Using beamforming and binaural synthesis for the psychoacoustical evaluation of target sources in noise

Wookeun Song[a)]
Sound Quality Research Unit, Department of Acoustics, Aalborg University, Fredrik Bajers Vej 7B, DK-9220 Aalborg East, Denmark and Brüel & Kjær Sound & Vibration Measurement A/S, Skodsborgvej 307, DK-2850 Nærum, Denmark

Wolfgang Ellermeier
Institut für Psychologie, Technische Universität Darmstadt, Alexanderstraße 10, D-64283 Darmstadt, Germany

Jørgen Hald
Brüel & Kjær Sound & Vibration Measurement A/S, Skodsborgvej 307, DK-2850 Nærum, Denmark

(Received 15 March 2007; accepted 18 November 2007)

The potential of spherical-harmonics beamforming (SHB) techniques for the auralization of target sound sources in a background noise was investigated and contrasted with traditional head-related transfer function (HRTF)-based binaural synthesis. A scaling of SHB was theoretically derived to estimate the free-field pressure at the center of a spherical microphone array and verified by comparing simulated frequency response functions with directly measured ones. The results show that there is good agreement in the frequency range of interest. A listening experiment was conducted to evaluate the auralization method subjectively. A set of ten environmental and product sounds were processed for headphone presentation in three different ways: (1) binaural synthesis using dummy head measurements, (2) the same with background noise, and (3) SHB of the noisy condition in combination with binaural synthesis. Two levels of background noise (62, 72 dB SPL) were used and two independent groups of subjects ($N=14$) evaluated either the loudness or annoyance of the processed sounds. The results indicate that SHB almost entirely restored the loudness (or annoyance) of the target sounds to unmasked levels, even when presented with background noise, and thus may be a useful tool to psychoacoustically analyze composite sources. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2822669]

PACS number(s): 43.66.Cb, 43.60.Fg, 43.66.Pn [RAL] Pages: 910–924

I. INTRODUCTION

The localization of problematic sound sources in a sound field is becoming increasingly important in areas such as automotive engineering and the aerospace, and consumer electronics industry. Typically, array techniques, such as near-field acoustic holography (NAH) (Maynard *et al.*, 1985; Veronesi and Maynard, 1987) and beamforming (Johnson and Dudgeon, 1993) have been employed to identify the noise sources of interest. In beamforming, a microphone array can be placed at a certain distance from the source plane and therefore it is easier to use in comparison with NAH, when there are obstacles close to the test object. Furthermore, the output of a beamformer is typically the sound pressure contribution at the center of the array in the absence of the array and this can be easily transformed to the sound pressure contribution at both ears by incorporating binaural technology (Møller, 1992). Hald (2005) proposed a scaling factor, which can be applied to the output of the delay-sum beamformer in order to obtain sound power estimates.

Since conventional physical measures, such as sound pressure or intensity, do not take into account how human listeners perceive sounds, there is growing interest in predicting specific psychoacoustic attributes from objective acoustical parameters (Ellermeier *et al.*, 2004b; Zwicker and Fastl, 2006). That also holds for microphone-array measurements in that it is desirable to identify problematic noise sources by mapping the sound fields of interest in terms of psychoacoustic attributes (Song, 2004; Yi, 2004) and by determining the directional contribution from individual sources (Song *et al.*, 2006).

Recently, spherical microphone arrays have been investigated for the recording and analysis of a sound field (Meyer, 2001; Meyer and Agnello, 2003; Petersen, 2004; Rafaely, 2004, 2005a). The major advantage of spherical microphone arrays where the microphones are distributed along the surface of a rigid sphere is that they permit steering a beam toward three-dimensional space with an almost identical beam-pattern, independent of the focused angle. Park and Rafaely (2005) validated the spherical microphone measurements in an anechoic chamber and measured the directional characteristics of reverberant sound fields. Rafaely (2005b) showed that spherical-harmonics and delay-sum beamforming provide similar performance when the highest spherical-harmonics order employed equals the product of the wave number and sphere radius. At lower frequencies, however, spherical harmonics beamforming allows the use of higher

a)Electronic mail: wksong@bksv.com

orders of spherical harmonics and thus better resolution. Note, though, that this improved resolution comes at the expense of robustness, i.e. the improvement of signal-to-noise ratio in the beamformer output.

Some studies examined the possibility of recording the higher-order spherical harmonics in a sound field and reproducing them by wavefield synthesis or ambisonics (Daniel *et al.*, 2003; Moreau *et al.*, 2006). But these methods require a large number of loudspeakers and a well-controlled environment such as an anechoic chamber. In order to render the recorded sound field binaurally, by contrast, the binaural signals obtained via either synthesis or recording can be played through a pair of headphones by feeding the left and right ear signal exclusively to each channel. Duraiswami and coworkers (Duraiswami *et al.*, 2005; Li and Duraiswami, 2005) studied theoretically how the free-field pressure obtained from spherical-harmonics beamforming (SHB) can be synthesized binaurally. The advantages of SHB, however, have not been demonstrated by means of psychoacoustic experiments in which subjective responses are collected to (a) validate the procedure, and (b) show that individual sources may successfully be isolated.

Therefore, the current study reports on a series of experiments to investigate the validity of using beamforming when auralizing a desired sound source in the presence of background noise or competing sources. The goals of this study are twofold:

1. To develop and verify the auralization of a desired source using beamforming. Procedures for estimating the pressure contribution of individual sources have already been suggested, but a scaling procedure will have to be developed to obtain the correct sound pressure level at the center of the array. To verify the procedure, the sound signals synthesized by beamforming will have to be compared with dummy head measurements.
2. To measure the effect of background noise suppression using beamforming on perceptual sound attributes, such as loudness and annoyance, derived from a listening experiment. To investigate the effects of noise suppression, the subjects' attention shall be controlled in such a way that they either judge the target sound (sound separated from background noise), or the entire sound mixture (including background noise).

To achieve these goals, the study employed ten stimuli from a study by Ellermeier *et al.* (2004a) which had been shown to cover a wide range with respect to loudness and annoyance. By playing them back in the presence of competing noise sources impinging from other directions, it may be investigated whether measuring the sound field with a spherical microphone array and processing it by SHB will recover the target source. Such a measurement protocol will be useful in situations in which only a desired source should be auralized, but in which background noise cannot be reduced or controlled during the measurement.

II. THEORETICAL BACKGROUND

A. Binaural synthesis

Reproduction of binaural signals via headphones is a convenient way of recreating the original auditory scene for the listener. The recording can be performed by placing a dummy head in a sound field, but it can also be synthesized on a computer. The binaural impulse response (BIR) from a "dry" source signal to each of the two ears in anechoic conditions can be described as (Møller, 1992):

$$h_{\text{left}}(t) = b(t) * c_{\text{left}},$$
$$h_{\text{right}}(t) = b(t) * c_{\text{right}}, \tag{1}$$

where the asterisk ($*$) represents convolution, b denotes the impulse response of the transmission path from a dry source signal to free-field pressure at the center of head position and c represents the impulse response of the transmission path from the free-field pressure to each of the two ears, i.e., head-related impulse response (HIR). The binaural signals can then be obtained by convolving a dry source signal with the binaural impulse response functions h. When using a spherical microphone array, SHB is able to approximate b for a given sound source by measuring the impulse response functions (IRF) from a dry source signal to each microphone of the array, and calculating the directional impulse response function (see Sec. II B 3) toward the dry source. The advantage of using SHB in comparison with a single-microphone measurement is the ability of focusing on a target source, i.e., obtaining the approximation of b, while suppressing background noise from other sources.

B. Spherical-harmonics beamforming

A theoretical description of SHB is presented in the following and a method to arrive at binaural auralization using SHB is proposed.

1. Fundamental formulation

For any function $f(\Omega)$ that is square integrable on the unit sphere, the following relationship holds (Rafaely, 2004):

$$F_{nm} = \oint f(\Omega) Y_n^{m*}(\Omega) d\Omega, \tag{2}$$

$$f(\Omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} F_{nm} Y_n^m(\Omega), \tag{3}$$

where the asterisk (*) represents complex conjugate, Y_n^m are the spherical harmonics, Ω is a direction, and $d\Omega = \sin\theta d\theta d\phi$ for a sphere. The spherical harmonics are defined as (Williams, 1999)

$$Y_n^m(\theta,\phi) = \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}} P_n^m(\cos\theta)\exp^{im\phi} \tag{4}$$

where n is the order, P_n^m are the associated Legendre polynomials, and $i=\sqrt{-1}$. Equation (3) shows that any square integrable function can be decomposed into spherical-harmonics coefficients. Rafaely (2004) defined the relation-

ship in Eqs. (2) and (3) as the spherical Fourier transform pair. The sound pressure on a hard sphere with radius $r=a$, $p(\Omega,a)$, and the directional distribution of incident plane waves, $w(\Omega)$, are square integrable and therefore we can introduce the two spherical transform pairs $\{p(\Omega,a),P_{nm}\}$ and $\{w(\Omega),W_{nm}\}$ according to Eqs. (2) and (3).

