Immediately after the HRTF measurements, the headphone transfer functions (HPTF) were measured with the ear shells remaining in place. The HPTF measurements were repeated up to ten times per subject (the headphone was taken off and on before each repetition) until a dynamic range of less than 25 dB for frequencies $2\,\mathrm{kHz} \leq f \leq 16\,\mathrm{kHz}$ was obtained. The inverted version of this particular HPTF was then used as the individual HPTF equalization filter, implemented as finite impulse response (FIR) filters with a length of 512 samples (≈11.6 ms).

As a first check, white noise stimuli filtered with the individual HRTFs from 24 directions in the horizontal plane were played dichotically via headphones after individual headphone equalization to each subject. All subjects were able to perceive the presented stimuli outside the head and correctly assigned the corresponding direction.

In order to limit the number of experiments to a manageable amount, four directions were chosen for experiments 1–3, with azimuth angles: $\theta = 0°$ (front), $\theta = 90°$ (left), $\theta = 225°$ (back right), and $\theta = 315°$ (front right). These directions were chosen to cover the main variability of HRTFs, including very small ($\theta = 0°$) and rather large ($\theta = 90°$) interaural differences and the varying monaural cues for the lateral front ($\theta = 315°$) and back ($\theta = 225°$). The directions for the fourth experiment (investigating the spatial dynamic range directly) were selected individually, cf. Sec. III C 4.

C. Procedure and stimuli

All experiments were designed using PSYLAB,[2] a set of MATLAB-scripts for designing various psychoacoustical detection, discrimination, and matching experiments. The signals were presented binaurally via a D/A-converter (ADI-8 DS, RME Audio) and headphones (K-240 Studio, AKG Acoustics) at an overall sound pressure level of 78 dB, calibrated with the binaural filters for the frontal direction $\theta = 0°$ using an artificial ear (type 43AA, G.R.A.S. Sound & Vibration). During the experiments, the subjects were seated in the same anechoic room that was used for the HRTF measurement.

In order to cover a wide frequency range and to include temporal cues, the digitally generated test signal consisted of short bursts of frozen white noise with a spectral content of $150\,\text{Hz} < f < 18\,050\,\text{Hz}$. The band limitation was achieved by a multiplication with a tapered Hann window (between 150 and 200 Hz and 18 000 and 18 050 Hz) in the frequency domain (cf. lower graph in Fig. 3). This window was used to avoid pitch cues due to sharp edges in the frequency domain. The test signal consisted of three segments, each with a noise burst of 0.15 s with 0.001 s onset-offset ramps followed by silence of 0.15 s, leading to a total length of the test signal within each interval of 0.75 s (cf. upper graph in Fig. 3).

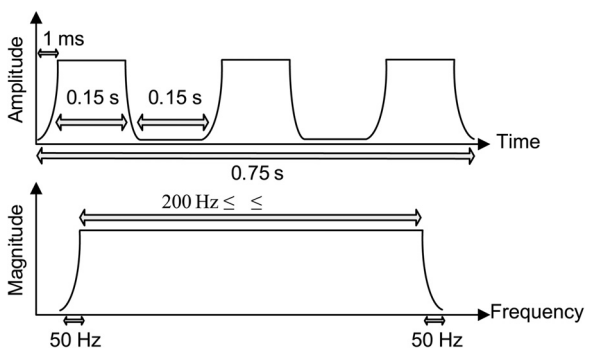

FIG. 3. Top: Temporal course of the used test signal consisting of three noise bursts with a total length of 0.75 s. Bottom: Spectral shape of the test signal with a constant magnitude for frequencies $200\,\text{Hz} \le f \le 18\,000\,\text{Hz}$.

A three alternative forced choice paradigm was applied to determine threshold values for the tested variables. Three intervals of filtered signals (each separated by 0.3 s silence) were presented to the subjects in a randomized order, where one interval contained the test signal filtered with the modified HRTFs and two intervals contained the test signal filtered with the original HRTFs (reference). Subjects were instructed to indicate the odd one of the three intervals. Feedback was presented after each trial. The modification of the variable parameter was adjusted adaptively according to the 1 up-1 down method, converging to a 50%-correct value on the psychometric function. This particular value on the psychometric function was chosen to represent the JND in all experiments. According to signal detection theory this threshold corresponds to a d' of 0.58 [cf. Hacker and Ratcliff (1979)]. The initial value, as well as the initial step size of the tested variable, were chosen for each experiment separately. In the familiarization phase of each experiment the step size of the tested variable was continuously halved at each upper reversal until a minimal step size was reached. As soon as the minimal step size was reached, the measurement phase started and the step size was kept constant. The threshold of the tested variable was defined as the median of six following reversals within the measurement phase. The specific details for each experiment are given in Secs. III C 1–III C 4.

1. Substitution of the individual HRTF phase by a mixture of the original phase and a linear phase

In this experiment, the original HRTF phase was substituted by $\phi_{\text{test}}(f)$, a mixture of the original phase $\phi_{\text{orig}}(f)$ and a linear phase $\phi_{\text{lin}}(f)$, cf. Eq. (1). The slope of the linear phase $\phi_{\text{lin}}(f)$ was computed by determining the delay of the maximum of the Hilbert envelope of the corresponding HRIR in the time domain (cf. Appendix A). The variable mixing gain L_ϕ ranged between $L_\phi = 0$ (original phase only) and $L_\phi = 1$ (linear phase only) and was independent of frequency.

The initial supra threshold value of the mixing gain was set to $L_\phi = 0.5$ and the initial step size to $\Delta L_\phi = 0.2$. The minimal step size was set to $\Delta L_\phi = 0.05$.

The length of the impulse response was kept constant throughout all experiments. No effort was made to compensate for acausalities or other artifacts associated with phase linearization.

2. Substitution of the individual HRTF phase by a linear phase above a certain cutoff-frequency

In this experiment, the original phase spectrum was preserved for frequencies $f \le f_c$, while the phase for frequencies $f > f_c$ was substituted by a linear phase [cf. Eq. (2)]. The linear phase ϕ_{lin} (constant phase slope) was applied starting at the frequency bin next to f_c (cf. Fig. 4), without any fading between ϕ_{orig} and ϕ_{lin} as, for instance, in Rasumow *et al.* (2012). Again, the slope of the linear phase was computed by determining the delay of the maximum of the Hilbert envelope of the corresponding HRIR (cf. Appendix A).

The initial supra threshold value of the cutoff-frequency was set to $f_c = 300\,\text{Hz}$ and the initial step size to $\Delta f_c = 160\,\text{Hz}$. The minimal step size was set to $\Delta f_c = 20\,\text{Hz}$.

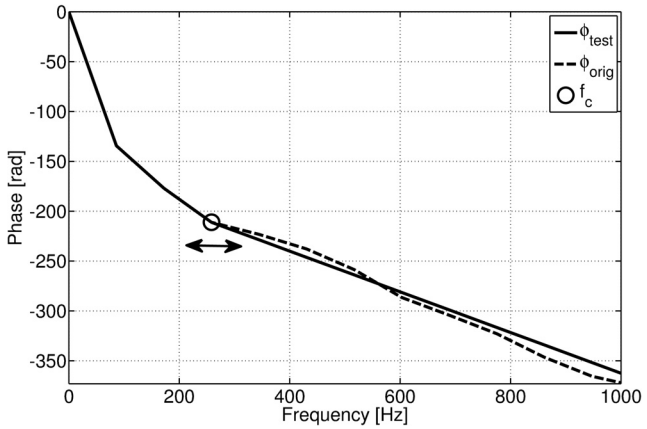

FIG. 4. Substitution of the original HRTF phase (ϕ_{orig}) by a linear phase (ϕ_{lin}) above a variable cutoff-frequency f_c (black circle): The resulting test phase ϕ_{test} is equal to ϕ_{orig} for $f \leq f_c$, and equal to ϕ_{lin} for $f > f_c$. The slope of ϕ_{lin} is computed from the maximum of the Hilbert envelope of the original HRIR.

