Beamforming applied to psychoacoustics - sound source localization based on psychoacoustic attributes and efficient auralization of 3D sound fields

Song, Woo-keun

Publication date:
2008

Document Version
Publisher's PDF, also known as Version of record

Beamforming applied to psychoacoustics

sound source localization based on psychoacoustic attributes and efficient auralization of 3D sound fields

Ph.D. thesis by
Wookeun Song

Beamforming applied to psychoacoustics

sound source localization based on psychoacoustic attributes

and efficient auralization of 3D sound fields

Ph.D. thesis

Wookeun Song

November 2007

Sound Quality Research Unit (SQRU)
Section of Acoustics, Department of Electronic Systems
Aalborg University, Denmark

Beamforming applied to psychoacoustics
— sound source localization based on psychoacoustic attributes
and efficient auralization of 3D sound fields

Wookeun Song

Section of Acoustics
Aalborg University
Fredrik Bajers vej 7-B5
DK-9220 Aalborg

Revised March 25, 2008

Preface

This thesis is submitted to the Faculty of Engineering, Science and Medicine at Aalborg University in partial fulfilment of the requirements for the Ph.D. degree. The work has been carried out in the period between October 2003 and November 2007 at the Section of Acoustics, Department of Electronic Systems at Aalborg University. The study was funded by Brüel & Kjær Sound & Vibration Measurement A/S, as part of the "Centerkontrakt on Sound Quality" which establishes participation in and funding of the "Sound Quality Research Unit" (SQRU) at Aalborg University. The participating companies are Bang & Olufsen, Brüel & Kjær, and Delta Acoustics & Vibration. Further financial support comes from the Ministry for Science, Technology, and Development (VTU), and from the Danish Research Council for Technology and Production (FTP).

I would like to thank my supervisor Wolfgang Ellermeier for his valuable support to the design and analysis of listening experiments, comments on my writing, and finalizing my PhD. A special thanks also goes to Jørgen Hald for providing me a lot of good ideas related to microphone array techniques and encouraging my PhD work. I also want to thank Pauli Minnaar who guided me to design listening experiments and contributed to some of my conference presentations. I appreciate all the help Finn Kryger Nielsen provided for my PhD including financial support as well as encouraging for me to start my PhD. I would like to thanks to the people at the Section of Acoustics and the Department of Innovation at Brüel & Kjær for their valuable contribution to my work, and to the subjects who participated in my listening experiments.

Lately, I would like to thank my parents for their unlimited support of my life, and my family for their patience during my PhD. Special thanks go to my wife, Meesuk, for allowing me to start my PhD even though it requires the family's sacrifice and to concentrate on my work whenever it is necessary. This thesis is dedicated to my family.

Wookeun Song
Copenhagen, November 2007

Table of Contents

Summary

Noise source identification correlates the location of sources with their physical measures, such as sound pressure level (SPL). A number of noise source identification techniques have been suggested, and typically they visualize noise contributions as a function of location. Such visualization can be performed by measuring a sound field with a microphone array or by scanning an area of interest using an intensity probe. However, for an efficient noise source identification, the perceptual quality of each noise source must be considered. Determining the perceptual quality of individual sources or a partial sound field is becoming important in many areas of sound engineering, for example the identification of annoying components in a wind turbine or a vehicle exterior/interior.

This PhD study investigated how beamforming can be utilized to quantify auditory attributes of sources, and to auralize a sound field, or a partial sound field, for further psychoacoustical investigation. These goals were achieved by deriving measurement concepts necessary for this study, and by performing a series of listening experiments, in which different psychophysical methods were used. The findings of the present investigations were then related to a number of acoustical applications from loudness measurements of products having multiple noise components to on-road vehicle testing.

As a first step, a sound quality metrics mapping based on beamforming was proposed, and this method makes use of both monophonic and binaural loudness algorithms. Binaural loudness mapping, in which monophonic beamforming pressure estimation is convolved with head-related transfer functions (HRTFs) in the corresponding direction to generate binaural signals, could optimize the location of multiple sources relative to listener's head rotation in order to minimize overall loudness. In addition, sound quality metrics mapping proved to be an efficient way of localizing problematic sources by directly relating auditory attributes to sources thus supplementing the traditional sound pressure mapping.

In the first experiment, the loudness of simultaneous sources was investigated in a simple loudspeaker setup. It showed that listeners perceived narrow-band noises to be equally loud independently of their spatial distribution, i.e. no matter whether they were focused (0°) or distributed ($\pm 10^\circ$ or $\pm 30^\circ$), provided the directional loudness sensitivity of individual sources was compensated for. Moreover, a 6-dB loudness summation rule accounted for the subjective loudness perception of multiple sources for the stimuli employed in this experiment. It was observed that some subjects ignored the loudness contribution from a secondary sound completely while judging overall loudness.

Therefore, the threshold below which a secondary sound does not contribute to overall loudness any longer was investigated in the second experiment. In general, this "loudness threshold" for a secondary sound was much higher than expected, indicating that the secondary sound was clearly audible but did not contribute to perceived loudness. This proved that there is a considerable loudness dominance of the primary sound in multiple-sound conditions, much like in the cocktail-party effect.

In the third experiment, the subjective loudness and annoyance of a target sound in background noise was derived for sound signals synthesized binaurally either using a head-and-torso simulator (HATS) or spherical-harmonics beamforming (SHB). The outcome of the analysis indicated that SHB largely reinstated the loudness (or annoyance) of the target sounds to unmasked levels, even in noisy conditions, while the effect of background noise was obvious for the traditional binaural synthesis using an artificial head.

Finally, auditory attributes of multi-channel reproduced sounds based on the two auralization methods, i.e. HATS versus SHB, were compared in the fourth experiment. The two auralization methods produced quite similar results showing that the SHB auralization could reproduce spatial perception close to the HATS auralization. Notice that a SHB measurement simplifies binaural 3D sound recordings significantly compared to HATS measurements. Based on the findings, a general procedure for deriving binaural signals using SHB was proposed.

Resumé (summary in Danish)

Støjkilde-bestemmelse (Noise Source Identification) korrelerer/sammenholder positionen af lydkilder med de tilhørende fysiske mål såsom lydtryk-niveau. Der er udviklet forskellige metoder til støjkilde-bestemmelse, og disse visualiserer typisk støjbidraget som funktion af a positionen. En sådan visualisering kan foretages ved at måle lydfeltet med et mikrofon-array eller ved at skanne kildeområdet med en intensitets-probe. En effektiv metode til støjkilde-bestemmelse bør imidlertid involvere den "perceptive quality" af de enkelte lydkilder. Bestemmelsen af den "perceptive quality" af de enkelte kilder eller en del af lydfeltet er vigtigt indenfor mange områder, heriblandt identificering af generende komponenter i en vind-turbine eller i et køretøj.

Dette PhD-studium omhandler undersøgelser af hvorledes beamforming kan benyttes til at bestemme "audotory attributes" af kilder og til at "aualize" et lydfelt (indimellem kun dele af lydfeltet) til yderligere psykoakustiske undersøgelser. Disse mål blev nået ved at udlede en den grundlæggende teori til studiet, og ved at udføre en række lytte-tests. Resultaterne af disse undersøgelser blev derefter relateret til et antal akustiske anvendelser fra loudness målinger fra produkter med multiple komponenter til test med køretøj på vej.