The goal of spherical-harmonics beamforming is to estimate the directional distribution $w(\Omega)$ of incident plane waves from the measured pressure on the hard sphere. To obtain a relation between the pressure on the sphere and the angular distribution of plane waves, we consider first the pressure on the hard sphere produced by a single incident plane wave. The pressure $p_\ell(\Omega_\ell,\Omega)$ on the hard sphere induced by a single plane wave with a unit amplitude and incident from the direction Ω_ℓ can be described as (Williams, 1999)

$$p_\ell(\Omega_\ell,\Omega)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}R_n(ka)Y_n^{m*}(\Omega_\ell)Y_n^m(\Omega), \tag{5}$$

where k is the wave number, and R_n is the radial function:

$$R_n=4\pi i^n\left[j_n(ka)-\frac{j_n'(ka)}{h_n^{(1)\prime}(ka)}h_n^{(1)}(ka)\right]. \tag{6}$$

Here, j_n is the spherical Bessel function, $h_n^{(1)}$ the spherical Hankel function of the first kind, and j_n' and $h_n^{(1)\prime}$ are their derivatives with respect to the argument. The total pressure $p(\Omega,a)$ on the hard sphere created by all plane waves can be found then by taking the integral over all directions of plane wave incidence. Using Eq. (5) and the spherical Fourier transform pair of $w(\Omega)$ we get

$$\begin{aligned}p(\Omega,r=a)&=\oint p_\ell(\Omega_\ell,\Omega)w(\Omega_\ell)d\Omega_\ell\\&=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}R_n(ka)Y_n^m(\Omega)\oint w(\Omega_\ell)Y_n^{m*}(\Omega_\ell)d\Omega_\ell\\&=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}W_{nm}R_n(ka)Y_n^m(\Omega).\end{aligned} \tag{7}$$

By comparing Eq. (7) with the spherical Fourier transform pair of $p(\Omega,a)$, the spherical Fourier transform coefficients of $w(\Omega)$ can be obtained as

$$W_{nm}=\frac{P_{nm}}{R_n(ka)}. \tag{8}$$

Substituting these coefficients in the spherical Fourier transform pair of $w(\Omega)$ results in

$$w(\Omega)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\frac{P_{nm}}{R_n(ka)}Y_n^m(\Omega). \tag{9}$$

This shows that the directional distribution of plane waves can be obtained by dividing the pressure coefficients P_{nm} by the radial function R_n in the spherical Fourier domain.

We now introduce a set of M microphones mounted at directions Ω_i, $i=1,\ldots,M$, on the hard sphere with radius a. The Fourier transform expression for P_{nm} has the form of a continuous integral over the sphere, but the sound pressure is known only at the microphone positions. Therefore, we must use an approximation of the form:

$$P_{nm}\approx\tilde{P}_{nm}\equiv\sum_{i=1}^{M}c_i p(\Omega_i)Y_n^{m*}(\Omega_i). \tag{10}$$

The weights c_i applied to the individual microphone signals and the microphone positions Ω_i are chosen in such a way that

$$H_{mn\mu\nu}\equiv\sum_{i=1}^{M}c_i Y_\nu^{\mu*}(\Omega_i)Y_n^m(\Omega_i)=\delta_{\nu n}\delta_{\mu m} \tag{11}$$

$$\text{for } n\leqslant N,\nu\leqslant N,$$

where N is the maximum order of spherical harmonics that can be integrated accurately with Eq. (10). The value of N will depend on the number M of microphones. Therefore, the beamformer response for the direction Ω is calculated by substituting Eq. (10) in Eq. (9) and by limiting the spherical harmonics order to N:

$$b(\Omega)\equiv\sum_{i=1}^{M}\left[\sum_{\nu=0}^{N}\frac{1}{R_\nu(ka)}\sum_{\mu=-\nu}^{\nu}c_i Y_\nu^{\mu*}(\Omega_i)Y_\nu^\mu(\Omega)\right]p(\Omega_i). \tag{12}$$

2. Pressure scaling

Equation (12) is the typical beamformer output, but does not provide the correct pressure amplitude of an incident plane wave. Therefore, the goal here is to derive a scaling factor that gives rise to the correct estimate of the pressure amplitude. Ideally one may derive the scaling factor for each focus direction by calculating the beamformer response to a plane wave incident from that direction. Such a procedure would, however, significantly increase the computational effort. In particular at the lower frequencies, where the spatial aliasing is very limited, the "in-focus plane wave response" is fairly independent of the focus angle of the beamformer. One could therefore calculate the in-focus plane wave response for a single focus direction and apply that quantity for scaling of the beamformer output for all focus directions. But as shown in the following, it is possible to derive an analytical expression for the angle-averaged in-focus plane wave response. Use of that simple analytical expression requires less computation and provides a scaling that is better as an average over all directions.

We assume now a plane wave incident with a unit amplitude from the direction Ω_ℓ. By inserting Eq. (5) in Eq. (12) followed by use of Eq. (11) we get the beamformer response for an arbitrary focus direction Ω:

$$b(\Omega,\Omega_\ell)=\sum_{\nu=0}^{N}\sum_{\mu=-\nu}^{\nu}Y_\nu^\mu(\Omega)\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\frac{R_n}{R_\nu}Y_n^{m*}(\Omega_\ell)\times\sum_{i=1}^{M}c_iY_\nu^{\mu*}(\Omega_i)Y_n^m(\Omega_i)=\sum_{\nu=0}^{N}\sum_{\mu=-\nu}^{\nu}Y_\nu^\mu(\Omega)\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\frac{R_n}{R_\nu}Y_n^{m*}(\Omega_\ell)H_{mn\mu\nu}. \tag{13}$$

Only the in-focus response is needed, i.e., in the direction of plane wave incidence, $\Omega=\Omega_\ell$. This response will have a fairly constant amplitude and phase independent of the angle of the plane wave incidence, so it can be well represented by the angle averaged response $\bar{b}$. When we perform such an averaging, we can make use of the following orthogonality of the spherical harmonics:

$$\oint Y_\nu^\mu(\Omega)Y_n^{m*}(\Omega)d\Omega=\delta_{\nu n}\delta_{\mu m}. \tag{14}$$

Use of Eq. (14) in connection with Eq. (13) leads to the following expression for the angle averaged in-focus response,

$$\bar{b}\equiv\frac{1}{4\pi}\oint b(\Omega_\ell,\Omega_\ell)d\Omega_\ell=\frac{1}{4\pi}\sum_{\nu=0}^{N}\sum_{\mu=-\nu}^{\nu}H_{\mu\nu\mu\nu}. \tag{15}$$

And if in Eq. (15) we use Eq. (11), we get

$$\bar{b}\equiv\frac{(N+1)^2}{4\pi} \tag{16}$$

provided N is not larger than the spherical-harmonics order the beamformer was designed for, see Eq. (11). Equation (16) provides the average beamformer output, when focusing at infinite distance toward an incident plane wave of unit amplitude. If we wish the response to equal the amplitude of the incident plane wave, we therefore have to divide the output by $\bar{b}$ of Eq. (16). Notice that Eq. (15) shows a general approach, which may be applied to frequencies higher than those the microphone array is designed for. However, assuming no spatial aliasing (i.e., $R_n(ka)=0$ for $n>N$) the array beam pattern is independent of the focused direction. This means that Eq. (16) may be derived directly by substituting Eq. (11) in Eq. (13) and subsequently by using the spherical harmonics addition theorem [Rafaely, 2004, Eq. (20)].

So far we have considered plane wave incidence and focusing at an infinite distance. Consider instead the case of a monopole point source and focusing of the beamformer at the distance r_0 of the point source. The free-field sound pressure produced at the origin by this monopole is

$$p_{\text{center}}=\frac{e^{ikr_0}}{kr_0}. \tag{17}$$

The sound pressure at the microphone positions on the hard sphere can be expressed in spherical harmonics as in Eq. (5), but now with the following radial function (Bowman *et al.*, 1987):

$$R_n(ka)=4\pi i h_n^{(1)}(kr_0)\left[j_n(ka)-\frac{j_n'(ka)}{h_n^{(1)\prime}(ka)}h_n^{(1)}(ka)\right]. \tag{18}$$

Using the radial function of Eq. (18) in the beamforming processing, and averaging over all directions for the point source, leads to the same average in-focus beamformer output as in Eq. (16). If we wish the output to be the free-field pressure at the center of the array [Eq. (17)], then we have to scale the beamformer output by the following factor:

$$\frac{4\pi e^{ikr_0}}{(N+1)^2kr_0}. \tag{19}$$

3. Binaural auralization using SHB

Scaling the beamformer output by Eq. (19) provides the directional free-field pressure contributions at the center position in the absence of the array. Beamforming measurement and processing should then be taken for each sound event to be reproduced by the loudspeaker setup (described in Sec. III C): The type of sound cannot be changed after the measurement is done. But performing the measurement and processing for each sound is very time consuming. For this reason, directional impulse response functions will be calculated and used for simulating the total transmission from each loudspeaker input to each of the two ears.

Provided we measured the frequency response function (FRF) $t(\Omega_i)$ from a loudspeaker input to each microphone position on the sphere, the coefficients of the loudspeaker FRF's spherical Fourier transform T_{mn} can then be obtained by replacing $p(\Omega_i)$ by $t(\Omega_i)$ in Eq. (10),

$$T_{nm}\equiv\sum_{i=1}^{M}c_it(\Omega_i)Y_n^{m*}(\Omega_i). \tag{20}$$

Substituting Eq. (20) in Eq. (9) yields the directional response of the beamformer,

$$s(\Omega)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\frac{T_{nm}}{R_n(ka)}Y_n^m(\Omega). \tag{21}$$

The directional impulse response can then be obtained by taking the inverse temporal fast Fourier transform (FFT) of $s(\Omega)$. If there is more than one loudspeaker, then the contribution from sound sources in other directions than the one focused on has to be taken into account and the total output of the beamformer at a particular direction Ω can be expressed as

$$y(\Omega)=\sum_{\ell=1}^{N_d}s_\ell(\Omega)x_\ell, \tag{22}$$

where N_d denotes the number of loudspeakers, $s_\ell(\Omega)$ represents the directional response of the ℓth loudspeaker in the focused direction Ω, and x_ℓ is the input signal of the ℓth loudspeaker. This will be a fairly good approximation since the contribution from other directions than those of the sources is negligible. Finally, the binaural signal can be obtained by multiplying $y(\Omega)$ with the HRTFs in the focused direction Ω.

Park and Rafaely (2005) suggested that the maximum spherical harmonics order in SHB should be limited to $N \leqslant ka$ in order to avoid noise originating from the high-order spherical harmonics. With the spherical microphone array used in this study, this would cause the beamformer output to become omnidirectional below 390 Hz. However, it was found that the order-limiting criterion can be relaxed in the following way without generating a high noise contribution:

$$N = \begin{cases} \lfloor ka \rfloor + 1, & \lfloor ka \rfloor + 1 \leqslant N_{\max} \\ N_{\max}, & \lfloor ka \rfloor + 1 > N_{\max} \end{cases} \tag{23}$$

where $\lfloor ka \rfloor$ represents the largest integer smaller than or equal to ka, and $N_{\max}$ is the maximum order of spherical harmonics for which the array can provide accurate integration [see Eq. (10)]. The number of spherical harmonics, $(N+1)^2$, should not exceed the number of microphones, and therefore $N_{\max}$ should be 7 in the current study where 64 microphones were used. Relaxing this condition by introducing higher spherical harmonic orders will reduce robustness and introduce greater uncertainties but our measurement and simulation experience shows that the use of spherical harmonic orders equal to $ka+1$ [as defined in Eq. (23)] produces only minor numerical instabilities.