3. Complex smoothing of the individual HRTFs into constant relative bandwidths

In this experiment, the individual HRTFs were smoothed complexly in the frequency domain using the smoothing algorithm proposed by Hatziantoniou and Mourjopoulos (2000). This smoothing algorithm achieves complex smoothing into constant relative bandwidths by replacing each frequency bin by a complex average of adjoining frequency bins within the respective bandwidth. Smoothing into constant relative bandwidths is equivalent to a truncation operation in the time domain with a frequency-dependent truncation length. To ensure that the smoothing algorithm processes the main parts of the HRIRs, it is therefore important to remove the overall delay from the HRIR before processing. Thus, the estimated overall delay of each HRIR was removed before smoothing and reconstructed afterwards (cf. Appendix B and Fig. 18).

As discussed above, complex-valued smoothing of HRTFs needs a prior treatment of phase spectra, which we chose to implement as a substitution of the original phase by a linear one above a certain cut-off frequency f_c, while maintaining the original phase below f_c. For practical reasons, the listening tests for the third experiment started before the second experiment (in which a threshold value for f_c was determined) had been completed. Therefore, a very conservative value of $f_c = 5$ kHz was chosen here, which proved to be much higher than the threshold value eventually determined by the second experiment, while at the same time being low enough to avoid difficulties related to steep and/or noisy phase spectra.

The initial supra threshold value of the relative bandwidth was set to $B_W = 2/3$ octave (approximately corresponding to the bandwidth of two auditory filters) and the initial step size to $\Delta B_W = 1/3$ octave. The minimal step size was set to $\Delta B_W = 1/24$ octave.

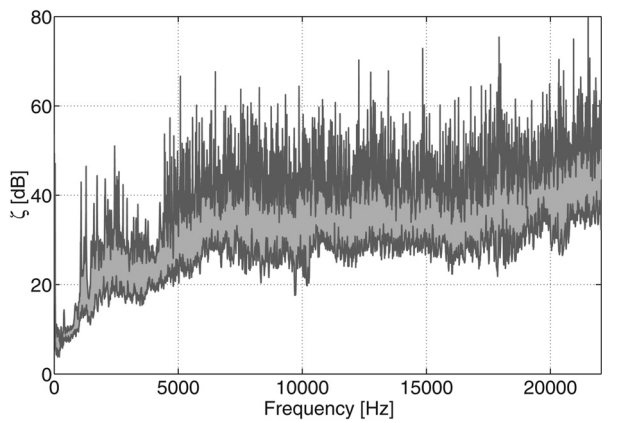

FIG. 5. Spatial dynamic range ζ of all left ear HRTFs (24 directions in the horizontal plane, all eight subjects), as a function of frequency. The minimum and maximum values of all subjects are shown as dark lines, the range in between (characterizing the remaining dynamic ranges) as a gray area. Similar ζ-values also result for HRTFs of the right ear (not shown).

4. Limiting the spatial dynamic range of individual HRTFs

In this experiment, the dynamic range ζ of the individual HRTFs over all directions of incidence θ (i.e., in the spatial domain) was limited in each frequency bin. As discussed in Sec. II, the spatial dynamic range ζ is defined as the dB value of the ratio between the largest and the smallest magnitude of a directivity pattern per frequency f [cf. Eq. (3)]. ζ will depend on the individual head geometry and the measurement accuracy. For the subjects and the measurement accuracy used in this study (NFFT = 512, $f_s = 44.1$ kHz and 15° resolution in the horizontal plane), the maximum observed spatial dynamic range approached $\zeta_{max} \approx 80$ dB at higher frequencies (cf. Fig. 5).

In order to limit the spatial dynamic range of the individual HRTFs in every of the 512 frequency bins, low levels in the directivity pattern (i.e., spatial notches) were boosted such that they were not less than ζ' lower compared to the direction with the highest level (cf. Fig. 6). The phase spectra of the manipulated directions were left unchanged. This procedure was applied to the directivity pattern in each frequency bin separately, but with the same ζ'.

As discussed above, the impact of limiting ζ was tested by comparing the altered to the original HRTF at a fixed set of four directions. The four directions were chosen individually such that for each subject the directions with the largest subjective change between original and altered HRTF were included. To this end, the subjects underwent a preliminary listening test in which they had to select those four directions in the horizontal plane where they perceived a reduction of their spatial dynamic range to $\zeta'_{pre} = 25$ dB most saliently. These directions (which differed from subject to subject) were then used in the subsequent experiment.

In the actual experiment, the initial supra threshold spatial dynamic range was set to $\zeta' = 20$ dB, with an initial step size of $\Delta\zeta' = 3$ dB. The minimal step size was set to $\Delta\zeta' = 1$ dB.

IV. RESULTS AND DISCUSSION

In the following, the results from eight subjects are shown as means and standard deviations of three to six runs of each experiment.

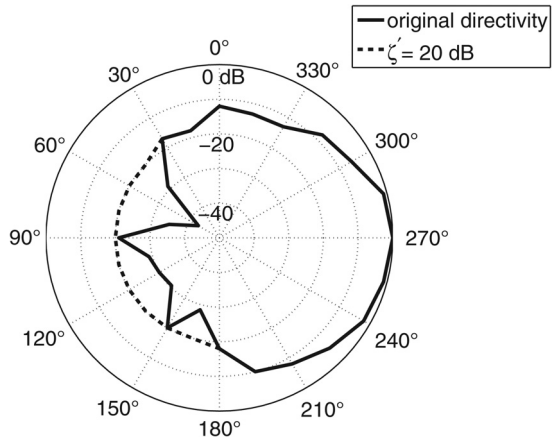

FIG. 6. Exemplary reduction of the spatial dynamic range of a normalized right ear HRTF magnitude in the horizontal plane from $\zeta \approx 43$ dB (original) to $\zeta' = 20$ dB (test situation).

A. Substitution of the individual HRTF phase by a mixture of the original phase and a linear phase

Means and standard deviations of the mixing gain L_ϕ at threshold (over three to six runs of the experiment), as a function of subjects and direction of incidence θ, are shown in Fig. 7.

It can be observed that the mean values of L_ϕ at threshold seem to vary more or less unpredictably with subject and direction of incidence. Many subjects (excluding subjects S1 and S8) perceived the modification of the individual HRTF phases for the frontal direction ($\theta = 0°$) at high L_ϕ (high proportion of the linear phase) only. In contrast, the modification of the HRTF phase for $\theta = 225°$ (gray diamonds) was already discriminable at small L_ϕ for most subjects. However, in both cases there were subjects who did not follow this trend.

None of the tested subjects showed a complete inability (i.e., $L_\phi = 1$) to discriminate the applied phase modification for the tested directions. Yet, some subjects indicated large L_ϕ

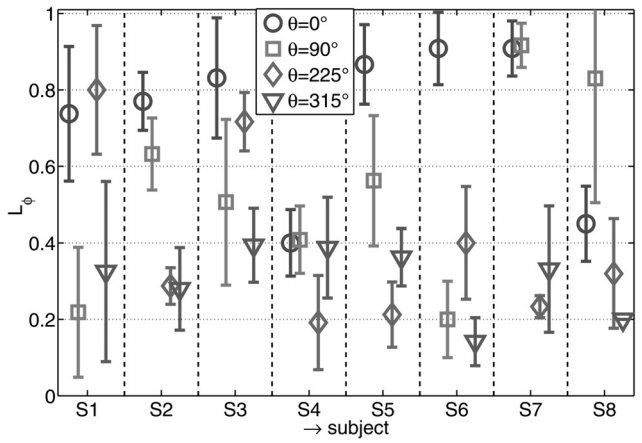

FIG. 7. Threshold data for experiment 1: means and standard deviations over three to six runs of the mixing gain L_ϕ at threshold, as a function of subjects and direction of incidence θ. Smaller values of L_ϕ, i.e., a smaller proportion of the linear phase, indicate a higher sensitivity to the broadband modification of the individual HRTF phase and vice versa.

and hence a pronounced insensitivity to phase modifications for particular directions (e.g., subject S6 at 0° and subject S7 at 0° and 90°).