"Sound quality metrics mapping" baseret på beamforming blev foreslået, og denne metode benytter både "monophonic" og binaural loudness algoritmer. Binaural loudness mapping, hvor monophonic beamforming tryk-estimering foldes med de hoved-relaterede overføringsfunktioner i den tilhørende retning for at generere binaurale signaler, kunne optimere positionen af multiple kilder i forhold til positionen af lytte-personens hoved for at minimere den samlede loudness. Derudover viste det sig, at "sound quality mapping metrics" var en effektiv måde til lokalisering af problematiske kilder ved direkte at relatere de "audiotory attributes" til kilder i modsætning til traditionel lydtryk mapping.

I det første eksperiment blev loudness fra simultane kilder undersøgt i et simpelt højtaler-setup. Det viste sig at lytte-personen opfattede smalbåndede kilder som værende ligelig i niveau uafhængigt af den rumlige fordeling, dvs. enten fokuseret ($0°$) eller fordelt ($\pm 10°$ eller $\pm 30°$), når der blev kompenseret for den retningsafhængige loudness følsomhed af de enkelte kilder. Derudover udgjorde den subjektive loudness perception af multiple kilder en 6dB loudness summering for den stimuli, der blev anvendt i eksperimentet.

Tærsklen, under hvilken en sekundær kilde ikke bidrager til den samlede loudness, blev derfor undersøgt i det næste eksperiment. Generelt var tærsklen for den sekundære kilde

meget højere end de forventede værdier tæt på høretærsklen, hvilket indikerer at den sekundære kilde var tydeligt "audible" men uden at bidrage til den opfattede loudness. Dette viste at der er en betydelig loudness dominans fra den primære lyd in multipel-lyd forhold, og dette kan forklares ved fænomenet bag cocktail-party effekten.

I det tredje eksperiment blev den subjektive loudness og annoyance fra en target-lyd i baggrundsstøj fundet for lydsignaler syntetiseret binauralt ved brug af enten head-and-torso simulator (HATS) eller sfærisk harmonisk beamforming (SHB). Resultatet af analysen indikerede at SHB oftest genindsatte loudness (eller annoyance) fra target-lydende til umaskerede niveauer, selv ved støjfyldte forhold, mens effekten fra baggrundsstøj var tydelig for den traditionelle binaurale syntese ved brug af et kunstigt hoved.

Reproduceret lyd fra multi-kanals auditory attributes baseret på to auralization metoder (HATS vs. SHB), blev sammenlignet i et fjerde eksperiment. De to auralization metoder gav rimelig ens resultater, hvilket viser at SHB auralization kan reproducere en rummelig perception, der er sammenlignelig med HATS auralization. Bemærk at SHB målinger simplificerer binaural 3D lyd optagelser markant sammenlignet med HATS målinger. Baseret på resultaterne blev der foreslået en generel procedure til at finde binaurale signaler ved brug af SHB.

Introduction and overview of the thesis

1 Introduction

In the field of automotive engineering and in the consumer electronics industry, it is often desired to identify the location of problematic noise sources as a means of "trouble shooting", and beamforming has been widely used for such purposes (Johnson and Dudgeon, 1993; Christensen and Hald, 2004, 2002). Beamforming is a signal processing technique based on measurements using an array of microphones placed at a medium to large distance from sound sources. Typically a delay-sum beamforming algorithm estimates how much of the pressure at the array position is incident from the focused direction by applying proper delays to each microphone position, and then summing the resulting signals (Johnson and Dudgeon, 1993). Thereby, it is possible to generate sound pressure maps by steering beams in such a manner that they cover the area of interest in a sound field. Moreover, Hald (2004) recently proposed a mathematical factor, which scales the output of the delay-sum beamformer in order to obtain an accurate sound power estimate for the sources. On the other hand, traditional physical measures, such as sound pressure and intensity, do not take into account how human subjects perceive sounds, and for that reason there is growing interest in predicting perceived psychoacoustic attributes from objective acoustical parameters (Zwicker and Fastl, 2006; Ellermeier *et al.*, 2004b).

In order to derive psychoacoustic attributes of sound fields of interest, e.g. a multichannel loudspeaker setup or a vehicle interior, it is frequently required to perform "blind" listening experiments to avoid biases, which might result from the visual appearance of the scene. Furthermore, different sound fields, e.g. a set of different cars, may have to be compared to each other directly in a listening experiment. Recently some studies measured binaural room impulse responses (BRIR) and convolved them with input signals according to listener's head movement using a head-tracking system (Horbach *et al.*, 1999; Mackensen *et al.*, 2000; Spikofski and Fruhmann, 2001). This type of auralization has been tested in a number of studies in which attributes of interior car noise or of multi-channel reproduced sound were estimated (Granier, 1996; Farina and Ugolotti, 1997; Christensen *et al.*, 2005; Bech *et al.*, 2005). However, the method requires measuring BRIRs at a large number of head rotation angles, and therefore is very time-consuming. Moreover, binaural signals may not be measurable at several head rotation angles, e.g. for on-road vehicle testing, where dummy head measurements are not repeatable and thus a new way of measurement overcoming these limitations has to be proposed.

Therefore, the present PhD study reports on a series of investigations applying beamforming techniques in conjunction with psychoacoustic experiments. The major goals of this study are twofold:

1. To identify and quantify problematic sound sources in a given sound field based on psychoacoustic attributes. Procedures deriving sound quality metrics maps on the basis of beamforming will be proposed. In addition, there should be guidelines to predict the loudness of simultaneous sound sources, and the interaction between multiple sound sources will have to be investigated, since traditional loudness models typically assume a single sound source in the free field or diffuse field.

2. To develop and verify a new binaural auralization method based on beamforming in order to utilize the advantages of beamforming such as background noise suppression and extracting partial sound fields. Two applications were chosen to validate the new method: (1) the auralization of target sounds in noise and (2) the auralization of the sound field generated by a multi-channel loudspeaker setup. As a first step, physical level differences between the new method and the traditional auralization based on dummy head measurements will be compared in a number of controlled conditions. Furthermore, the perceptual differences between the two technologies will be investigated in a series of listening experiments.

2 Organization of the thesis

[**Manuscript A**] Song, W. (2004). Sound quality metrics mapping using beamforming. Portions of this work have been presented at the Internoise, Prague, Czech Republic, August 22-25.

[**Manuscript B**] Song, W. & Ellermeier, W. (2006). Loudness assessment of simultaneous sounds using beamforming. Portions of this work have been presented at the Forum Acusticum Congress, Budapest, Hungary, 2005 August 29 - September 2 and the Annual Congress of the Society of Automotive Engineers of Japan (JSAE), Yokohama, Japan, 2006 May 24-26.

[**Manuscript C**] Song, W. & Ellermeier, W. (2007). Loudness threshold for a secondary sound source. To be submitted.