C. Psychoacoustical considerations

The goal of the empirical part of the present study is to validate the beamforming method proposed, and—more specifically—to show how its use will help to psychoacoustically characterize target signals in a background of noise.

While, from a methodological perspective, it may be interesting to investigate the *detectability* of a target source in the presence of noise, in practice, the sources of interest are almost always well above threshold, or at best partially masked. Often, the focus of industrial applications is restricted to identifying the most problematic source in a mixture (Hald *et al.*, 2007; Nathak *et al.*, 2007), and to modify it to reduce its negative impact. Therefore, from a psychoacoustical perspective, some kind of suprathreshold subjective quantification of the salience of the target source in the background noise is called for. For the present investigation, the suprathreshold attributes of loudness and annoyance were chosen, since the former has been extensively studied (for reviews, see Moore, 2003; Zwicker and Fastl, 2006), and the latter is of particular relevance for noise control engineering (e.g., Marquis-Favre *et al.*, 2005; Versfeld and Vos, 1997).

As will be detailed in Sec. III, a between-subjects design was employed, investigating the two attributes in two independent groups of listeners. This was done in order to avoid potential carry-over effects that might produce artifactual correlations between loudness and annoyance.

Measuring the loudness or annoyance of the target stimuli under various conditions of partial masking required a scaling method that is relatively robust with respect to changes in context. A two-step category scaling procedure that uses both initial verbal labels to "anchor" the judgments and subsequent numerical fine-tuning possesses this property. It has been shown (Ellermeier *et al.*, 1991; Gescheider, 1997) to largely preserve the "absolute" sensation magnitudes even if the experimental context is changed. It was felt that the most widespread suprathreshold scaling procedure, namely Stevens' magnitude estimation, by virtue of the instructions to judge ratios of successive stimuli would encourage "relative" judgment behavior which might make it hard to compare the results across the different auralization methods used. Finally, the chance that in some conditions the target sounds might be entirely masked (yielding judgments of zero or undefined ratios), appeared to make ratio instructions unfeasible.

III. METHOD

A. Subjects

Twenty-eight normal-hearing listeners between the age of 21 and 34 (12 male, 16 female) participated in the experiment. All listeners were students at Aalborg University except for one female participant. The subjects' hearing thresholds were checked using standard pure-tone audiometry in the frequency range between 0.25 and 8 kHz and it was required that their pure-tone thresholds should not fall more than 20 dB below the normal curve (ISO 1998) at more than one frequency. The subjects were also screened for known hearing problems and paid an hourly wage for their participation. The subjects were not exposed to the sounds employed prior to the experiment.

B. Apparatus

The experiment was carried out in a small listening room with sound-isolating walls, floors, and ceiling. The room conforms with the ISO (1992) standard. The listeners were seated in a height-adjustable chair with a headrest. They were instructed to look straight ahead and were not allowed to move their head during the experiments. Their head movement was monitored by a camera installed in the listening room. Two monitors, one in the control room and the other in the listening room, were displayed at the same time with the help of a VGA splitter. A small loudspeaker placed in the control room played the same sound as the subject listened to so the experimenter could monitor the sound playback and the listener's behavior.

A personal computer with a 16-bit sound card (RME DIGI96) was used for D/A conversion of the signals. The sound was played with a sampling rate of 48 kHz and delivered via an electrostatic headphone (Sennheiser HE60) connected through an amplifier (Sennheiser HEV70) with a fixed volume control to assure constant gain. An external amplifier (t.c. Electronic Finalizer) between the headphone amplifier and the sound card controlled the playback level.

Playback and data collection were controlled by a customized software developed in C#. The software read the session files to assign a subject to the defined session, played the stimuli using the ASIO driver, collected subjects' responses, and wrote the responses into text files.

C. Measurements

The three different types of measurements, i.e., microphone, dummy head, and spherical microphone array, were

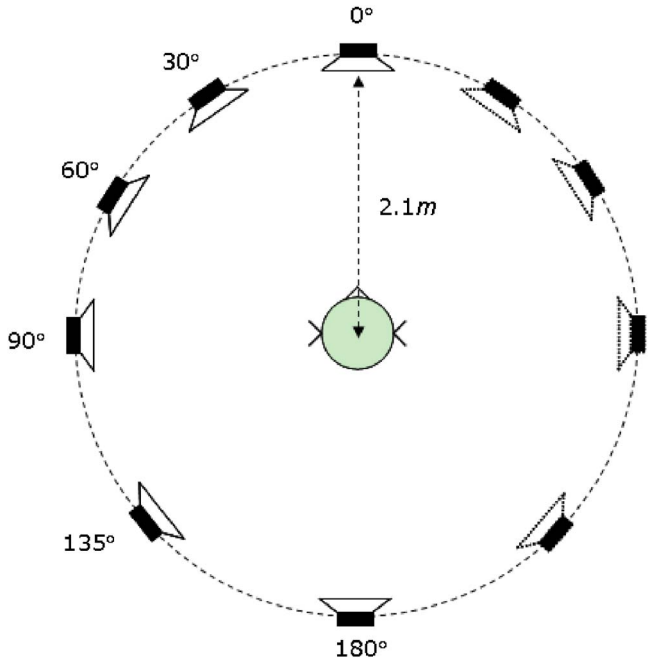

FIG. 1. (Color online) The loudspeaker setup in the anechoic chamber.

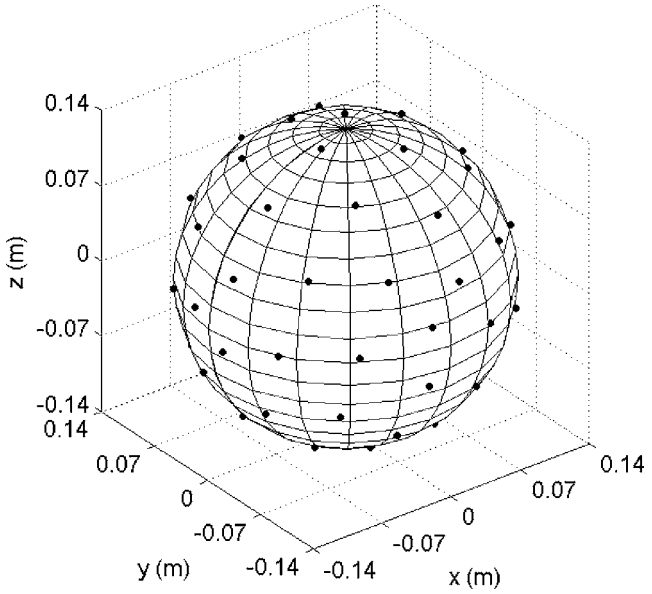

FIG. 2. The array consisting of 64 microphones placed on the hard surface of a sphere having a 14-cm radius. The dots on the sphere indicate the microphone positions.

performed in an anechoic chamber. Six loudspeakers were positioned at 2.1 m away from the center of the setup and their positions are shown in Fig. 1 (placed on the left-hand side). A setup of ten loudspeakers was simulated by flipping the four loudspeakers to the right-hand side. The loudspeaker in the frontal direction was used as the desired source through which the recorded sounds were synthesized and the rest of the loudspeakers served to create background noise sources. Since the microphone array and the required hardware was available for a very limited time, it was decided to record time data to permit changing some of the parameters without repeating the measurements. The input and output time data were recorded by means of the Data Recorder in the Brüel & Kjær software (type 3560) with a frequency range of 6.4 kHz. The loudspeakers were excited by random pink noise. The IRFs between speaker excitations and microphone responses were calculated using the autospectrum and cross spectrum of input and output and taking the inverse FFT of the calculated frequency response function in MATLAB. In order to remove the influence of reflections caused by the supporting structure and by other loudspeakers than the measured one, an 8-ms time window was applied to the calculated IRFs.

The loudspeaker responses were measured at the center position of the setup using a 1/2-in. pressure field microphone (Brüel & Kjær type 4134). The microphone was placed at 90° incidence to the loudspeakers during the measurement with the help of three laser beams mounted in the room. The measured IRFs were compared with the simulated ones to validate the recorded sound field using SHB. Responses of each loudspeaker at each ear of a dummy head were measured by placing the artificial head VALDEMAR (Christensen and Møller, 2000) at the center of the loudspeaker setup. Care was taken that the IRFs in both ears have the same delay when measuring the loudspeaker response in the frontal direction. The dummy head measurements were compared with the ones synthesized from SHB. The HRTFs employed in this study to perform binaural synthesis using SHB were taken from a database containing artificial-head HRTFs measured at 2° resolution (Bovbjerg *et al.*, 2000; Minnaar, 2001).

IRFs of each loudspeaker at the microphones of the array were obtained by positioning a spherical microphone array at the center of the setup. The position of the microphone array was adjusted carefully so that the beamformed sound pressure mapping could localize the correct angular position of each loudspeaker. The microphone array with a radius of 14 cm consisted of 64 microphones (1/4-in. microphone, Brüel & Kjær type 4951) that were evenly distributed on the surface of the hard sphere in order to achieve the constant directivity pattern in all directions. Figure 2 displays the position of microphones marked by dots on a sphere. In an earlier study, the array was applied to the issue of noise source localization, and the detailed specifications and characteristics of the array are described in Petersen (2004). In total, six loudspeaker positions and 64 microphones produced 384 IRFs.

The headphone transfer functions (PTFs) were measured in the listening room with the same dummy head and equipment used for the IRF measurement. The PTF measurement was repeated five times and after each measurement the headphone was repositioned. The upper panel of Fig. 3 shows that the repetitions have similar spectral shape in the frequency range of the investigation. An average of these five measurements was taken and smoothed in the frequency domain by applying a moving average filter corresponding to the 1/3 octave bands. The inverse PTF was calculated from the average PTF using fast deconvolution with regularization (Kirkeby *et al.*, 1998) (see the lower panel of Fig. 3).