Can the seemingly inconsistent results be explained by the fact that phase modifications yielded individual (binaural) spectra that were objectively different from the original ones, despite the fact that the same kind of (monaural) modification was applied? In an attempt to do so, we hypothesize that a large alteration of the interaural phase difference (IPD) from the original HRTF pair to the completely manipulated (i.e., 100% linear) HRTF pair corresponds to a rather salient discrimination. This in turn should correspond to small L_ϕ, and vice versa. In order to quantify the IPD alteration, we introduce a model for the individual discrimination ability L_{mod} as

$$L_{\text{mod}} = \nu \sum_{f=150\,\text{Hz}}^{1500\,\text{Hz}} \frac{\Delta f}{\text{ERB}(f)} \frac{|\text{IPD}_{\text{orig}}(f) - \text{IPD}_{\text{lin}}(f)|}{\phi_{\text{JND}}(f)},$$
$$\text{with} \sum_{f=150\,\text{Hz}}^{1500\,\text{Hz}} \left(\frac{\Delta f}{\text{ERB}(f)}\right) \nu = 1, \quad (4)$$

where ν is a normalization constant, $\Delta f = f_s/\text{NFFT}$ is the frequency resolution of the FFT, given by the sampling frequency f_s and the number of frequency bins NFFT, IPD(f) is the frequency-dependent interaural phase difference at a given frequency, ERB(f) is the frequency-dependent equivalent rectangular bandwidth according to Glasberg and Moore (1990), and $\phi_{\text{JND}}(f)$ is the interaural time difference threshold [taken from Klumpp and Eady (1956)]. Note that in the summation of Eq. (4) the difference between the original and the linear phase is weighted with the reciprocal value of the interaural time difference thresholds to account for the frequency-dependence of $\phi_{\text{JND}}(f)$. In addition, the division by ERB(f) ensures that each auditory filter is equally weighted in the summation. Moreover, we assumed that the ability to discriminate phase modifications is dominated by lower frequencies. Therefore, we chose to take into account frequencies between $150\,\text{Hz} \le f \le 1.5\,\text{kHz}$, because interaural phase differences in the stimulus fine-structure are processed by the auditory system up to about 1.5 kHz only [cf. Perrott and Nelson (1969)]. The lower limit is determined by the stimulus frequency content (cf. Sec. III C).

The relationship between the modeled individual discrimination ability L_{mod} and the measured L_ϕ (mean values) is shown in Fig. 8. As can be seen, high values of the individual discrimination ability correspond to low L_ϕ (i.e., to a high sensitivity to phase changes). The linear correlation coefficient [$\rho(L_{\text{mod}}, L_\phi) = -0.83$] is rather high for this type of experiment, indicating that L_{mod} is quite well suited to predict the individual sensitivity to broadband HRTF phase linearizations. On a side note, the observed correlation between L_{mod} and L_ϕ decreases when the upper frequency range of L_{mod} is extended to higher frequencies (not shown here). This fact again emphasizes the perceptual importance of the individual IPD at lower frequencies, which is well in line with the literature [cf. Kulkarni *et al.* (1999)]. Furthermore, it countenances the approach of the second experiment by substituting the

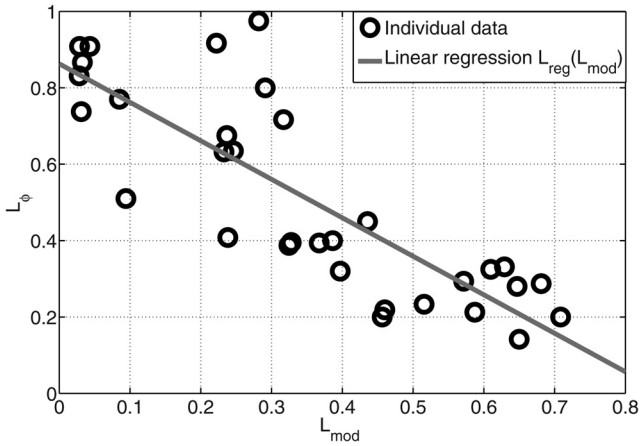

FIG. 8. Experimentally observed mean values of L_ϕ, as a function of the individual discrimination ability L_{mod} [Eq. (4)]. This relation can be characterized by a linear regression $L_{\mathrm{reg}}(L_{\mathrm{mod}}) = -1.0077\,L_{\mathrm{mod}} + 0.8627$ (gray line), with a linear correlation coefficient of $\rho(L_{\mathrm{mod}}, L_\phi) = -0.83$.

phase only at higher frequencies while maintaining the measured phase at lower frequencies.

B. Substitution of the individual HRTF phase by a linear phase above a certain cutoff-frequency

The mean cutoff-frequencies f_c at threshold and corresponding standard deviations are shown in Fig. 9. Both mean values and standard deviations vary with subject and direction of incidence. The highest f_c, i.e., the most sensitive thresholds, are associated with different directions for different subjects.

There is, however, a trend towards larger standard deviations for higher f_c. The most plausible explanation would be that for high f_c, the artifacts introduced by the applied phase modification will become so hard to discern that the subjects will have difficulty to judge consistently.

Relating these results to those of the first experiment (Sec. IV A), one would expect that the conditions associated with a poor sensitivity to broadband phase modification (large L_ϕ in Fig. 7) correspond to small cutoff-frequencies f_c

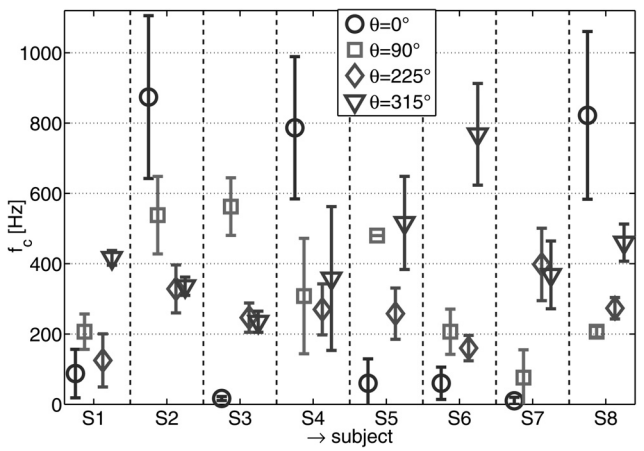

FIG. 9. Threshold data for experiment 2: means and standard deviations over three to six runs of the cutoff-frequency f_c at threshold between the original (for $f \le f_c$) and the linear phase (for $f > f_c$). Lower f_c indicate that a larger proportion of the original phase can be substituted with the linear phase still yielding a discrimination at threshold and vice versa.

in the second experiment. In other words, if a subject is completely unable to discriminate a broadband phase modification, this should also result in a (very) low cutoff-frequency f_c. On the other hand, a high f_c is not necessarily related to a high sensitivity to a broadband phase linearization (i.e., a low L_ϕ in experiment 1) because for higher f_c the two experiments refer to different phase manipulations.

In fact, when taking a closer look at Figs. 7 and 9 it can be seen that the lowest f_c values (reaching down to $f_c \approx 0$ Hz) for subjects S3 (0°), S5 (0°), S6 (0°), and S7 (0° and 90°) indeed correspond to L_ϕ values close to 1. In addition, the large L_ϕ values, e.g., for S1 (225°) and S8 (90°), imply a lack of discriminability of broadband phase modifications. For these conditions, the according cutoff-frequencies reach down to $f_c = 120$ Hz and $f_c = 205$ Hz, respectively. Considering that the test signal had a steep roll-off for frequencies $f < 200$ Hz, these cutoff-frequencies can also be regarded as "close to zero Hz." For all other conditions, the relation between L_ϕ and f_c does not follow a simple pattern.