[**Manuscript D**] Song, W., Ellermeier, W., & Hald, J. (2008) Using beamforming and binaural synthesis for the psychoacoustical evaluation of target sources in noise., J. Acoust. Soc. Am. **123**, 910-924.

[**Manuscript E**] Song, W., Ellermeier, W., & Hald, J. (2007) Psychoacoustic evaluation of multi-channel reproduced sounds using binaural synthesis and beamforming. To be submitted.

The PhD thesis consists of five manuscripts, some of which are revised versions of conference papers, some being submitted or intended for publication in journals. Manuscript A

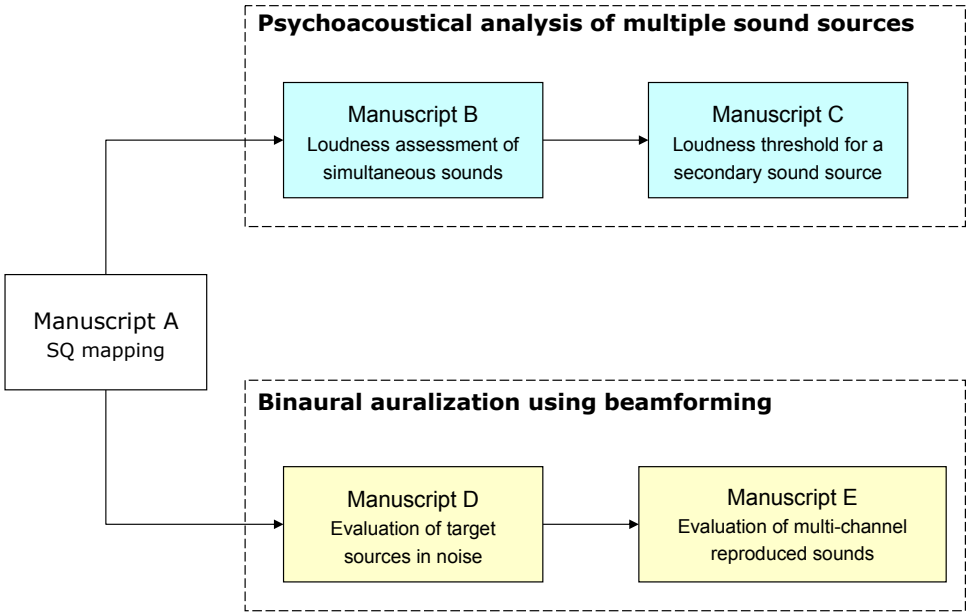


Figure 1: Schematic overview of the relationship between the five manuscripts.

describes the proposed procedure of sound quality metrics mapping, and demonstrates the problems, which have to be investigated in the following four manuscripts. Manuscript B and C focus on the perceived loudness of simultaneous sources in listening experiments, and propose a method of calculating the loudness of combined sources identified by beamforming. Manuscript D is dedicated to the psychoacoustical evaluation of target sources in noise while Manuscript E describes and validates the binaural auralization of multi-channel reproduced sound by means of beamforming and binaural synthesis. Fig. 1 shows the relations between the five manuscripts schematically. In the following chapter, an overview of each manuscript will be given.

Some of these papers involve significant contributions by Wolfgang Ellermeier, the academic supervisor of my PhD work, and by Jørgen Hald, a senior researcher at Brüel & Kjær. Wolfgang Ellermeier contributed mainly to the design and statistical analysis of the listening experiments, and aided in improving the write-up of the results. Jørgen Hald contributed to designing beamforming algorithms, and helped to refine the mathematical derivations related to those.

3 State of the art

In this section, the state of the art of employing acoustical measurement techniques for further psychoacoustical evaluation will be reviewed. In part 1, physical measurement techniques for psychoacoustical analysis using a single microphone, i.e. monophonic measurements, and a head-and-torso simulator (HATS), i.e. binaural measurements, are introduced. It is further shown how objective measures can be derived using such measurements. Part2 introduces typical beamforming techniques employed in the study, and

explains their main characteristics and applications.

3.1 Part 1. Physical measurement techniques for psychoacoustical analysis

Measurements for psychoacoustical analysis are closely related to estimating the sound transmission from a source in a free field through the outer ear to the eardrum, which is largely influenced by the size and shape of the torso, pinnae, and the ear canals. In general, the goal of such measurements is to reproduce the physical sound pressure levels at each of the ear drums as they were recorded in a sound field assuming that the compete auditory experience can be reproduced including timbral and spatial aspects. The sound transmission from a free field to a point in the ear canal of a human subject, namely the Head-Related Transfer Function (HRTF) (Shaw, 1974; Møller, 1992; Blauert, 2001), accounts for the filtering of a sound source due to the physical shape of human subjects, and thus depends on the direction of sound. Further physical sound transmissions, e.g. from the entrance of the ear canal to the eardrum, has been shown to be independent of the sound incident angle (Hammershøi and Møller, 1996).

When measuring HRTFs, sound pressures at each ear can be obtained at a number of positions in the outer ear, such as at the ear drum, at the blocked or open entrance of the ear canal (Møller, 1992), and Hammershøi and Møller (1996) revealed that the blocked entrance is the most suitable and stable point for measurements of HRTFs and for binaural recordings, due to the fact that sound at this point contains the complete spatial information, as well as the minimum amount of individual variation.

Møller *et al.* (1996) compared the source localization performance measured when asking subjects to listen a real sound field and binaural recordings of the same sound field, and found out that the localization performance was preserved with individual recordings compared to real life exposure. Furthermore, localization performance was found to be worse with binaural recordings made with a head-and-torso simulator when compared to individual recordings (Minnaar *et al.*, 2001a). Moreover, the results also demonstrated significant differences between currently available head-and-torso simulators. Despite these facts, binaural recordings may not be made individually for each subject in most of applications since they are very time-consuming and not practical. Therefore, in this PhD study, recordings and analyses based on HATS measurements will be studied.

3.1.1 Monophonic measurement

The auditory system of humans is binaural in that sound arrives at the two ears and the inputs to each ear are combined and processed by the system. In spite of this fact, most objective measures of perceived sound quality, such as loudness, sharpness, roughness, are developed based on monophonic measurements, i.e. measured by a single microphone. This may be due to the fact that monophonic measurements can easily be standardized since they measure free-field pressure and avoid measuring the filtering of a sound source caused by the presence of HATS, which is dependent on the sound incident direction. As

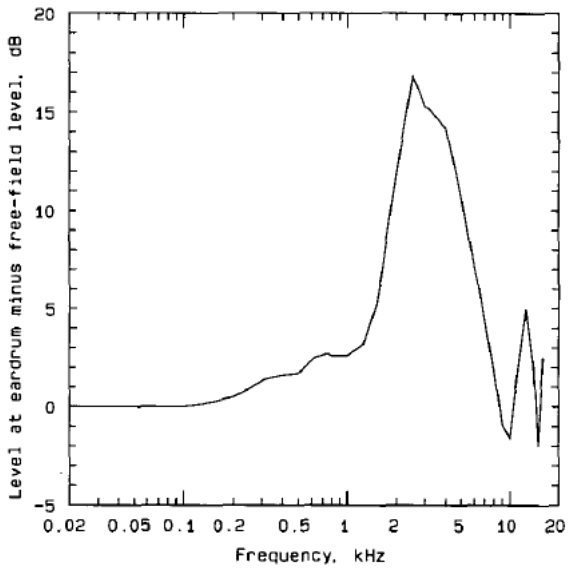


Figure 2: Transfer function from free-field pressure to sound pressure at the eardrum, taken from Moore *et al.* (1997)

reported in Minnaar *et al.* (2001a), currently available HATS are quite different and may results in different results when predicting sound quality metrics of the same sound source.