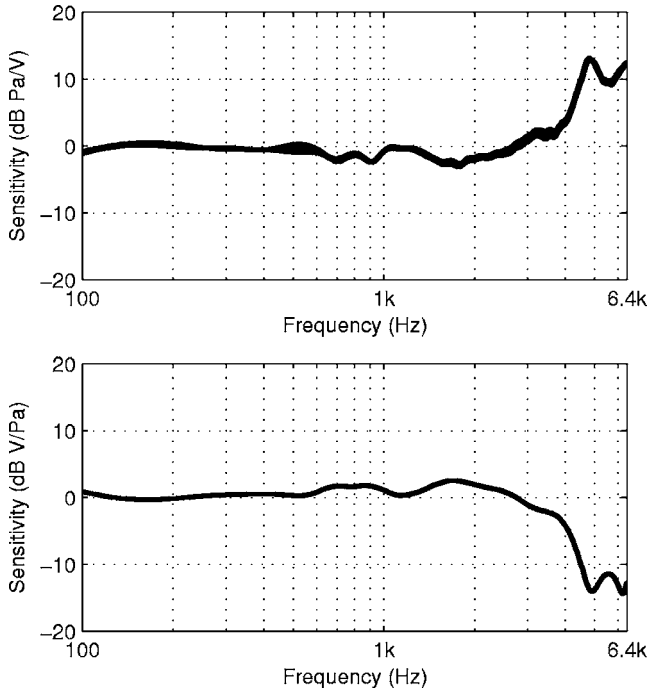

FIG. 3. Five headphone transfer functions (upper panel) measured at the left ear of the dummy head and the inverse filter derived (lower panel).

D. Stimuli

A set of 10 environmental and product sounds was selected from the 40 stimuli used by Ellermeier *et al.* (2004a). The ten sounds chosen were recorded in a sound-insulated listening room, except for two outdoor recordings of automotive sounds. About half of them were everyday sounds (e.g., door knocking, water pouring) and the rest were product sounds (e.g., kitchen mixer, razor, car). Both the perceived loudness and the annoyance of the selected sounds were almost equally spaced according to the attribute scales obtained in the reference study (Ellermeier *et al.*, 2004a). The length of stimuli varied from 0.8 to 5 s, and their overall sound pressure level at the recorded position ranged from 45 to 75 dB SPL. The sounds had a sampling rate of 44.1 kHz originally, but were resampled to 48 kHz in order to meet the requirements of the listening test program.

The desired source was synthesized to be located in the frontal direction and the remaining nine loudspeakers generated background noise. The selected sounds were convolved with the dummy head IRFs in the frontal direction to obtain the desired stimuli, and white noise having the same duration as the target sounds was convolved with the dummy head IRFs corresponding to the other nine directions. For each loudspeaker position, a new random sequence of white noise was created, and the signals convolved with the BIR at each ear were simply added to obtain the background noise. By doing so, the generated background noise was perceived to be diffuse. Two different levels of background noise were employed. The low level of background noise was adjusted to have the same sound pressure level as the bell sound (62 dB SPL), which was located in the middle of the attribute scale and the high level was defined to be 10 dB higher than the low one. In this way, the effect of the background noise level could be investigated. It was expected that some of the sounds would be partially masked by the background noise thereby affecting the attribute-scale responses.

The directional pressure contribution was obtained by recording the sound field using the spherical microphone array and applying SHB to the recorded data. Thus directional impulse response functions were calculated by using the IRFs at each of the microphone positions on the sphere as input to SHB processing. The resulting directional impulse response functions were convolved with HRTFs in the frontal direction to obtain the binaural IRFs, which still contain the contributions from background noise sources, though greatly reduced by the beamforming. In this case, the perception of the background noise is different from that with traditional binaural synthesis in that the noise is perceived to originate from the frontal direction. Thus in this study the influence of the level and perceptual quality of the background noise are confounded.

Subjects were asked to judge either the annoyance or the loudness of 50 stimuli, which were produced by combining three different processing modes (original, original+noise, SHB+noise), with two different noise levels, for the ten sounds selected. The same calibration tone as in the reference study (Ellermeier *et al.*, 2004a) was used and the level at the center position of the loudspeaker setup was adjusted to be 88 dB SPL when playing the calibration tone. A 100-ms ramp was applied to the beginning and end of each stimulus in order not to generate impulsive sounds. The inverse PTF was applied to the stimuli as a final step of the processing.

E. Procedure

The subjects were randomly assigned to one of two groups, one judging the loudness, the other the annoyance of the sounds. During the experiment, the participants were instructed to judge the entire sound event in one session, and the target sound only in the other. When judging the target sound only, they were asked to ignore the background noise and not to give ratings based on the direct comparison between the target sound and the background noise. The listeners were instructed to combine any of the components they heard for rating the entire sound mixture. These two ways of judging the sound attributes were chosen to check whether the effect of suppressing the background noise by SHB processing is different dependent on which part of a stimulus is being judged.

In each group, half of the subjects started judging the target sound only and proceeded to judge the entire sound (target plus background). The other half completed those two tasks in the opposite order. Note that each subject made but a single rating of each of the 50 experimental stimuli, i.e., there were no repetitions. The subjects spent approximately 1.5 h to complete the experiment. The participants were asked to judge either the loudness or the annoyance of the sounds by using a combined verbal/numerical rating scale, i.e., category subdivision (see Ellermeier *et al.*, 1991), shown in Fig. 4.

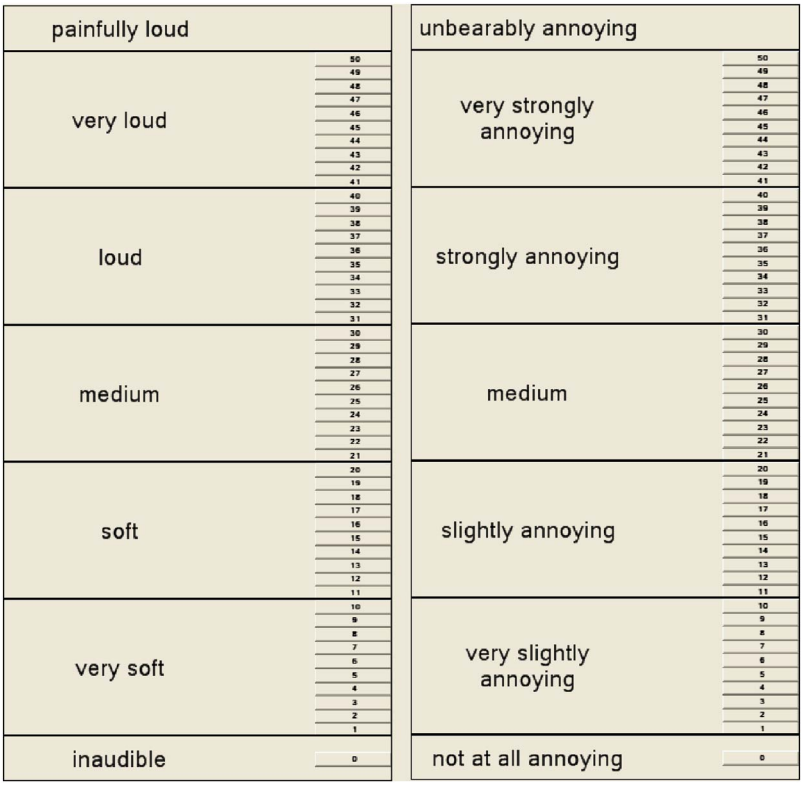

FIG. 4. (Color online) Category subdivision scales for loudness (left) and annoyance (right).

1. Training

There were two types of training prior to the main experiment. The goal of the first training unit was to give the subjects an opportunity of listening to the target sounds and to get an idea on what they had to focus, if the target was presented in background noise. To that effect, 20 buttons were displayed on a PC screen in two columns. The first column was labeled "target sound" and the second one "target sound+noise." The noise level was randomly selected from either the high- or the low-level condition. The participants were asked to first listen to the target sound only and then to the target sound with noise. During the training, the experimenter was present in the listening room and the subjects could ask any questions related to the understanding of the task. During the second training unit, the subjects received practice with rating the attribute, e.g., loudness or annoyance, of either target sound only or the entire sound dependent on which session they started with. The aim was to familiarize the participants with the procedure. This training unit consisted of only ten stimuli sampled to cover the entire range of sound pressure levels.

If the subjects started with judging the entire sound, they completed the training on the rating procedure first and were practiced in distinguishing target and background before starting with the second part of the experiment. Subjects, who judged the target sound in the first block, finished the two training units in a sequence prior to the main experiment.

2. Loudness scaling

For loudness scaling, the scale shown in Fig. 4 was displayed on a computer screen together with a reminder indicating whether they have to judge the target sound or the entire sound. The scale consisted of five verbal categories which were subdivided into ten steps and labeled "very soft" (1–10), "soft" (11–20), "medium" (21–30), "loud" (31–40), and "very loud" (41–50). The end points of the scale were used and labeled as "inaudible" (0) and "painfully loud" (beyond 50). On each trial, one sound was presented at a time, and the subjects were asked to decide which category the sound belonged to and then to fine-tune their judgment by clicking a numerical value within that category. That input started the next trial with a 1-s delay. The subjects were not allowed to make their rating while a sound was played. In order to avoid the situation where subjects rated the target sounds based on identifying them and recalling previous ratings, they were told that the level of the target sound might vary between trials.

3. Annoyance scaling

The format of the annoyance scale used was the same as that of the loudness scale (see Fig. 4). The five verbal categories were "very slightly annoying" (1–10), "slightly annoying" (11–20), "medium" (21–30), "strongly annoying" (31–40), and "very strongly annoying" (41–50). The lower end point was labeled as "not at all annoying" (0) and the higher one "unbearably annoying" (beyond 50). In the target sound only session, an "inaudible" button was placed below the category scale and subjects were asked to press it when they could not detect the target sound due to strong background noise.