In conclusion, the experimental results show that for $f > 1000$ Hz, the HRTF phase of the tested population could be substituted by a suitable linear phase without causing any discriminable artifacts for any of the eight subjects. In order to give a first impression of the consequences of the proposed phase manipulation for the resulting HRIRs, two exemplary HRIRs are plotted in Fig. 10. The black solid graph shows a measured HRIR and the gray dashed graph shows this HRIR when its phase is linearized for $f \ge 1$ kHz. It can be seen that the phase linearization alters the envelope of the impulse response, yielding a more symmetrical counterpart. This effect was observed for all HRIRs. It is worth noting that the peak of the magnitude of the HRIRs remains at a fixed delay which is encoded in the slope of the linear phase.

C. Complex smoothing of the individual HRTFs into constant relative bandwidths

The mean relative bandwidths B_W at threshold and corresponding standard deviations when smoothing individual

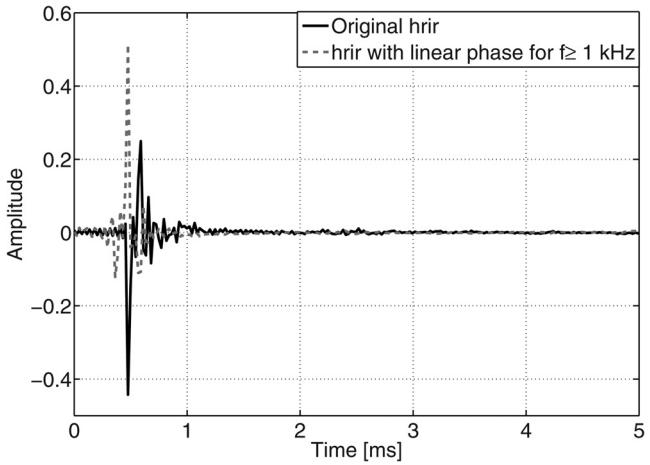

FIG. 10. Exemplary course of a measured HRIR (black solid line) and a manipulated HRIR (gray-dashed line) for the left ear of subject S8 and $\theta = 105°$ as a function of time. The phase of the latter HRIR is linearized for $f \ge 1$ kHz, which results in a rather symmetrical envelope of the corresponding HRIR in the time domain. Note that the delay of the peak magnitude remains unchanged.

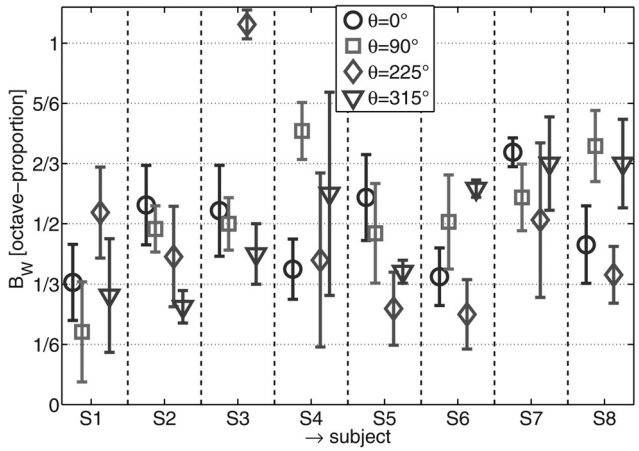

FIG. 11. Threshold data for experiment 3: means and standard deviations over three to six runs of the relative bandwidth B_W at threshold expressed as fractions of one octave when complexly smoothing HRTFs in the frequency domain. Prior to smoothing, the HRTF phase (for $f > 5$ kHz) was substituted with the linear phase ϕ_{lin}. Higher B_W indicate more smoothing at threshold and thus less sensitive thresholds and vice versa.

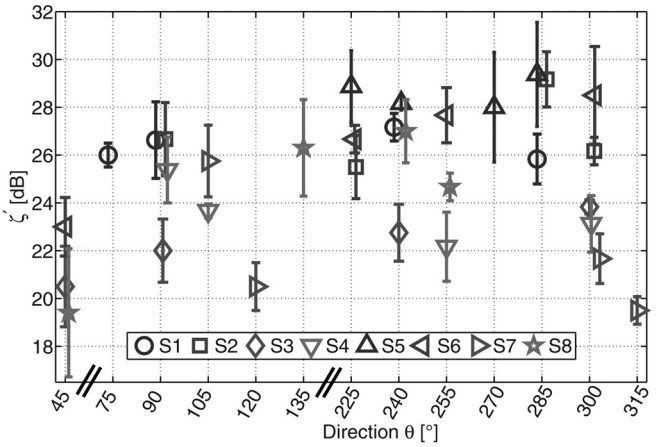

FIG. 12. Threshold data for experiment 4: means and standard deviations over three to six runs of the spatial dynamic range ζ' as a function of azimuthal direction θ. Note the non-equidistant spacing. Here the individual subjects are depicted as various symbols. Smaller spatial dynamic ranges ζ' indicate larger manipulations of the directivity patterns and thus less sensitive thresholds and vice versa.

(complex-valued) HRTFs in the frequency domain (with the original phase substituted by a linear phase for $f > 5$ kHz) are shown as fractions of one octave in Fig. 11. Again, both means and standard deviations seem to vary with subject and direction of incidence in an unpredictable manner. This is not surprising, since HRTFs have individual spectral shapes, and smoothing into relative bandwidths will therefore yield individually varying artifacts.

The lowest relative bandwidths at threshold (most sensitive cases) were at $B_W \approx 1/5$ octave, i.e., between a major second and a minor third. The highest bandwidth at threshold (least sensitive case) was at $B_W \geq 1$ octave, this occurred however just for one subject at one direction of incidence (subject S3, 225°). The vast majority of all conditions resulted in bandwidths at threshold between one and two third octaves. These bandwidths approximately correspond to 2 to 4 ERBs at high frequencies [cf. Glasberg and Moore (1990)].

In conclusion the results indicate that the HRTFs of the tested population may be smoothed complexly with a bandwidth of $B_W \approx 1/5$ octave (corresponding to roughly 1 ERB at high frequencies) without causing a perceptual difference to the original HRTFs. These results also confirm the validity of the proposed phase modifications in Sec. IV B ($f_c = 5$ kHz), since if the phase linearization was discriminable the subjects would have indicated a subjective difference in any case (independent of the bandwidth), which would have resulted in bandwidths at threshold of $B_W \approx 0$.

D. Limiting the spatial dynamic range of the individual HRTFs

The spatial dynamic ranges ζ' (means and standard deviations) at threshold are shown in Fig. 12.

First of all, the directions chosen by the subjects were not equally distributed in the horizontal plane. Instead, they were all in the ranges $90° \pm 45°$ or $270° \pm 45°$, i.e., at lateral directions. This can be explained by the fact that lateral directions exhibit more spectral notches, which results in larger modifications when the spatial dynamic range is limited.

The experimentally determined spatial dynamic ranges at threshold vary between $19\,\text{dB} \leq \zeta' \leq 29\,\text{dB}$. The standard deviation varies approximately between $0.1\,\text{dB} \leq \sigma(\zeta') \leq 3\,\text{dB}$. Again, the variation of individual ζ' is hypothesized to be due to spectral characteristics of the HRTFs which are individual for each subject.

As a general conclusion, the spatial dynamic range of the HRTFs in the horizontal plane of the tested population can be reduced to $\zeta' > 29$ dB without leading to a discriminable alteration compared to the original HRTFs.

V. DISCUSSION

Four different methods for smoothing HRTFs were investigated. The first two addressed phase smoothing which is a pre-processing step required to do complex spectral smoothing. The third method was a complex spectral smoothing method, and the last method was a spatial smoothing method.