There are a number of examples, which try to derive objective metrics based on monophonic measurements. The most popular example is loudness, and it is developed in Moore *et al.* (1997); Zwicker and Fastl (2006). The models assume measured sound fields to be either free or diffuse. When they are free field, sounds are assumed to be presented in the frontal direction, meaning that subjects listen to sounds diotically (the same sound at the two ears). This allows to build a model based on two transfer functions, i.e. one for free field and the other for diffuse field, measured from sound pressure recorded in the absence of the listener at the center position of the listener's head to ear drum sound pressure. This function was measured in a number of studies (Killion *et al.*, 1987; Kuhn, 1979; Shaw, 1974), and Fig. 2 shows the function measured in the free field. If a 3.5-kHz pure tone is presented in the frontal direction, the sound pressure level of the sound will be increased by approximately 17 dB. On the other hand, the level of a 100-Hz tone will be almost unchanged after the filtering.

In practical applications, a lot of measurements should be done within a very limited time due to for example availability of facilities, and this makes it difficult to measure a sound field both monophonically (using a microphone) for objective measures and binaurally (using a HATS) for performing listening experiments. In such cases, binaural recordings are made and monophonic measurements usually are omitted. Then the inverse filter of the HATS response is applied to the signal at each ear, and objective measures are calculated on the filtered signals. The results for both ears may be averaged or the maximum value may be taken for the final result.

To perform such post-processing, we need to measure the free-field as well as the diffuse-field response of the HATS used for the measurement. For the free field, the response is

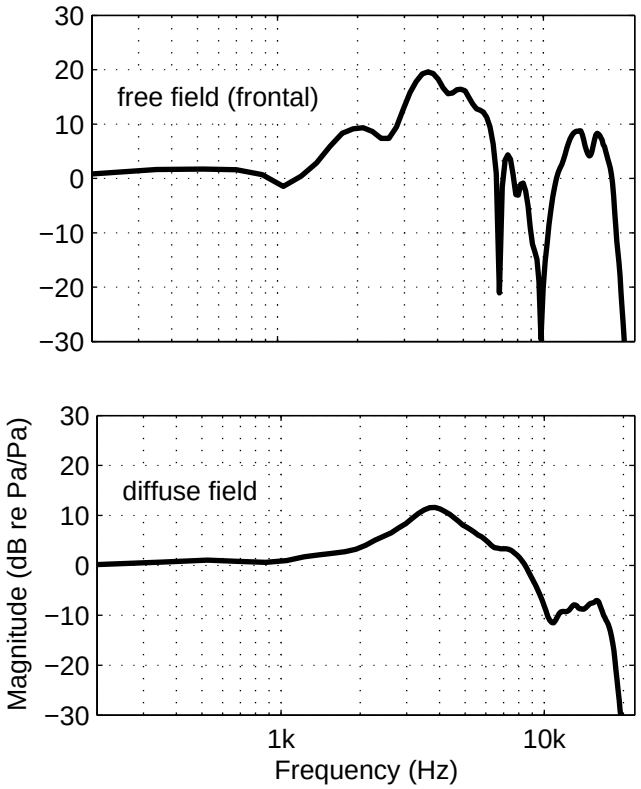


Figure 3: Transfer function of VALDEMAR from free-field pressure to sound pressure at the blocked entrance to the left ear canal. The upper plot shows the transfer function for sound incident at the frontal direction in the free field, and the lower plot illustrates that in the diffuse field.

the same as the HRTFs in the frontal direction. For the diffuse field, one may average HRTFs in all directions to obtain an approximated response or take the measurement in the diffuse field directly. An example of functions measured for the artificial head VALDEMAR (Christensen *et al.*, 2000) is shown in Fig. 3. It is evident that the HATS diffuse-field response is more smoothed than that of the free field, and the peak value of the transfer function is higher in the free field.

Even though monophonic measurements may be more straightforward for predicting the perceived quality of sounds, assuming diotic listening is still far from reality. In real-life situations, most of sound sources are perceived to be dichotic (different sounds at the two ears), and may be localized by listeners. Therefore, in many situations sound fields should be recorded binaurally and binaural objective measures are need to predict the perceived quantity more precisely.

3.1.2 Binaural measurement

Recording with a HATS is a most common way of measuring a sound field in many sound quality applications. The complete auditory experience is assumed to be reproduced

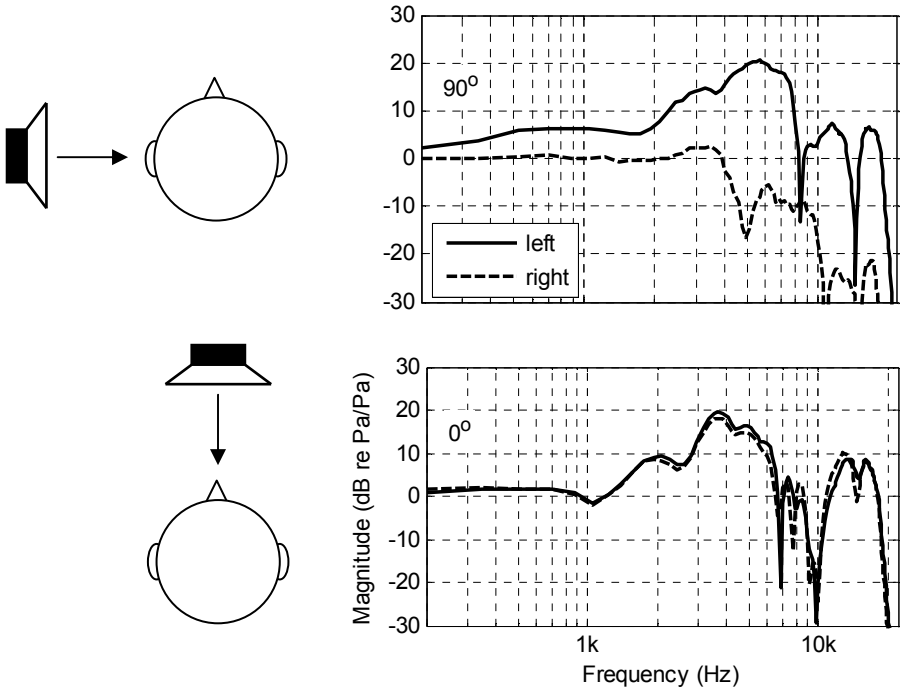


Figure 4: Head-related transfer functions measured at the blocked entrance to the ear canal of VALDEMAR for two incidence angles in the horizontal plane. Measured angles are indicated in the plots, and source positions are depicted in the left-hand side of each plot.

exactly if sounds are recorded at each ear and played back the same as they were. The 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):

$$h_{left}(t) = b(t) * c_{left}$$

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

where 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 Fourier transformation of HIR results in HRTF, and a thorough description of typical HRTF measurement is given by Møller *et al.* (1995). The binaural signals can then be obtained by convolving a "dry" source signal with the binaural impulse response functions h.