The annoyance instructions were based on proposals by Berglund *et al.* (1975) and Hellman (1982). That is, a scenario was suggested, leading the participants to imagine a

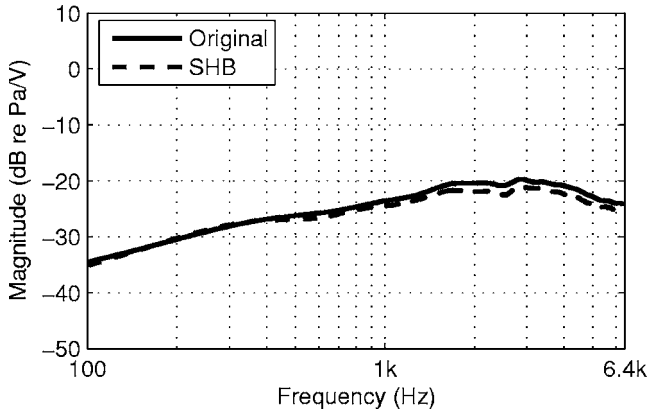

FIG. 5. Free-field loudspeaker response (30°): Measured (solid line) and synthesized (dashed line) using SHB.

situation in which the sounds could interfere with their activity: "After a hard day's work, you have just been comfortably seated in your chair and intend to read your newspaper."

IV. RESULTS

Here, the simulated sound field using SHB is compared with both the microphone and the dummy head measurements to illustrate the expected level difference induced by the beamforming in monaural and binaural responses. Moreover, the discrepancies in perceptual quality among the processing modes are demonstrated in both loudness and annoyance ratings obtained in the listening experiments.

A. Recording the sound field using SHB

In order to evaluate the success of the SHB simulation, the simulated and measured loudspeaker responses were compared. The loudspeaker responses at the 64 microphones placed on the sphere were measured and used as the input to the SHB calculation. The directional impulse response function toward each loudspeaker was calculated and compared with the direct measurement using a microphone positioned at the center position of the setup. The simulated and measured responses were compared in the frequency range of interest from 0.1 to 6.4 kHz, and an example for the loudspeaker placed at 30° is displayed in Fig. 5.

Generally, the agreement between the simulated and measured responses was good and the maximum discrepancy was approximately 2 dB in all loudspeaker directions. There was a tendency for the error to increase at high frequencies. In the current investigation where N_{max} is 7 and the radius of the array is 14 cm, spatial aliasing is expected above 2.7 kHz and thereby corrupts the spatial response. This could be the main reason for the inaccuracies at high frequencies.

The binaural response to the six loudspeakers was simulated by convolving the directional impulse response with the HRTF for the same direction as the loudspeaker (see Sec. II B 3). Subsequently, the simulated responses were compared with those measured with a dummy head and an example of the results is displayed in Fig. 6. The graphs represent the combination of the free-field loudspeaker response and the HRTF. In general, the two curves have similar shape and amplitude and the same tendency as for the free-field

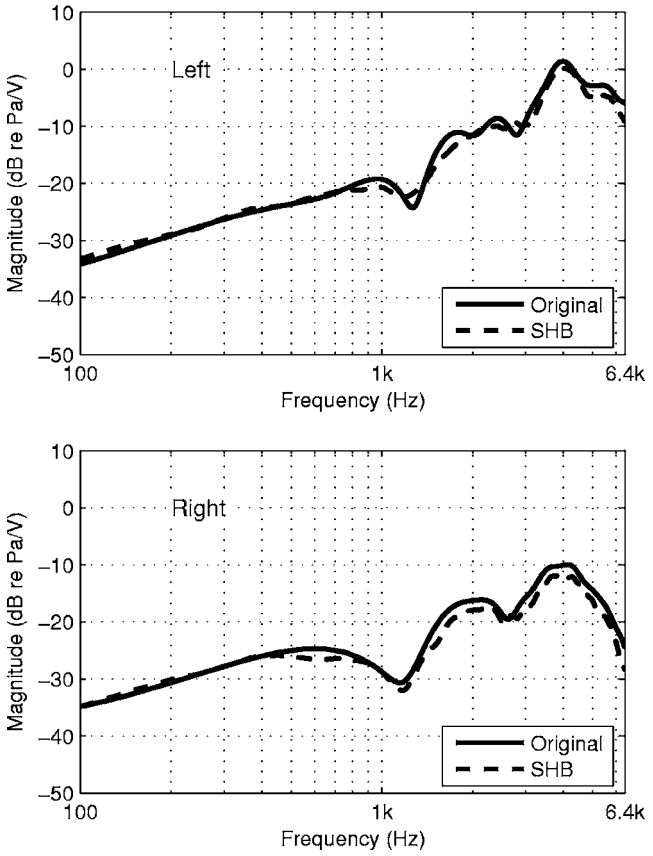

FIG. 6. Loudspeaker response (30°) at both ears: Measured (solid line) and synthesized (dashed line) using SHB.

response was observed, i.e., that the error grows slightly at high frequencies. These investigations confirm that the proposed method of combining SHB and binaural synthesis can generate binaural signals physically close to the measured ones.

B. Signal-to-noise ratio

The two measurement techniques, i.e., based on a dummy head and SHB, respectively, may be compared physically in terms of their monophonic signal-to-noise (S/N) ratios for each sound sample in the noisy conditions. Since the monophonic response for each loudspeaker was estimated both with a single microphone and with a microphone array, it was possible to separate the pressure contribution of the sound samples presented in the frontal direction from that of the noises in other directions. The monophonic S/N ratio for each sound sample was calculated simply by dividing the rms pressure of the signal by that of the noise.

Figure 7 shows the resulting S/N ratios of dummy head (original+noise) and SHB synthesis in both background noise conditions. The lower panel indicates the results of the low level noise condition and the upper panel the high level one. Notice that the S/N ratio of the bell sound is 0 and −10 dB in the original+noise condition for the low and high background noise levels, respectively (see Sec. III D). In general, the S/N ratio increases monotonically for the sounds ordered along the abscissa and there is a constant 10 dB difference between the low and high background noise con-

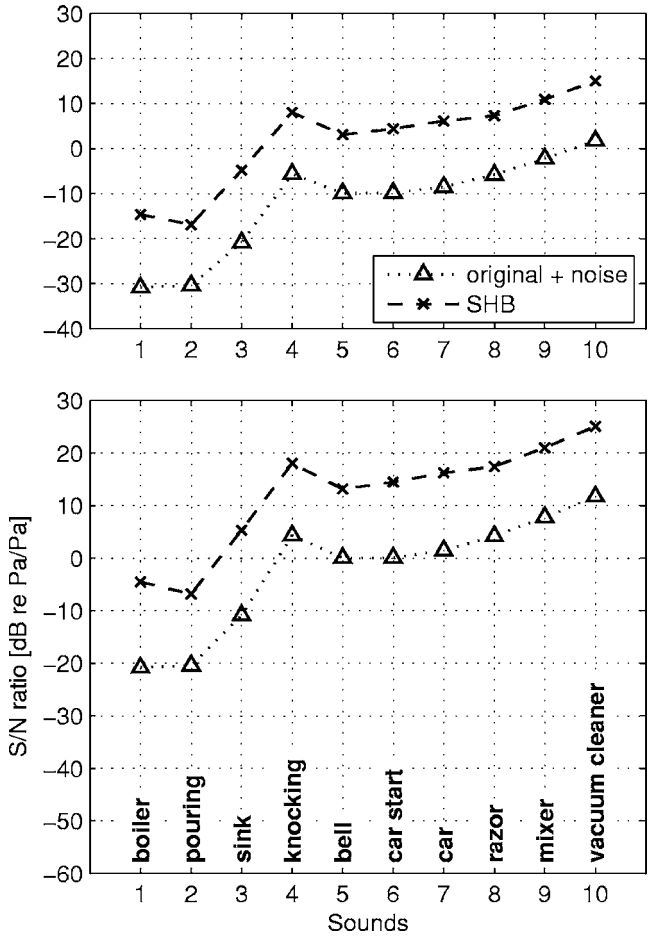

FIG. 7. Monophonic S/N ratio of dummy head (original+noise) and SHB measurements in the low (lower panel) and high (upper panel) background noise conditions.

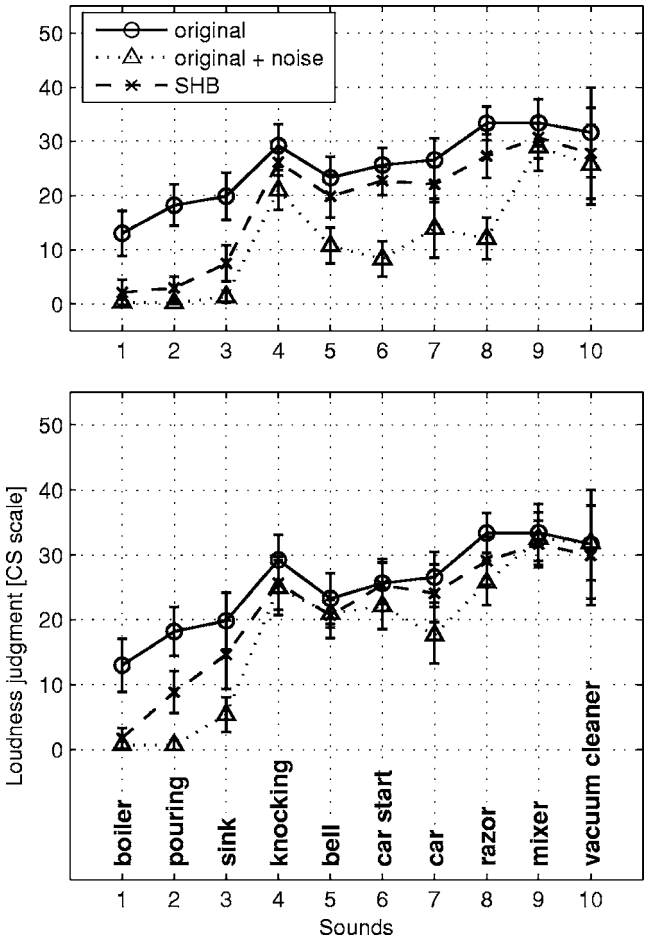

FIG. 8. Loudness judgments of the ten test sounds in the low (lower panel) and high (upper panel) background noise condition. The target sounds are labeled along the abscissa and the error bars indicate 95%-confidence intervals. Listeners focused on the target sound only.

ditions. Thus, the effect of noise on the psychoacoustical scales is expected to be dominant in the low level sounds, e.g., for sound 1 to 3, for both measurement techniques. SHB increases the S/N ratio by approximately 15 dB for all sound samples, and thus the effect of the noise on loudness will be smaller for SHB in comparison with the dummy head technique.