A. Perceptual effects of phase linearization

We found that the binaural discriminability of a broadband phase linearization of individual HRTFs (while preserving the delay) largely varies with subject and direction. This agrees well with the investigations by Kulkarni *et al.* (1999). We assumed the individual characteristic of the observed thresholds to be due to the differences in the individual phase spectra that are modified by the broadband linearization. In fact, although the sensitivity to monaural phase differences is rather poor [cf. Kulkarni *et al.* (1999)], the sensitivity to interaural phase differences is high for low frequencies [cf. Klumpp and Eady (1956) and Yost (1974)]. We assumed that the discriminability is proportional to the individual IPD alteration at these frequencies. Using a simple model representing the individual discrimination ability

[cf. Eq. (4)], it was possible to approximately quantify and predict the majority of the individual thresholds (cf. Fig. 8). More clearly than in Kulkarni *et al.* (1999), the current results indicate a pronounced (individual) sensitivity to a broadband phase linearization for some subjects/directions (cf. Fig. 7, $\theta = 315°$). Possible explanations for this difference may be the differing HRTF directions [$\theta = 0°$, 90°, −90°, 180° in Kulkarni *et al.* (1999)] and the higher number of subjects in the current study [eight over four in Kulkarni *et al.* (1999)], which increases the chance to spot individually higher sensitivities. On the basis of these findings, a broadband linearization seems to be an inappropriate preprocessing operation for smoothing complex-valued HRTFs in the frequency domain.

Based on the previous arguments, we assume the phase linearization to be less discriminable when the monaural phase and hence the IPD is preserved for lower frequencies. The observed thresholds (cf. Fig. 9) indicate that a phase linearization is not discriminated when the original phase is preserved for $f < 1000$ Hz. This cutoff-frequency complies well with literature on interaural phase processing (phase locking) in the auditory system, which is limited up to about 1.5 kHz [cf. Perrott and Nelson (1969)]. Moreover, these findings are in line with Kulkarni *et al.* (1999) where no significant phase discrimination was reported with high-pass stimuli which only contained frequencies $f > 2$ kHz.

Interestingly, even the highest cutoff-frequencies at threshold are lower than those observed in a previous study [Rasumow *et al.* (2012)]. This difference turned out to be due to a detail of the phase modification: In the current study, the original phase ϕ_{orig} was substituted by a linear phase ϕ_{lin} above f_c *without* any fading between the two, whereas in the previous study, the transition between original and linear phase was accompanied by a spectral windowing function extending over five frequency bins. This result indicates that a larger proportion of the measured phase (lower f_c) can be substituted when *no* fading between the original phase and the linear phase is applied. Hence, we assume the fading between the two phases in Rasumow *et al.* (2012) to have created cues that are not present when no fading is used.

B. Perceptual effects of smoothing HRTFs in the frequency domain

In the current study, the ability to discriminate complexly smoothed HRTFs with variable relative bandwidths was investigated after applying a linear phase for $f > 5$ kHz (cf. Fig. 2). In general, the obtained bandwidths at threshold approximately corresponded to one to two third octaves (approximately 2 to 4 ERBs at high frequencies). These results are in good agreement with general investigations on frequency grouping [cf. Patterson and Nimmo-Smith (1980) and Glasberg and Moore (1990)] and comparable with similar investigations [cf. Xie and Zhang (2010)], where the magnitude of HRTFs could be smoothed with bandwidths of 1 to 3.5 ERBs for higher frequencies ($f > 5000$ Hz). Furthermore, these results also compare well with findings from Breebaart and Kohlrausch (2001) where unprocessed HRTFs could not be discriminated from those smoothed using a gammatone filter bank (with its bandwidth approximately corresponding to the bandwidth of one ERB). Interestingly, those studies used a separate smoothing of magnitude and phase, whereas in the current study the HRTFs were complexly smoothed after applying a phase linearization above 5 kHz. It thus seems that both approaches are equally well suited to spectrally smooth HRTFs.

In the most sensitive condition, the relative bandwidth at threshold was at about a minor third, indicating that the individual HRTFs of the tested population could safely be smoothed into constant relative bandwidths of $B_W = \frac{1}{5}$ octave for any of the eight subjects.

C. Influence of the test signal

In the current study we used broadband noise bursts, which, on the one hand, have a broadband spectrum and, on the other hand, a temporal structure. Such signals have successfully been used in a number of discrimination experiments on the effect of HRTF manipulations [cf. Kulkarni *et al.* (1999)]. Also, in a study on the effect of various headphone equalizations, noise signals have been shown to give more sensitive thresholds than music stimuli [cf. Lindau and Brinkmann (2012)]. Since headphone equalization is comparable to spectral HRTF smoothing, these results should apply to the third experiment in a similar manner. Compared to continuous noise stimuli, we assume the discrimination to be still more sensitive when noise bursts are used, which is particularly important for the discrimination of phase manipulations.

In the fourth experiment, we modified HRTFs frequency bin by frequency bin. A test signal which would exhibit dominant components in the particular frequency bin only could then possibly give more salient discrimination cues. Such test signals are, however, extremely artificial and it is hard to imagine a scenario in which they will be relevant for real-world applications of the VAH.

D. Impact of the proposed smoothing methods on the HRTFs

To illustrate the functioning and to give a first impression regarding the two proposed smoothing methods, an exemplary HRTF (subject S2, right ear, top box in Fig. 13) was complexly smoothed into $B_W = \frac{1}{5}$ octave bands (second box in Fig. 13) and the spatial dynamic range of the measured HRTF was reduced to $\zeta' = 29$ dB (third box in Fig. 13). Comparing the first two boxes, it is apparent that the complex smoothing clearly reduced the spectral resolution of the measured HRTF, especially at higher frequencies ($f \geq 6$ kHz). Furthermore, this example demonstrates that the complex smoothing leads to a smooth magnitude spectrum, also at high frequencies, which would not be the case if no prior phase linearization would have been applied (cf. Fig. 2). The reduction of the spatial dynamic range in the third box is best visible for contralateral directions ($0° \leq \theta \leq 180°$) where the most spatial notches (light areas) are adjusted per frequency bin. This manipulation

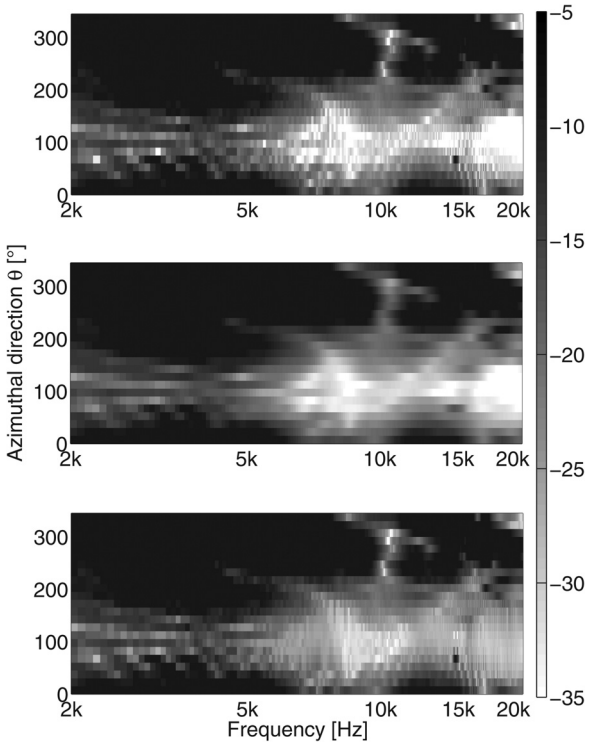

FIG. 13. Effects of the proposed smoothing methods. Top: Exemplary HRTF (subject S2, right ear) as a function of frequency in kHz on the x axis and as a function of azimuthal direction θ on the y axis. Middle: the same HRTF, processed through a complex smoothing as described in Sec. III C 3 with a bandwidth of $B_W = \frac{1}{3}$ octave. Bottom: the same HRTF, with the spatial dynamic range limited to $\zeta' = 29$ dB as described in Sec. III C 4. The latter manipulation is primarily apparent for contralateral directions ($0° \leq \theta \leq 180°$) where the spatial notches (light areas) are adjusted. In this illustration, levels were limited between -5 and -35 dB to highlight the differences of the particular smoothing methods.

is primarily effective for spatial notches which mainly occur for frequency $f \geq 5000$ Hz.