The sound pressure level at each ear is very much affected by the incident direction of the sound, and this is the reason why the frontal incidence of the sound in the free field or

the diffuse field assumption is not enough to describe the sound fields of interest. Fig. 4 shows the HRTFs of VALDEMAR (Christensen *et al.*, 2000) used in this study. The sound pressure at each ear was determined by measuring it at the blocked entrance of the ear canal, and the results are from two sound incident angles (0° and 90°) in the horizontal plane. When the source is presented in the frontal direction, the sound signal arrive at the ears with the almost same delay, i.e. no interaural time difference, and the level at each ear is almost equal, i.e. no interaural level difference. On the other hand, for the 90° incidence, the sound signals arrive at the left ear earlier than the right ear, and interaural level differences of up to approximately 30 dB are observed. In this case, assuming diotic listening may cause significant errors in predicting objective metrics.

Most objective metrics, e.g. sharpness, roughness, and fluctuation strength, are based on loudness estimates of the sounds in question. For this reason, understanding binaural loudness models may lead us to find out how binaural metrics can be obtained from binaural recordings. Recently, Moore *et al.* (1997) suggested a loudness model for steady sounds and a perfect loudness summation by simply summing the monaural loudness values at each ear. This way of calculating binaural loudness is standardized in ANSI S3.4 (2005).

Perfect loudness summation is a concept suggested by experiments in which the level at each ear is controlled independently, and thus does not account for the effect of the binaural impulse response function, i.e. h in Eq. 1, it thus may be said to be ecologically invalid. Therefore, Sivonen and Ellermeier (2006) modeled binaural loudness based on more realistic stimulation, which employed a single loudspeaker excitation in an anechoic condition or in reverberant environment (Sivonen, 2007). Directional loudness was measured for a number of sound incident angles and frequency bands. As a result of these investigations, they suggested a "3-dB binaural loudness summation rule", i.e. $\beta = 3$ in Eq. 2:

$$L_{dio} = \beta log_2(2^{L_{left}/\beta} + 2^{L_{right}/\beta}) - \beta \tag{2}$$

where L_{dio} is the equivalent sound pressure level needed for diotic stimulation to match any binaural combination of left-ear (L_{left}) and right-ear (L_{right}) input levels. The equation has the same form as one earlier proposed by Robinson and Whittle (1960). Once the binaural loudness of a sound is calculated, other binaural metrics based on loudness may be derived.

Although binaural metrics may be calculated based on loudness summation rules, binaural recordings do not provide a possibility of separating sources and thereby calculating the objective metrics of individual sources. For example, if there are two sources located at 0°, i.e. the frontal direction, and 90°, see Fig. 4, the binaural recording of such a sound field will capture the sum of transformations from two sources, i.e. the upper and lower graphs in Fig. 4, to the listener's ears. However, the goal of most measurement applications is to determine objective measures of sources rather than of the entire sound field. Therefore, in this PhD study a method of determining source metrics rather than the metrics of the entire sound field is proposed based on beamforming techniques.

3.2 Part 2. Beamforming

3.2.1 Review of beamforming techniques

Propagating sounds carry much information concerning the sound sources that generate them. By measuring a sound field with more than one acoustic transducer, the nature of the source, i.e. its temporal as well as spatial characteristics, may be determined based on physics involved. If each source produced sound in a different frequency range, the simplest way of separating the sources would be to apply linear filtering. On the other hand, such ideal conditions rarely occur, and simultaneous sources contain similar frequency ranges in many applications. Thus, a spatial filtering technique is required to localize noise sources and to determine the contribution of each sound source. The most popular of these techniques is beamforming. Beamforming is a signal processing technique employing an array of transducers that controls the directivity of, or sensitivity to, a focused direction. When measuring a sound field, beamforming can increase the receiver sensitivity in the focused direction by decreasing the sensitivity in the direction of interference or noise.

Beamforming algorithms may be categorized into fixed and adaptive beamforming (Veen and Buckley, 1988). Typically, fixed beamformers have a fixed spatial directivity (not dependent on the acoustical environment), and focus on a wanted sound source, thereby reducing the influence of background noise. Examples of fixed beamformers are delay-and-sum beamforming (Johnson and Dudgeon, 1993; Christensen and Hald, 2004), weighted-sum beamforming (Gallaudet and de Moustier, 2000), superdirective beamforming (Kates, 1993), and frequency-invariant beamforming (Ward *et al.*, 1994). On the other hand, adaptive beamforming may change its directivity dependent on the acoustical enviornment in which the beamformer is located. Doclo and Moonen (2003) designed a fixed beamformer, which makes use of a FIR filter-and-sum beamformer structure, to achieve a broadband beamformer having a given arbitrary spatial directivity for a given arbitrary microphone array configuration.

An alternative approach to beamforming was recently suggested by Liu *et al.* (2000) to reduce the number of employed sensors and the physical dimensions of an array, and the method may localize multiple sources simultaneously. It is based on the fact that human beings can communicate effectively in the presence of background noise as well as concurrent speakers due to the properties of directional hearing (see Blauert, 2001, for a review). The study made use of two microphones and the Jeffress model (Jeffress, 1948), which is based on interaural cross-correlation, and showed that their broadband localization technique works well in complex auditory scenes containing four simultaneous talkers in an anechoic chamber. Furthermore, the algorithm was applied to extract the desired signals in noisy environments (Liu *et al.*, 2001). On the other hand, localization performance, e.g. spatial resolution, was not compared with traditional beamforming techniques, and thus it is hard to see the improvement in the new algorithm. The method also does not distinguish sources that are only differently located in elevation since the model does not take into account the filtering effects of head, body, and pinnae.

Beamforming arrays may be divided into planar arrays and spherical arrays. Planar microphone arrays are the most widely used beamforming array type in many applications, and easy to build by placing each microphone physically on a plane. Typical shapes of

Figure 5: A spherical array with 50 microphones and 11 small cameras mounted flush on a rigid spherical surface with a diameter of 19.5 cm.

planar arrays are rectangular and circular, and microphones are positioned in a way that optimize the spatial characteristics of beamforming. However, these arrays cover only approximately $\pm 30^\circ$ from on-axis (see details in section 3.2.3), and thereby require a lot of measurements to cover all directions. For this reason, spherical microphone arrays became an alternative way of capturing a full 3D sound field in a single-shot measurement. A typical spherical array is shown in Fig. 5. Microphones are distributed evenly on a transparent sphere for delay-and-sum beamforming and typically on a rigid sphere for spherical harmonics beamforming (Rafaely, 2004; Park and Rafaely, 2005). Such a microphone array often consists of a number of cameras flush mounted on a sphere in order to overlay the captured sound field with the corresponding picture.