C. Loudness scaling

The subjective loudness judgments were averaged across the 14 subjects for each sound in the three processing modes (original, original+noise, SHB) and 95%-confidence intervals were determined. The outcome is plotted in Fig. 8, for judgments of the target sound only, and in Fig. 9, for judgments of the entire sound event. The upper graph in Figs. 8 and 9 represents the high background noise condition and the lower graph the low background noise condition. Both graphs share the same ratings for the original condition plotted with solid lines. The sounds on the abscissa were arranged in the order of the mean ratings obtained in the reference study (Ellermeier *et al.*, 2004a). It appears that the present sample of subjects judged the knocking sound to be somewhat louder than in the reference study.

In the "target sound only" conditions (see Fig. 8), the target loudness was considerably reduced by adding noise to the target sound (compare the dotted and solid line) due to partial masking. It appears that SHB (dashed line in Fig. 8) partially restored the loudness of the target sounds. This was confirmed by performing a three-factor analysis of variance[1] (ANOVA) (Montgomery, 2004) with the two processing modes (SHB; original+noise), the two noise levels, and the ten sounds all constituting within-subjects factors. The analysis showed a highly significant main effect of processing mode [$F(1,13)=44.5$, $p<0.001$], as well as significant interactions ($p<0.001$) of processing mode with all other factors. That suggests that SHB did indeed suppress the background noise, thereby partially restoring loudness to the original levels. With the low-level masking noise (lower panel of Fig. 8) that was true for relatively "soft" target sounds (pouring and sink) while with the high-level masking noise (upper panel) the "loud" targets were the ones benefiting most from the release from masking produced by the SHB auralization. Notice on the other hand the difference between the two synthesis techniques in terms of S/N ratio is almost constant across different sounds and not dependent on the background noise level (see Fig. 7). Thus, a simple objective measure such as S/N ratio may not be suitable for predicting the effect of background noise suppression using beamforming on psychoacoustic attributes. Most subjects, however, could not de-

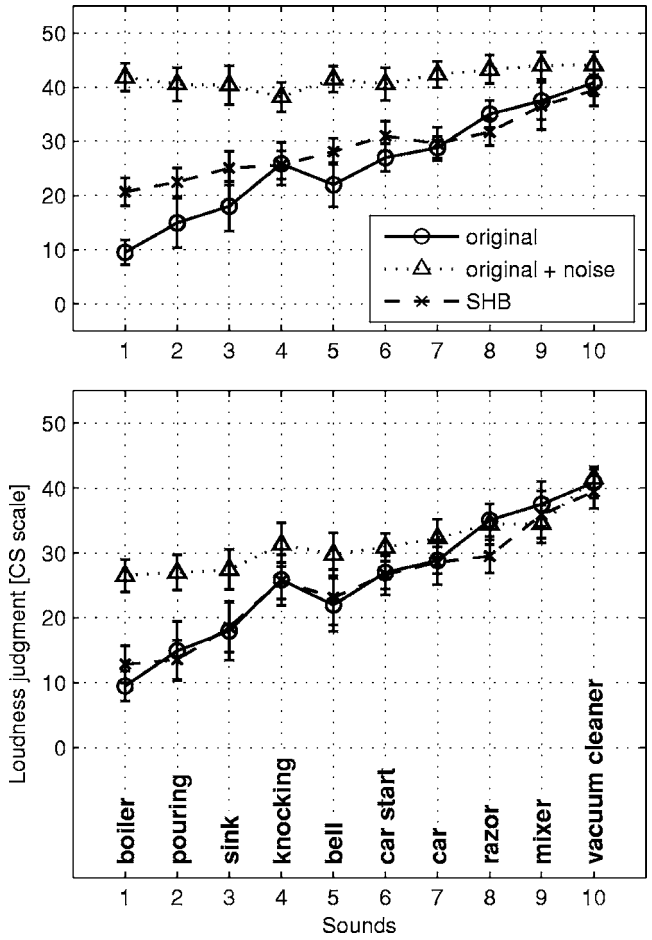

FIG. 9. Loudness judgments of the ten test sounds in the low (lower panel) and high (upper panel) background noise condition. The target sounds are labeled along the abscissa and the error bars indicate 95%-confidence intervals. Listeners judged the entire sound event.

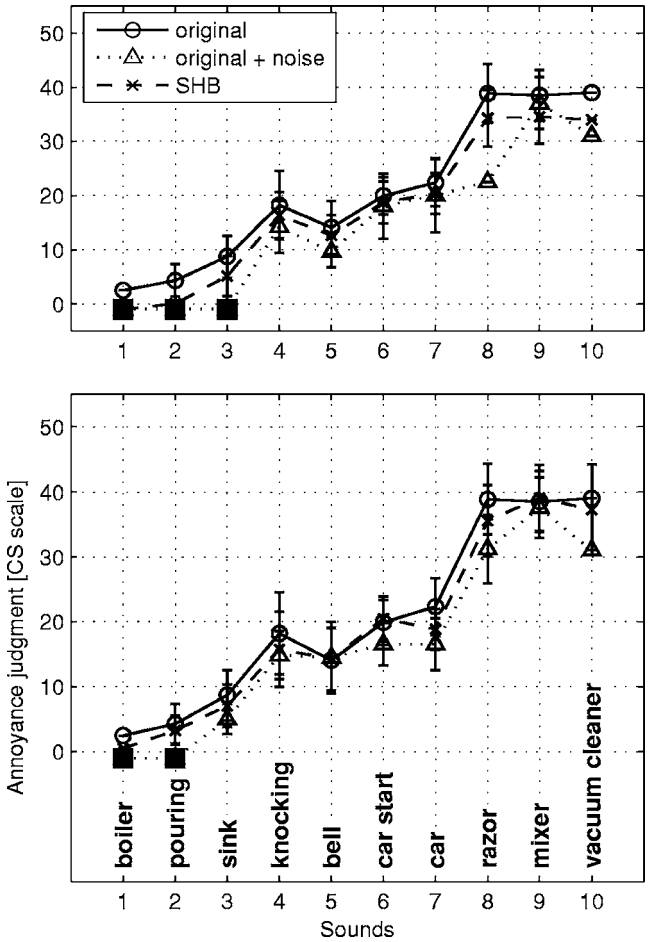

FIG. 10. Annoyance judgments of the ten test sounds in the low (lower panel) and high (upper panel) background noise condition. The target sounds are labeled along the abscissa and the error bars indicate 95%-confidence intervals. Listeners focused on the target sound only. If the majority of the participants did not hear the target, the data points were marked with closed squares.

tect the "boiler" sound in both noise conditions since this sound was completely masked by background noise. This may be seen in Fig. 7 in that for the boiler sound a very low S/N ratio was obtained, even after the processing. Furthermore, the subjective ratings of the "knocking" sound almost coincided with those of the original sound, revealing that the subjects extracted this impulsive sound from the background much easier than other sounds. The high confidence intervals obtained for the vacuum-cleaner sound occurred because the target sound was so similar to background noise that it was difficult to distinguish one from the other.

Judging the entire sound event (see Fig. 9) made the suppression of the masker even more obvious in that the loudness functions for the original and SHB conditions almost coincide. That is, the SHB processing, though simulating a "noisy" listening situation, sufficiently suppresses the noise to approximate listening to the original targets in quiet. The significance of that effect was confirmed by a three-factor ANOVA showing a highly significant main effect of processing mode $[F(1,13)=229.7,\ p<0.001]$, and a processing mode × sound interaction $[F(9,117)=20.94,\ p<0.001]$. Only when the background noise level is high (upper panel in Fig. 9) and the target level is low, one can observe some noise "leaking" into the SHB condition, and the ratings to fall between those of the original sounds in quiet, and of the original sounds with noisy background.

These results imply that an evaluation of individual target sound sources in a background of noise or competing sources can be achieved by steering the beam toward the target sound source using SHB. The results are not dependent on whether listeners are asked to judge the loudness of the target sound or the entire sound event.

D. Annoyance scaling

The average annoyance data are depicted in Fig. 10 (target sounds rated) and Fig. 11 (entire sound rated) with the sound samples ordered in the same way as in Figs. 8 and 9. The lower plot shows the low noise condition and the upper the high noise condition. In the experimental condition in which the participants were asked to judge the annoyance of the target only (Fig. 10), and did not hear it (i.e. pressed the inaudible button, which occurred in 11.9% of all annoyance trials), a "−1" was recorded. To account for this qualitatively different response reflecting a lower, but indeterminate level of annoyance, the median of all responses was substituted for

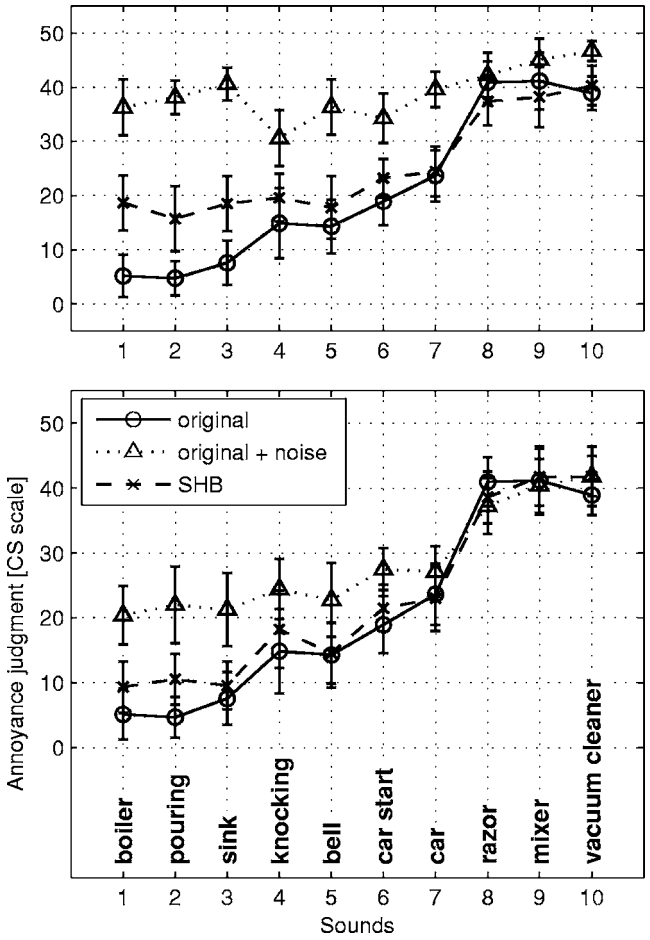

FIG. 11. Annoyance judgments of the ten test sounds in the low (lower panel) and high (upper panel) background noise condition. The target sounds are labeled along the abscissa and the error bars indicate 95%-confidence intervals. Listeners judged the entire sound event.

the mean in all graphical depictions when a judgment of "not heard" had occurred. It is evident in Fig. 10 that in three (respectively, two) cases the majority of the participants did not hear the target when presented in background noise of high (respectively, low) level. In one instance, the target (boiler sound in high-level noise; top panel of Fig. 10) was not even detected after SHB processing.