E. Impact of the proposed smoothing methods on the accuracy of the VAH

The main motivation of this study was to enhance the performance of the VAH by smoothing the desired HRTFs prior to the computation of the VAH filters. In order to illustrate the benefits associated with the complex smoothing of HRTFs in the frequency domain and the reduction of the spatial dynamic range, we computed VAH filters for individual HRTFs using analytical steering vectors (i.e., pure delays) of 24 microphones positioned according to the topology used by Rasumow *et al.* (2011). The filters were calculated by minimizing a least-squares cost function linking the re-synthesized directivity patterns $\text{HRTF}_{\text{synth}}$ to the desired directivity patterns HRTF_{des} for each frequency bin separately according to Eq. (5) in Rasumow *et al.* (2013), with a desired white noise gain of $\text{WNG} = 3$ dB [cf. Eq. (3) in Rasumow *et al.* (2013)].

First, to illustrate the effect of complexly smoothing HRTFs in the frequency domain an example (subject S2, right ear, $\theta = 210°$) is shown in Fig. 14. The complex spectral smoothing of the HRTF_{des} becomes more and more

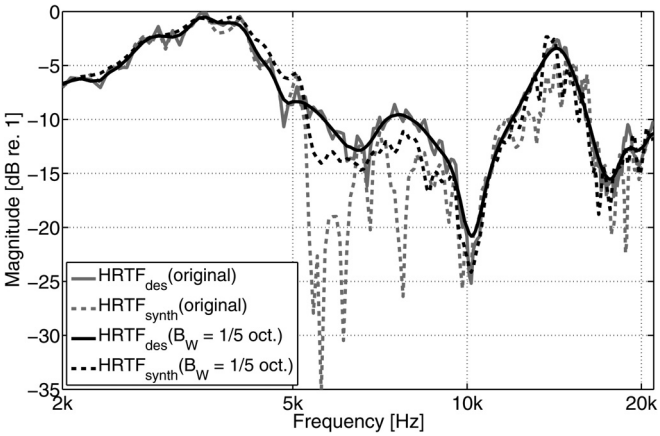

FIG. 14. Desired HRTF_{des} (solid lines) and re-synthesized $\text{HRTF}_{\text{synth}}$ using the VAH (dashed lines) are depicted for the original (gray) and for the spectrally smoothed case (black) with $B_W = \frac{1}{5}$ for an exemplary HRTF (subject S2, right ear, $\theta = 210°$), as a function of frequency.

visible with increasing frequency (cf. black and gray solid lines), which is expected for smoothing into constant relative bandwidths. Furthermore, the re-synthesis using the VAH clearly changes when the HRTF_{des} is smoothed. Especially for frequencies $5\,\text{kHz} \leq f \leq 10\,\text{kHz}$ and $f \geq 16\,\text{kHz}$ the re-synthesis using the VAH improves considerably when the HRTF_{des} is smoothed spectrally prior to the re-synthesis.

Second, to illustrate the benefits of reducing the spatial dynamic range, an exemplary right-ear HRTF_{des} (subject S2, right ear, $f \approx 13, 5$ kHz) and its re-synthesis $\text{HRTF}_{\text{synth}}$ using the VAH are depicted in Fig. 15 as a function of direction θ. There the spatial notch of the original HRTF_{des} at 105° is only poorly re-synthesized by the VAH (gray lines, maximum error of >20 dB). The performance of the VAH, however, improves considerably (at $90° \pm 30°$) when the spatial dynamic range of the HRTF_{des} is reduced to $\zeta' = 29$ dB (black lines, maximum error of ≈5 dB).

In order to quantify the synthesis accuracy for each set of HRTFs (one ear, P directions of incidence), we used the absolute VAH-error E,

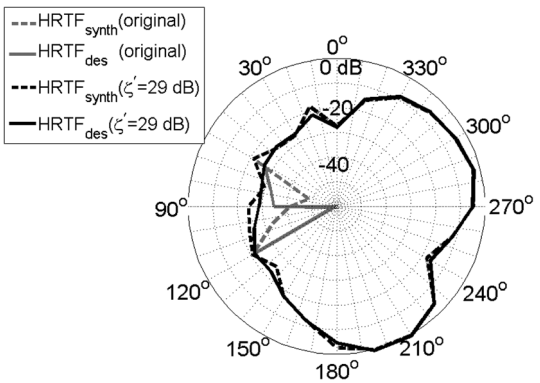

FIG. 15. The HRTF_{des} (solid lines) and their re-synthesis $\text{HRTF}_{\text{synth}}$ using the VAH (dashed lines) are depicted for the original HRTF_{des} (gray) and for the HRTF_{des} with reduced spatial dynamic range ($\zeta' = 29$ dB, black lines) for an exemplary HRTF (subject S2, right ear, $f \approx 13, 5$ kHz).

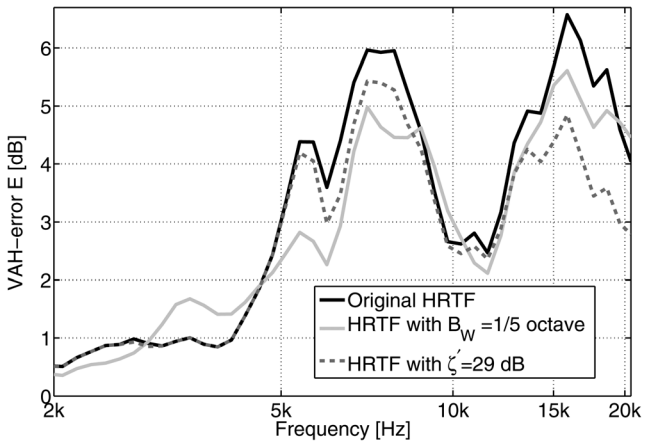

FIG. 16. Effects of the proposed smoothing methods on the accuracy of the VAH. Plotted is the VAH-error E [Eq. (5)] of the right ear HRTF of subject S2, as a function of frequency, for the original $\mathrm{HRTF}_{\mathrm{des}}$ (black solid line), the spectrally smoothed $\mathrm{HRTF}_{\mathrm{des}}$ ($B_W = \frac{1}{5}$ octave, gray solid line) and the $\mathrm{HRTF}_{\mathrm{des}}$ with reduced spatial dynamic range ($\zeta' = 29$ dB, gray solid line).

$$E(\mathrm{CB}) = \frac{1}{P}\sum_{i=1}^{\mathrm{P=24}}\left|20\log_{10}(|\delta(\theta_i)|)\mathrm{dB}\right|,$$
$$\text{with}\quad \delta(\theta_i) = \frac{1}{\mathrm{M}}\sum_{f=f_1}^{f_{\mathrm{M}}}\frac{\mathrm{HRTF}_{\mathrm{des}}(f,\,\theta_i)}{\mathrm{HRTF}_{\mathrm{synth}}(f,\,\theta_i)}, \qquad (5)$$

where $\delta(\theta_i)$ is computed in critical bands (CB) with 50% overlap [cf. Rasumow *et al.* (2011)]. f_1 and f_{M} indicate the lowest and highest frequency bin within each CB.

This error is shown in Fig. 16 as a function of frequency for an exemplary subject (subject S2, right ear, $P = 24$). As can be seen, the re-synthesis with the VAH works best for lower frequencies up to $f \approx 5$ kHz. At low frequencies ($f < 2$ kHz), the VAH-error E is still lower and therefore not shown here. Larger errors primarily occur at higher frequencies. Compared to the re-synthesis of the original HRTF, the VAH-error decreases especially in the frequency range of $5\,\mathrm{kHz} \le f \le 10\,\mathrm{kHz}$ for the spectrally smoothed HRTF. On the other hand, the error is slightly higher around 4 kHz. This can be explained by the fact that a spectral smoothing alters the $\mathrm{HRTF}_{\mathrm{des}}$ for all directions separately and hence may generate directivity patterns which are harder to re-synthesize than the original HRTFs. Yet, in general the VAH-error clearly decreases when the $\mathrm{HRTF}_{\mathrm{des}}$ is smoothed in the frequency domain.