The spherical microphone array depicted in Fig. 5 covers up to 4 kHz with low sidelobe levels (see 3.2.2), and above 4 kHz sidelobe levels start to increase. To cover a wide frequency range and at the same time to be able to localize noise sources in the presence of reflections, e.g. interior noise measurements, it may be possible to combine a spherical microphone array with a back-screened array shown in Fig. 6 (Hald *et al.*, 2007). The array diameter equals to 0.45 m, and its back is covered by a 0.6 m diameter screen. Delay-and-sum beamforming with the back-screened array achieves better sidelobe suppression above 4-kHz up to approximately 14 kHz.

In this PhD study, delay-and-sum beamforming in planar arrays and spherical harmonics beamforming were used to measure sound fields of interest. The following section will provide a description of fundamental theories on these two beamforming techniques.

3.2.2 Fundamental formulations

Delay-sum-beamforming using a planar microphone array is a simple and robust method. In the left figure of Fig. 7, plane waves arrive at a planar array of M microphones at locations $\mathbf{r_m}$, which is a vector defined from the origin. This assumption is true for far-field measurement. The output of beamforming is calculated by applying delays dependent on

Figure 6: A planar array with a back screen. The array consists of 36 microphones.

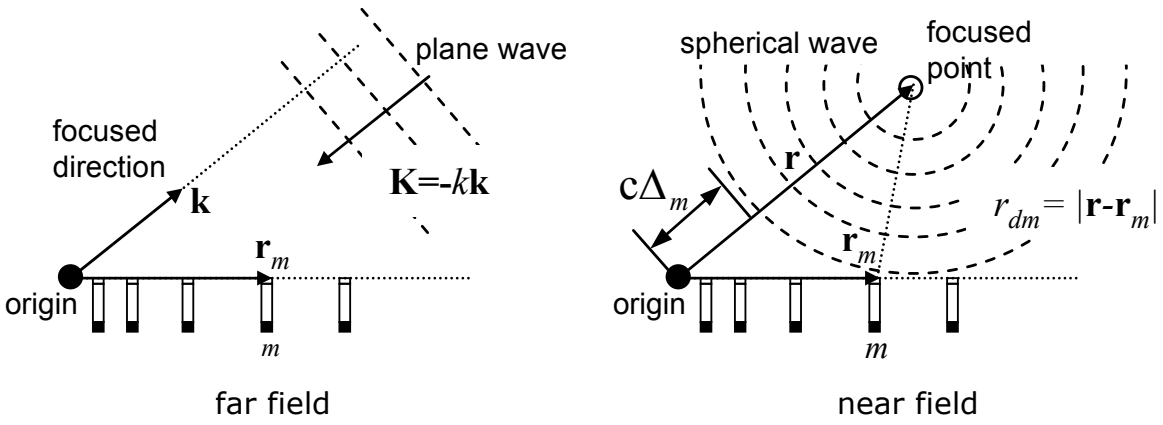


Figure 7: Incident waves to a microphone array in the far field and in the near field. In the far field, plane waves are incident from a focused direction to an array and microphone signals differ only in terms of phase. In the near field, spherical waves are emitted from a monopole source and reach the microphone array, and signals at each microphone are different both in their amplitude and phase, after Christensen and Hald (2003).

microphone positions to the signals recorded with the microphones. In this way, the acoustical waves from the focused direction are added coherently in the output, and pressure contributions from other directions are reduced. This may be formulated as:

$$b(\mathbf{k}, t) = \sum_{m=1}^{M} p_m(t - \Delta_m(\mathbf{k})) \tag{3}$$

where b is the beamformer output, p_m is the microphone signal, $\mathbf{k}$ is a unit vector in the focused direction, and Δ_m is an individual time delay on each microphone signal. In order to align signals associated with a plane wave in the focused direction, the delay in each microphone can be selected:

$$\Delta_m = \mathbf{k} \cdot \mathbf{r_m}/c \tag{4}$$

where c is the propagation speed of sound. In the frequency domain, a time delay is shown as a phase shift, and the counterpart of Eq. 3 is

$$B(\mathbf{k}, \omega) = \sum_{m=1}^{M} P_m(\omega) e^{-j\omega\Delta_m} = \sum_{m=1}^{M} P_m(\omega) e^{j\mathbf{K}\mathbf{r}_m} \tag{5}$$

Here, ω is the temporal angular frequency, $\mathbf{K} = -k\mathbf{k}$ is the wave number vector of a plane wave incident from the direction $\mathbf{k}$ in the focused direction (see the left figure of Fig. 7).

In the case of near field measurements, a distribution of monopole point sources on the focused plane may be assumed (see the right figure of Fig. 7), and Eq. 5 also holds in this case. Now, the delay Δ_m applied to each microphone should be changed as the following form for spherical incident waves:

$$\Delta_m = |\mathbf{r} - \mathbf{r_m}|/c = r_{dm}/c \tag{6}$$

As shown in this section, the fundamental formulation of delay-and-sum beamforming is rather simple, and that is why the calculation is computationally robust and easy to use. Even though Eq. 5 helps to add contributions from the focused direction coherently, there will be "leakage" from plane waves incident from other directions into the calculation of the main lobe direction $\mathbf{k}$. These are called "sidelobes", and may be clearly visible in an array directivity pattern. Fig. 8 shows the main lobe in the focused direction and sidelobes in other directions. For example, if there is a monopole source located in 75°, its sound pressure level will be reduced by approximately 16 dB in the main lobe direction.

In comparison with a planar array, a spherical array with microphones evenly distributed on a sphere may obtain directional characteristics independent of focused directions. This is particularly useful for measurements in an enclosed space, such as an interior car cabin, where sources are placed in 3D space and reflections are incident from almost all directions. The remaining part of this section is taken from II.B.1 in Manuscript D. For any function f(Ω) that is square integrable on the unit sphere, the following relationship holds (Rafaely, 2004).

$$F_{nm} = \oint f(\Omega) Y_n^{m*}(\Omega) d\Omega \tag{7}$$

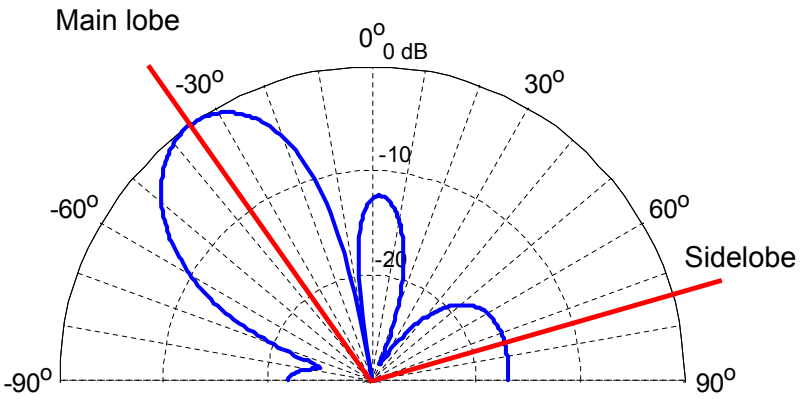


Figure 8: The array directivity pattern of a 66-channel wheel array at 1 kHz.

$$f(\Omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} F_{nm} Y_n^m(\Omega) \tag{8}$$

where "*" 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)

$$Y_n^m(\theta, \phi) = \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(cos\theta) \exp^{im\phi} \tag{9}$$

where n is the order, P_n^m are the associated Legendre polynomials, and $i = \sqrt{-1}$. Eq. 8 shows that any square integrable function can be decomposed into spherical-harmonics coefficients. Rafaely (2004) defined the relationship in Eq. 7 and 8 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 Eq. 7 and 8.