When the subjects were asked to focus on the annoyance of the target sound only (see Fig. 10), it appears that the different processing conditions do not affect the ratings very much: The three curves in Fig. 10 (upper and lower panel) are hardly distinguishable. Furthermore, the level of background noise does not seem to affect the annoyance ratings significantly: $F(1,13)=2.2$, $p=0.166$. This indicates that even though the sounds were contaminated by noise, the subjects were able to judge the annoyance of the target sound consistently by identifying the target's annoying features. Therefore, the advantage of using SHB cannot be shown in this case, because in contrast to the results of the loudness scaling there is hardly a background noise effect in the first place. A four-factor analysis of variance with the two attributes (loudness and annoyance) constituting an additional between-subjects factor revealed that the annoyance ratings of the target sounds were significantly different from the corresponding loudness judgments, as was evident in the significant interactions of the attribute judged with the processing mode [$F(1,26)=5.22$, $p=0.03$], and the three-way interaction with processing mode and sound [$F(9,234)=2.36$, $p=0.014$].[2]

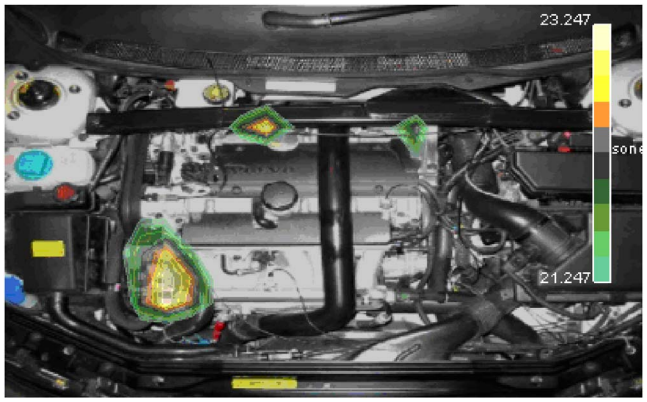

FIG. 12. (Color online) Loudness mapping of an engine compartment between 15 and 18 bark at 4000 rpm. See the text for details.

When the annoyance of the entire sound event is judged (see Fig. 11), the results are quite similar to those obtained for loudness. The effect of SHB processing is highly significant [$F(1,13)=158.43$, $p<0.001$], and the ratings obtained with SHB resemble those of the original sounds, with discrepancies emerging for the low-level sounds only. When loudness and annoyance are contrasted with respect to judgments of the entire sound, the interaction of the attributes with processing mode, sound level, and their combinations are no longer statistically significant (compared to judgments focusing on the targets, see the previous discussion), suggesting that the general pattern is quite similar for loudness and annoyance. This indicates that the annoyance percept is largely based on loudness if the subjects' attention is drawn to the entire sound mixture.

V. DISCUSSION

In an earlier investigation (Song, 2004), a comparison between traditional sound pressure maps and loudness maps derived from microphone array measurements was made and it was found that source identification in terms of psychoacoustic attributes improves the detectability of problematic sources. On the other hand, the mapping of some attributes cannot be derived due to the lack of metrics algorithms. Hence there is a need for auralizing the target sound identified as being devoid of background noise for further listening experiments.

Figure 12 shows the loudness map of an engine compartment of a passenger car with a five-cylinder, four-stroke engine. The engine was running at constant 4000 rpm without any external load applied. A 66-channel wheel array of 1 m diameter was mounted parallel to the car engine compartment at a distance of 0.75 m. In Fig. 12, it is obvious that the blank hole placed at the opposite side of the oil refill cap and the power steering pump at the lower left corner were the dominant sources in this operating condition. One might want to investigate attributes other than loudness, e.g., the

annoyance of those two sound sources, i.e., an attribute for which no agreed-upon objective metric exists. This could be done by having subjects judge the annoyance of the binaurally auralized sound of each target source at a time. This is a typical scenario for the use of source localization in practical applications in the automotive and consumer electronics industries.

Thus, the theoretical scaling of the SHB output derived in this paper and its experimental validation can be utilized for deriving a procedure to measure the auditory effects of individual sound sources. Since the method is based on steering the beam of a microphone array in three-dimensional space, no physical modifications of the sound field need to be made in contrast to typical dummy-head measurements. The details of the procedure proposed here will be discussed in the following.

A block diagram of the procedure for auralizing a target sound source binaurally is depicted in Fig. 13. This can easily be implemented together with classical beamforming applications in order to investigate problematic sources. Sound pressure signals are first measured at each microphone position on a rigid sphere, and converted to the frequency domain. Spherical harmonics beamforming is applied to steer a beam toward the target source (S_n) in each frequency band. A limited number of spherical-harmonics orders are used in SHB in order to avoid noise from the high-order spherical harmonics [see Eq. (23)].

The output of SHB, $P_{\mathrm{SHB}}(f)$, is scaled according to Eq. (19) to obtain the free-field pressure, $P_s(f)$, in the absence of the array with the assumption of a point source distribution on the source plane. The corresponding pressure time data, $P_s(t)$, are calculated by taking the inverse FFT of the scaled free-field pressure, $P_s(f)$. Finally, the binaural pressure signal can be acquired by convolving the free-field pressure with the HRTF in the source direction. Since HRTF databases are usually measured at discrete points on a full sphere, it is required to take either the nearest functions if the HRTFs are measured with a fine spatial resolution, or to interpolate between nearby points. The detailed procedure for interpolating HRTFs is described by Algazi *et al.* (2004) with respect to reproducing the measured sound field binaurally with the possibility of head tracking.

In the present paper, the analysis was restricted to the pressure contribution from a single direction. But, in many situations, such as in the professional audio industry, it is required to auralize distributed sources, i.e., the contribution from an area, and even the entire sound field as authentically as possible. An example of this kind of sound reproduction is the recording of sound fields in a car cabin while driving and reproducing it for head-tracked listening tests. In such situations, the measurements with a dummy head will have to be repeated many times in a well-controlled environment, which is very time-consuming, and may even be impossible due to lack of repeatability. Applying the procedure developed here to more than one direction enables the recording of full three-dimensional sound fields by one-shot array measurements and therefore allows listeners to turn their head while preserving the spatial auditory scene.

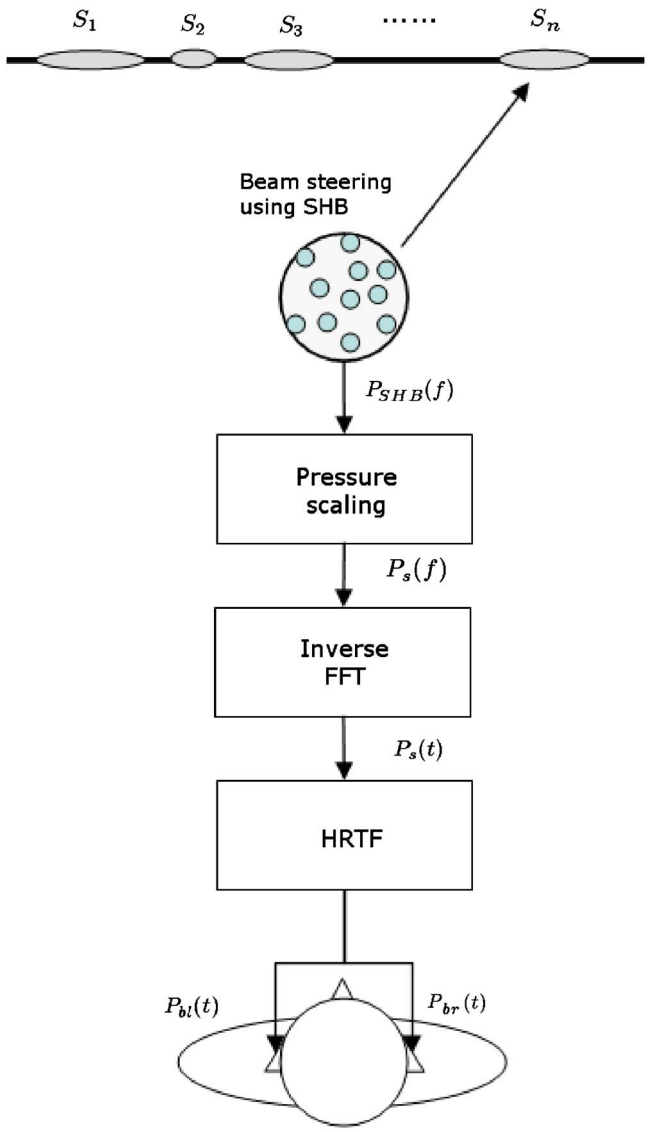

FIG. 13. (Color online) Binaural auralization of a desired sound source. Sound pressure signals are measured at each microphone position, and converted to the frequency domain. Spherical-harmonics beamforming (SHB) is applied to steer the beam toward a desired sound source and the output, $P_{\mathrm{SHB}}(f)$, is scaled to generate the free-field pressure, $P_s(f)$. The HRTF in the source direction is convolved with the pressure time signal, $P_s(t)$, obtained from the inverse FFT, and this results in binaural signals, $P_{bl}(t)$ and $P_{br}(t)$, at each ear.