The VAH-error E for the re-synthesis of the $\mathrm{HRTF}_{\mathrm{des}}$ with a spatially limited dynamic range shows improvements over the re-synthesis of the original $\mathrm{HRTF}_{\mathrm{des}}$ for frequencies $5\,\mathrm{kHz} \le f \le 9\,\mathrm{kHz}$ and $f \ge 10\,\mathrm{kHz}$. At these frequencies, spatial notches occurred frequently and were thus modified by the reduction of the spatial dynamic range. It is worth noting that the VAH-error for the re-synthesis of the HRTF with limited spatial dynamic range never exceeds the VAH-error for the original HRTF.

To put these re-synthesis accuracies into one number, the mean error $\bar{E}$ for all directions θ and frequencies f can be used. It decreased from 3.8 to 3.4 dB when the measured HRTFs were smoothed spectrally and from 3.8 to 3.0 dB when the spatial dynamic range was reduced according to the obtained thresholds. Although, the benefits of the VAH re-syntheses highly depend on the individual HRTFs, the applied optimization strategy etc.

VI. SUMMARY AND CONCLUSIONS

In this study we investigated HRTF smoothing in the spectral and spatial domains as a preprocessing step for the virtual artificial head. It was found that:

(1) Subjects are sensitive to a broadband phase linearization of HRTFs.
(2) The individual sensitivity to a broadband phase linearization can be predicted by a simple model that is based on interaural phase differences at frequencies $f \le 1,5$ kHz.
(3) The original phase can be substituted by a linear phase above $f > 1000$ Hz without introducing noticeable artifacts.
(4) After substituting the original phase by a linear phase above $f \ge 5$ kHz, HRTFs may be smoothed complexly into constant relative bandwidths of $B_W \le 1/5$ octave, without introducing noticeable artifacts.
(5) Spatial notches in the directivity pattern do not need to be retained in detail if they are less than 29 dB below the maximum value.

These findings permit us to efficiently smooth individual HRTFs in the spectral and spatial domains. It must, however, be kept in mind that they are limited to HRTFs in the horizontal plane.

In the future we will investigate extensions to non-horizontal HRTFs and evaluate the consequences of the proposed preprocessing methods on the re-synthesis of individual HRTFs using the VAH. In this context the threshold values found in this study may be used as a starting point to optimize the overall performance of the VAH.

ACKNOWLEDGMENTS

This project was partially funded by Bundesministerium für Bildung und Forschung under grant no. 17080X10 and the Cluster of Excellence 1077 "Hearing4All" of the German Research Foundation.

APPENDIX A: ESTIMATION OF THE DELAY AND THE RESULTING LINEAR PHASE $\phi_{\mathrm{lin}}(f)$

The main idea of this method is that the estimated linear phase should maintain the delay of the largest magnitude (loudest part) of a HRIR (cf. Fig. 10). Consequently, we estimated the delay of the largest magnitude associated with the largest magnitude of the Hilbert-envelope of a HRIR. The Hilbert-envelope of a HRIR is defined as the magnitude of the analytic signal, which was calculated using the function **hilbert** in MATLAB (with the HRIR-length of $n = 512$ samples and $f_s = 44\,100$ Hz). The analytic signal is calculated in a three-step algorithm: First, the n-point fast Fourier transformation (FFT) of the HRIR is calculated. Second, this FFT-sequence is multiplied with the factor of 2 for frequency bins i ranging from $2 \le i \le (n/2)$ and with the factor of 0 for frequency bins ranging from $(n/2)+2 \le i \le n$. Third, the inverse fast Fourier transformation of the manipulated spectrum is calculated, where the first n elements represent the analytic signal.

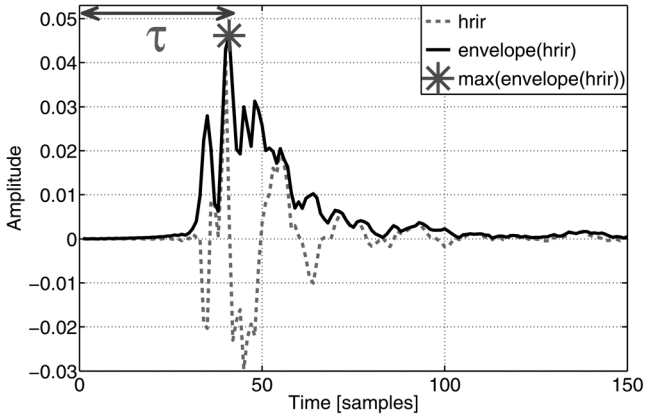

FIG. 17. An exemplary HRIR is depicted (dashed gray line) as a function of time in samples on the x axis. The black solid line depicts the corresponding Hilbert-envelope of this HRIR. The gray asterisk is the maximum of the Hilbert-envelope, characterizing the estimated delay τ in samples.

An exemplary HRIR and its corresponding Hilbert envelope are depicted in Fig. 17 as a function of time in samples. The maximum of the Hilbert-envelope was used to define the estimated delay τ of the HRIR. Based on this approach, the estimated linear phase was given by

$$\phi_{\text{lin}}(f) = e^{-i2\pi(f/f_s)\tau} \tag{A1}$$

with f being the discrete frequency vector (with the length of $n = 512$ bins) equidistantly ranging from zero to f_s.

APPENDIX B: APPLIED METHOD FOR THE SPECTRAL SMOOTHING OF HRTFS

Initially the delay of the measured HRIR is estimated (step I in Fig. 18) according to the method described in Appendix A. This step yields a linear phase ϕ_{lin}, which is used to continue the phase of the HRTF for frequencies $f \geq 5$ kHz (step II). Generally, a smoothing of HRTFs into constant relative bandwidths in the frequency domain is analogous to a truncation of the HRIRs in the time domain using frequency-dependent truncation lengths. Hence, in order to ensure that the smoothing algorithm processes the main parts of the HRIRs, it is important to remove the overall delay τ before smoothing. Analogous to Eq. (A1), the removal of the delay τ is done in step III by multiplying the HRTF with the phase term $e^{i2\pi(f/f_s)\tau}$. It is worth noting that due to the preprocessing in steps I–III the phase of the HRTF is set to zero for frequencies $f \geq 5$ kHz (cf. phase in step III in Fig. 18). By this means the phase manipulations in steps I–III reduce a complex-valued HRTF to a real valued HRTF for higher frequencies, while the phase characteristics for the lower frequencies (except for the delay τ) are maintained. In step IV the so-preconditioned HRTF is complexly smoothed according to the algorithm proposed in Hatziantoniou and Mourjopoulos (2000) with the bandwidth B_W. In the final step (V) the original delay τ of the measured HRTF is reconstructed. Analogous to the previous procedure the delay τ is reconstructed by multiplying the smoothed HRTF with the phase term $e^{-i2\pi(f/f_s)\tau}$.