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)

$$p_\ell(\Omega_\ell, \Omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} R_n(ka) Y_n^{m*}(\Omega_\ell) Y_n^m(\Omega) \tag{10}$$

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

$$R_n = 4\pi i^n \left[j_n(ka) - \frac{j_n^{'}(ka)}{h_n^{(1)'}(ka)} h_n^{(1)}(ka) \right] \tag{11}$$

Here, j_n is the spherical Bessel function, $h_n^{(1)}$ the spherical Hankel function of the first kind, and $j_n^{'}$ and $h_n^{(1)'}$ 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. 10 and the spherical Fourier transform pair of $w(\Omega)$ we get:

$$p(\Omega, r = a) = \oint p_\ell(\Omega_\ell, \Omega) w(\Omega_\ell) d\Omega_\ell \tag{12}$$

$$= \sum_{n=0}^{\infty} \sum_{m=-n}^{n} R_n(ka) Y_n^m(\Omega) \oint w(\Omega_\ell) Y_n^{m*}(\Omega_\ell) d\Omega_\ell \tag{13}$$

$$= \sum_{n=0}^{\infty} \sum_{m=-n}^{n} W_{nm} R_n(ka) Y_n^m(\Omega) \tag{14}$$

By comparing Eq. 14 with the spherical Fourier transform pair of $p(\Omega, a)$, the spherical Fourier transform coefficients of $w(\Omega)$ can be obtained as

$$W_{nm} = \frac{P_{nm}}{R_n(ka)} \tag{15}$$

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

$$w(\Omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{P_{nm}}{R_n(ka)} Y_n^m(\Omega) \tag{16}$$

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

We now introduce a set of M microphones mounted at directions $\Omega_i, i = 1, ..., M$, on the hard sphere with radius a. The Fourier transform expression for P_{mn} 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:

$$P_{nm} \approx \tilde{P}_{nm} \equiv \sum_{i=1}^{M} c_i p(\Omega_i) Y_n^{m*}(\Omega_i) \tag{17}$$

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

$$H_{mn\mu\nu} \equiv \sum_{i=1}^{M} c_i Y_\nu^{\mu*}(\Omega_i) Y_n^m(\Omega_i) = \delta_{\nu n} \delta_{\mu m} \quad for \quad n \leq N, \nu \leq N \tag{18}$$

where N is the maximum order of spherical harmonics that can be integrated accurately with Eq. 17. 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. 17 in Eq. 16 and by limiting the spherical harmonics order to N:

$$b(\Omega) \equiv \sum_{i=1}^{M} \left[\sum_{\nu=0}^{N} \frac{1}{R_\nu(ka)} \sum_{\mu=-\nu}^{\nu} c_i Y_\nu^{\mu*}(\Omega_i) Y_\nu^{\mu}(\Omega) \right] p(\Omega_i) \tag{19}$$

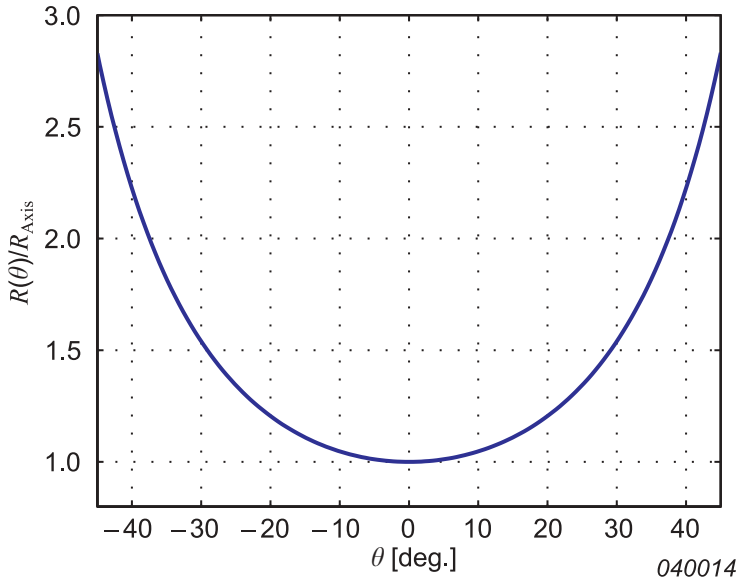


Figure 9: The ratio between on-axis and off-axis resolution as a function of the focused angle. The figure is from Christensen and Hald (2004).

3.2.3 Spatial resolution

Spatial resolution is one of the most crucial characteristics of beamforming, especially when the method is used for noise source identification, and shows its ability to distinguish waves incident from directions close to each other. The resolution is defined as the smallest angular separation that can distinguish two adjacent sources. The spatial resolution of planar arrays is defined (Christensen and Hald, 2004)

$$R(\theta) = \frac{a}{cos^3\theta}\frac{z}{D}\lambda \tag{20}$$

where R is spatial resolution, θ is the off-axis angle, z is the distance to the source, D is the array diameter, λ is the wave length of the incident wave, and a is the coefficient determined by the employed criterion. There are three criteria used in beamforming to determine the coefficient a. They are the *Rayleigh criterion* ($a = 1.22$) when the main lobe first falls into "null", a *3-dB criterion* ($a = 1.03$) when the amplitude of the main lobe becomes 3 dB lower than the peak level, and a *6-dB criterion* ($a = 1.41$) when the main lobe amplitude is 6 dB lower than the peak level. The 3 dB criterion is used when two sources to be separated are incoherent, and the 6 dB criterion when they are coherent.

The on-axis resolution can be calculated by setting $\theta = 0$ in Eq. 20 and the ratio between on-axis and off-axis resolution is depicted in Fig. 9. One may see that the spatial resolution in the focused direction of 30° is twice bigger than the one in on-axis. Therefore, the angular range of the focused direction in beamforming is set to ±30° in practice, and that was the reason why Manuscript B and C used a maximum sound incident angle of ±30°.