VI. CONCLUSION

(1) A theoretical proposal was made for scaling the output of a spherical-harmonics beamformer, in order to estimate the free-field pressure at the listener's position in the absence of the microphone array. The comparison of measured and simulated responses (both monaural and binaural) to an array of loudspeakers showed that there is good agreement in the frequency range between 0.1 and 6.4 kHz. Notice that the simulated binaural responses were generated using an HRTF database, which was based on measurements using different instruments, physical structures, and a different anechoic chamber. Therefore, any differences between the two sets of responses contain the discrepancies between the earlier and current measurements.

(2) When the subjects judged target sounds partially masked by noise, their loudness was greatly reduced, but spherical harmonics beamforming managed to largely restore *loudness* to unmasked levels, except at low S/N ratios. By contrast, judgments of target *annoyance* were hardly affected by noise at all, suggesting that annoying sound features are extracted regardless of partial masking.
(3) When the subjects were asked to judge the entire sound events, SHB led to ratings close to those obtained in the original unmasked condition for both loudness and annoyance by suppressing background noise. The subjective judgments were largely explained by the percept of loudness: The loudness and annoyance data sets were highly correlated.
(4) The background noise level had significant effects by either producing partial masking (of targets) or contributing to the overall loudness (when the entire sound was judged). Judgments of target annoyance constituted an exception in that they were not affected by overall level.
(5) Implications of the study for sound-quality applications were sketched and a general procedure of deriving binaural signals using SHB was illustrated. The procedure can be used for evaluating the loudness and annoyance of individual sources in the presence of background noise.

ACKNOWLEDGMENTS

The experiments were carried out while the first two authors were at the "Sound Quality Research Unit" (SQRU) at Aalborg University. This unit was funded and partially staffed by Brüel & Kjær, Bang & Olufsen, and Delta Acoustics and Vibration. Additional financial support came from the Ministry for Science, Technology, and Development (VTU), and from the Danish Research Council for Technology and Production (FTP).

[1]Since the prerequisite normal-distribution assumption was met in the vast majority of experimental conditions—as verified by a Kolmogorov–Smirnov goodness-of-fit test—standard parametric analyses of variance were performed.

[2]In the ANOVAs, all not heard judgments were treated as values of −1. The pattern of statistical significances remained essentially the same when these problematic cases were excluded from the analysis.

Algazi, V. R., Duda, R. O., and Thompson, D. M. (**2004**). "Motion-tracked binaural sound," J. Audio Eng. Soc. **52**, 1142–1156.

Berglund, B., Berglund, U., and Lindvall, T. (**1975**). "Scaling loudness, noisiness, and annoyance of aircraft noise," J. Acoust. Soc. Am. **57**, 930–934.

Bovbjerg, B. P., Christensen, F., Minnaar, P., and Chen, X. (**2000**). "Measuring the head-related transfer functions of an artificial head with a high directional resolution," in Audio Engineering Society, 109th Convention, Los Angeles, preprint 5264.

Bowman, J. J., Senior, T. B. A., and Uslenghi, P. L. E. (**1987**). *Electromagnetic and Acoustic Scattering by Simple Shapes* (Hemisphere, New York).

Christensen, F., and Møller, H. (**2000**). "The design of VALDEMAR—An artificial head for binaural recording purposes," in Audio Engineering Society, 109th Convention, Los Angeles, preprint 5253.

Daniel, J., Nicol, R., and Moreau, S. (**2003**). "Further investigations of high order ambisonics and wavefield synthesis for holophonic sound imaging," in Audio Engineering Society, 114th Convention, Amsterdam, The Netherlands, preprint 5788.

Duraiswami, R., Zotkin, D. N., Li, Z., Grassi, E., Gumerov, N. A., and Davis, L. S. (**2005**). "High order spatial audio capture and its binaural head-tracked playback over head-phones with HRTF cues," in Audio Engineering Society, 119th Convention, New York, preprint 6540.

Ellermeier, W., Westphal, W., and Heidenfelder, M. (**1991**). "On the 'absoluteness' of category and magnitude scales of pain," Percept. Psychophys. **49**, 159–166.

Ellermeier, W., Zeitler, A., and Fastl, H. (**2004a**). "Impact of source identifiability on perceived loudness," in ICA2004, 18th International Congress on Acoustics, Kyoto, Japan, pp. 1491–1494.

Ellermeier, W., Zeitler, A., and Fastl, H. (**2004b**). "Predicting annoyance judgments from psychoacoustic metrics: Identifiable versus neutralized sounds," in Internoise, Prague, Czech Republic, preprint 267.

Gescheider, G. A. (**1997**). *Psychophysics: The Fundamentals* (Erlbaum, London, NJ).

Hald, J. (**2005**). "An integrated NAH/beamforming solution for efficient broad-band noise source location," in SAE Noise and Vibration Conference and Exhibition, Grand Traverse, MI, preprint 2537.

Hald, J., Mørkholt, J., and Gomes, J. (**2007**). "Efficient interior NSI based on various beamforming methods for overview and conformal mapping using SONAH holography for details on selected panels," in SAE Noise and Vibration Conference and Exhibition, St. Charles, IL, preprint 2276.

Hellman, R. P. (**1982**). "Loudness, annoyance, and noisiness produced by single-tone-noise complexes," J. Acoust. Soc. Am. **72**, 62–73.

ISO. (**1992**). "Audiometric test methods. 2. Sound field audiometry with pure tone and narrow-band test signals," ISO 8253-2, Geneva, Switzerland.

ISO. (**1998**). "Reference zero for the calibration of audiometric equipment. 1. Reference equivalent threshold sound pressure levels for pure tones and supra-aural earphones," ISO 389-1, Geneva, Switzerland.

Johnson, D. H., and Dudgeon, D. E. (**1993**). *Array Signal Processing: Concepts and Techniques* (Prentice Hall, London).

Kirkeby, O., Nelson, P. A., Hamada, H., and Orduna-Bustmante, F. (**1998**). "Fast deconvolution of multichannel systems using regularization," IEEE Trans. Speech Audio Process. **6**, 189–194.

Li, Z., and Duraiswami, R. (**2005**). "Hemispherical microphone arrays for sound capture and beamforming," in IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New York, pp. 106–109.

Marquis-Favre, C., Premat, E., and Aubre, D. (**2005**). "Noise and its effects—A review on qualitative aspects of sound. II. Noise and Annoyance," Acust. Acta Acust. **91**, 626–642.

Maynard, J. D., Williams, E. G., and Lee, Y. (**1985**). "Nearfield acoustic holography. I. Theory of generalized holography and the development of NAH," J. Acoust. Soc. Am. **78**, 1395–1413.

Meyer, J. (**2001**). "Beamforming for a circular microphone array mounted on spherically shaped objects," J. Acoust. Soc. Am. **109**, 185–193.

Meyer, J., and Agnello, T. (**2003**). "Spherical microphone array for spatial sound recording," in Audio Engineering Society, 115th Convention, New York, preprint 5975.

Minnaar, P. (**2001**). "Simulating an acoustical environment with binaural technology—Investigations of binaural recording and synthesis," Ph.D. thesis, Aalborg University, Aalborg, Denmark.

Møller, H. (**1992**). "Fundamentals of binaural technology," Appl. Acoust. **36**, 171–218.

Montgomery, D. C. (**2004**). *Design and Analysis of Experiments* (Wiley, New York).

Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing* (Academic, London).

Moreau, S., Daniel, J., and Bertet, S. (**2006**). "3D sound field recording with higher order ambisonics—Objective measurements and validation of a 4th order spherical microphone," in Audio Engineering Society, 120th Convention, Paris.

Nathak, S. S., Rao, M. D., and Derk, J. R. (**2007**). "Development and validation of an acoustic encapsulation to reduce diesel engine noise," in SAE Noise and Vibration Conference and Exhibition, St. Charles, IL, preprint 2375.

Park, M., and Rafaely, B. (**2005**). "Sound-field analysis by plane-wave decomposition using spherical microphone array," J. Acoust. Soc. Am. **118**, 3094–3103.

Petersen, S. O. (**2004**). "Localization of sound sources using 3D microphone array," Master's thesis, University of Southern Denmark, Odense, Denmark.

Rafaely, B. (**2004**). "Plane-wave decomposition of the sound field on a sphere by spherical convolution," J. Acoust. Soc. Am. **116**, 2149–2157.

Rafaely, B. (**2005a**). "Analysis and design of spherical microphone arrays," IEEE Trans. Speech Audio Process. **13**, 135–143.

Rafaely, B. (**2005b**). "Phase-mode versus delay-and-sum spherical microphone array processing," IEEE Signal Process. Lett. **12**, 713–716.
Song, W. (**2004**). "Sound quality metrics mapping using beamforming," in Internoise, Prague, Czech Republic, preprint 271.
Song, W., Ellermeier, W., and Minnaar, P. (**2006**). "Loudness estimation of simultaneous sources using beamforming," in *JSAE Annual Congress*, Yokohama, Japan, preprint 404.
Veronesi, W. A., and Maynard, J. D. (**1987**). "Nearfield acoustic holography (NAH). II. Holographic reconstruction algorithms and computer implementation," J. Acoust. Soc. Am. **81**, 1307–1322.
Versfeld, N. J., and Vos, J. (**1997**). "Annoyance caused by sounds of wheeled and tracked vehicles," J. Acoust. Soc. Am. **101**, 2677–2685.
Williams, E. G. (**1999**). *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, London).
Yi, S. (**2004**). "A study on the noise source identification considering the sound quality," Master's thesis, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea.
Zwicker, E., and Fastl, H. (**2006**). *Psychoacoustics: Facts and Models*, (Springer, Berlin), 3rd Ed.