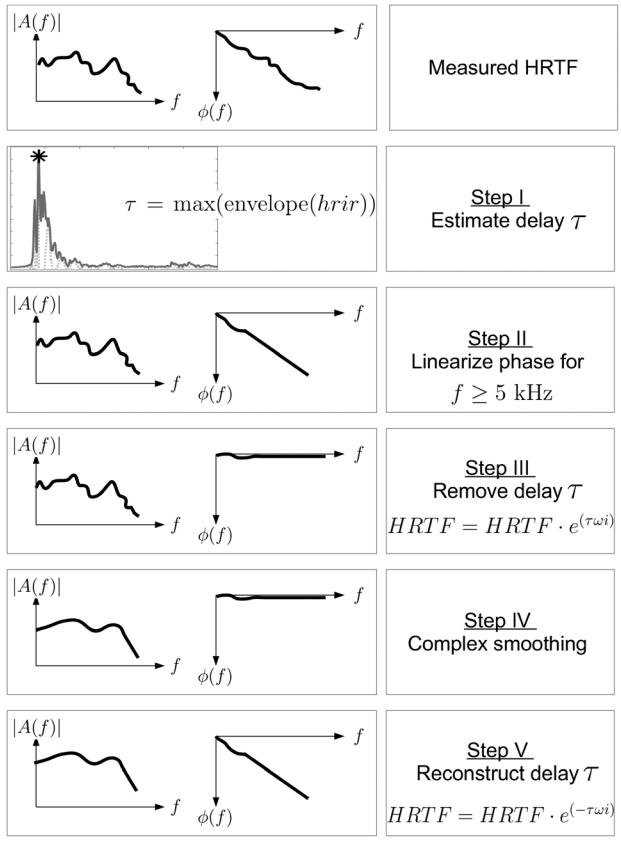

FIG. 18. Block diagram characterizing the procedure for the applied complex smoothing of HRTFs. The process can be divided into five separate steps: In step I the delay τ is estimated as described in Appendix A. The resulting linear phase is applied to the HRTF for frequencies $f \geq 5$ kHz in step II. In step III the delay τ is removed from the measured HRTF. In step IV the preconditioned HRTF is smoothed complexly according to the algorithm proposed in Hatziantoniou and Mourjopoulos (2000). In step V the initial delay τ is reconstructed for the smoothed HRTF.

[1]The IRCAM HRTF-database is available at http://recherche.ircam.fr/equipes/salles/listen (date last viewed 3/6/2014).
[2]Psylab is available at http://www.hoertechnik-audiologie.de/psylab (date last viewed 3/6/2014).

Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (**2001**). "The CIPIC HRTF Database," in *Proceedings of the 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2001)* (IEEE, New Paltz, NY), pp. 99–102.

Atkins, J. (**2011**). "Robust beamforming and steering of arbitrary beam patterns using spherical arrays," in *Proc. 20 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2011)* (IEEE, New Paltz, NY), pp. 237–240.

Breebaart, J., and Kohlrausch, A. (**2001**). "Perceptual (ir)relevance of HRTF magnitude and phase spectra," in *Audio Engineering Society Convention*, Amsterdam, Netherlands, Vol. 110, pp. 1–9.

Breebaart, J., Nater, F., and Kohlrausch, A. (**2010**). "Spectral and spatial parameter resolution requirements for parametric, filter-bank-based HRTF processing," J. Audio Eng. Soc. **58**, 126–140.

Castaneda, C. D. S., Sakamoto, S., Lopez, J. A. T., Li, J., Yan, Y., and Suzuki, Y. (**2013**). "Accuracy of head-related transfer functions synthesized with spherical microphone arrays," in *Proceedings of Meetings on Acoustics 2013 (ICA)*, ASA No. 055085, Montreal, Canada, Vol. 19.

Chen, J., Veen, B. D. V., and Hecox, K. E. (**1992**). "External ear transfer function modeling: A beamforming approach," J. Acoust. Soc. Am. **92**, 1933–1944.

Duraiswami, R., Zotkin, D. N., and Gumerov, N. A. (**2004**). "Interpolation and range extrapolation of HRTFs [head related transfer functions]," in

Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004 (ICASSP '04), IEEE, Montreal, Canada, Vol. 4, pp. iv-45–iv-48.
Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–65.
Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
Hacker, M., and Ratcliff, R. (**1979**). "A revised table of d' for m-alternative forced choice," Percept. Psychophys. **26**, 168–170.
Hammershoi, D., and Moller, H. (**1996**). "Sound transmission to and within the human ear canal," J. Acoust. Soc. Am. **100**, 408–427.
Hatziantoniou, P. D., and Mourjopoulos, J. N. (**2000**). "Generalized fractional-octave smoothing of audio and acoustic responses," J. Audio Eng. Soc. **48**, 259–280.
Huopaniemi, J., and Karjalainen, M. (**1996**). "HRTF filter design based on auditory criteria," in *Proceedings of the Nordic Acoustical Meeting (NAM'96)*, Acoustical Society of Finland, Helsinki, Finland, pp. 323–330.
Huopaniemi, J., Zacharov, N., and Karjalainen, M. (**1999**). "Objective and subjective evaluation of head-related transfer function filter design," J. Audio Eng. Soc. **47**, 218–239.
Kahana, Y., Nelson, P. A., Kirkeby, O., and Hamada, H. (**1999**). "A multiple microphone recording technique for the generation of virtual acoustic images," J. Acoust. Soc. Am. **105**, 1503–1516.
Klumpp, R. G., and Eady, H. R. (**1956**). "Some measurements of interaural time difference thresholds," J. Acoust. Soc. Am. **28**, 859–860.
Kulkarni, A., and Colburn, S. (**1998**). "Role of spectral detail in sound-source localization," Lett. Nature **396**, 747–749.
Kulkarni, A., Isabelle, S. K., and Colburn, S. (**1999**). "Sensitivity of human subjects to head-related transfer-function phase spectra," J. Acoust. Soc. Am. **105**, 2821–2840.
Lindau, A., and Brinkmann, F. (**2012**). "Perceptual evaluation of headphone compensation in binaural synthesis based on non-individual recordings," J. Audio Eng. Soc. **60**, 54–62.
Mehrgardt, S., and Mellert, V. (**1977**). "Transformation characteristics of the external human ear," J. Acoust. Soc. Am. **61**, 1567–1576.
Mitchell, L. D. (**1982**). "Improved methods for the fast Fourier transform (FFT) calculation of the frequency response function," J. Mech. Des. **104**, 277–279.
Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic Press, London), Chap. 3.
Patterson, R. D., and Nimmo-Smith, I. (**1980**). "Off-frequency listening and auditory-filter asymmetry," J. Acoust. Soc. Am. **67**, 229–245.
Paul, S. (**2009**). "Binaural recording technology: A historical review and possible future developments," Acta Acust. Acust. **95**, 767–788.
Perrott, D. R., and Nelson, M. A. (**1969**). "Limits for the detection of binaural beats," J. Acoust. Soc. Am. **46**, 1477–1481.
Rasumow, E., Blau, M., Doclo, S., Hansen, M., van de Par, S., Püschel, D., and Mellert, V. (**2013**). "Least squares versus non-linear cost functions for a vitual artificial head," in *Proceedings of Meetings on Acoustics 2013 (ICA)*, ASA No. 055082, Montreal, Canada, Vol. 19.
Rasumow, E., Blau, M., Hansen, M., Doclo, S., van de Par, S., Mellert, V., and Püschel, D. (**2011**). "Robustness of virtual artificial head topologies with respect to microphone positioning errors," in *Proceedings of Forum Acusticum 2011*, European Acoustics Association, Aalborg, Denmark, pp. 2251–2256.
Rasumow, E., Blau, M., Hansen, M., Doclo, S., van de Par, S., Püschel, D., and Mellert, V. (**2012**). "Smoothing head-related transfer functions for a virtual artificial head," in *Proceedings of the Acoustics 2012 Nantes Conference*, European Acoustics Association, Nantes, France, pp. 1019–1024.
Romigh, G. D., Brungart, D., Stern, R. M., and Simpson, B. D. (**2013**). "The role of spatial detail in sound-source localization: Impact on HRTF modeling and personalization," in *Proceedings of Meetings on Acoustics 2013 (ICA)*, ASA No. 050170, Montreal, Canada, Vol. 19.
Tohtuyeva, N., and Mellert, V. (**1999**). "Approximation of dummy-head recording technique by a multimicrophone arrangement," J. Acoust. Soc. Am. **105**, 1101.
Xie, B., and Zhang, T. (**2010**). "The audibility of spectral detail of head-related transfer functions at high frequency," Acta Acust. Acust. **96**, 328–339.
Yost, W. A. (**1974**). "Discriminations of interaural phase differences," J. Acoust. Soc. Am. **55**, 1299–1303.
Zotkin, D. N., Duraiswami, R., and Gumerov, N. A. (**2009**). "Regularized HRTF fitting using spherical harmonics," in *2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2009)* (IEEE, New Paltz, NY), pp. 257–260.