In spherical harmonics beamforming, the directivity of the sound field is calculated using Eq. 16 and the maximum order of spherical harmonics is limited to N. Assuming a plane

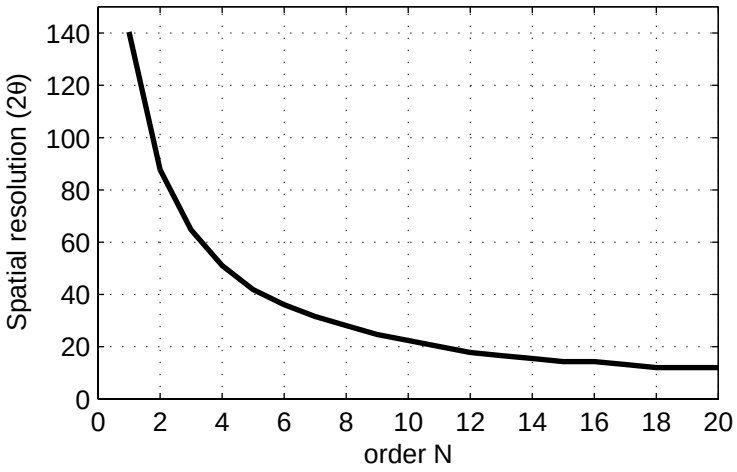


Figure 10: Spatial resolution of spherical harmonics beamforming as a function of spherical harmonics order. The 6-dB criterion is used for the calculation.

wave with a unit amplitude arriving from Ω_ℓ, the directivity can be calculated as follows (Rafaely, 2004)

$$w(\Omega) = \sum_{n=0}^{N} \sum_{m=-n}^{n} Y_n^{m*}(\Omega_\ell) Y_n^m(\Omega) \tag{21}$$

$$= \sum_{n=0}^{N} \frac{2n+1}{4\pi} P_n(cos\Theta) \tag{22}$$

$$= \frac{N+1}{4\pi(cos\Theta - 1)} \left[P_{N+1}(cos\Theta) - P_N(cos\Theta)\right] \tag{23}$$

where Θ is the angle between the focused and the considered direction. Fig. 10 shows the spatial resolution of SHB as a function of spherical harmonics order and the 6-dB criterion is used for the calculation. The plot indicates that higher orders of spherical harmonics should be used to achieve better spatial resolution. To obtain a spatial resolution of 20°, the spherical harmonics up to the 11th order should be taken into account according to Fig. 10.

3.2.4 Pressure scaling

In the case of delay-sum-beamforming, the beamformer output (see Eq. 5) delays each microphone signal according to the focused point and sums it across all microphones. This means that the output should be normalized by the number of microphones in order to obtain the average pressure contribution from the array microphones, and the normalized output may be formulated as (Christensen and Hald, 2004)

$$B_N(\mathbf{k}, \omega) = \frac{1}{M} \sum_{m=1}^{M} P_m(\omega) e^{j\mathbf{K}\mathbf{r}_m} \tag{24}$$

Binaural auralization requires to measure precise free-field pressure at the center of the array, i.e. corresponding to the center of the head position without the head presence. Typical fixed beamforming including delay-sum-beamforming does not provide a functionality to cover an area on the source plane to perform the auralization of a partial sound field, and the beam width changes dependent on frequency.

Hald (2005) derived a mathematical expression for a factor to scale the beamformed maps as sound intensity in such a way that area integration provides good estimates of partial area sound power data. The suggested scaling factor is

$$\alpha \approx \frac{2.94}{\rho c}\left(\frac{D}{\lambda}\right)^2 \tag{25}$$

where ρ is the density of the medium. A similar approach may be applied to obtain a pressure scaling factor that gives good estimates of free-field pressure generated by a partial sound field by taking area integration. First, we calculate the sound power of a partial sound field by taking area integration over the desired area on the beamformed map. Then we set the sound power equal to the square of the unknown free-field pressure P_f multiplied by the area of a sphere with a radius equal to distance to source, and divided by the free-field acoustic wave impedance. Then we can obtain the following scaling factor for the estimation of free-field pressure:

$$|P_f|^2 = \frac{2.94}{2\pi}\left(\frac{D}{L\lambda}\right)^2 \int\int |B_N|^2 \, dS \tag{26}$$

The scaling factor was derived with the help of the author in Hald (2005). To validate the scaling factor, a simulation with a monopole source placed in on-axis was performed with a beamforming wheel array with 66 microphones having a diameter of 1 m. When taking the area integration, 10 dB dynamic range was applied to minimize the effect of sidelobes. Fig. 11 shows the difference between the estimated and the exact free-field pressure at the center of the array in dB. It may be seen that the error reduces by increasing the distance to source and frequency.

The output of spherical harmonics beamforming (see Eq. 19) does not provide the correct pressure amplitude of an incident wave at the center of the array, and the scaling factor was derived to compensate for the error in Song *et al.* (2008), i.e. Manuscript D. Considering the case of a monopole point source and focusing of the beamformer at the distance r_0 of the point source, the derived scaling factor is

$$\frac{4\pi e^{ikr_0}}{(N+1)^2 kr_0} \tag{27}$$

Applying the scaling factor results in obtaining the correct free-field pressure generated by acoustical waves in the focused direction. However, to calculate the pressure contribution of a partial as well as an entire 3D sound field, the integration of beams over 3D space needs to be performed. The detailed procedures are described in Manuscript E. In Manuscript E, it is proposed that the response error caused by different overlapping beams should be calculated, and the inverse function of that needs to be compensated for during the

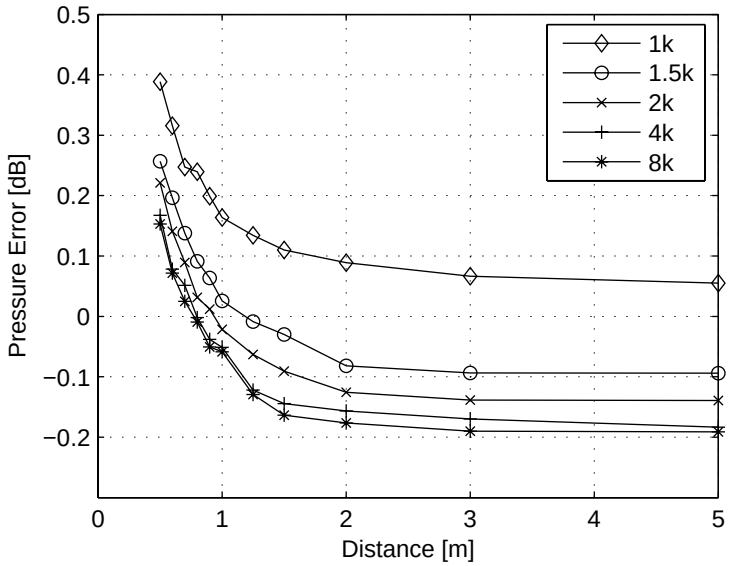


Figure 11: Difference in dB between estimated and true sound pressure. Estimated values are calculated using a delay-and-sum beamformer with the pressure scaling. The source is located in a monopole on the array axis.

integration of beams. The calculated response error as a function of frequency is shown in Fig. 12 when integrating 132 directions using a spherical array with 64 microphones having a radius of 14 cm. The error increases at low frequencies as a result of the greater beam overlap, and the curve has a staircase shape due to different orders of spherical harmonics being applied dependent on frequency.

3.2.5 Applications

There are two major applications of beamforming: using it as an auralization technique and using it as a noise source identification method. Many speech communication ap-

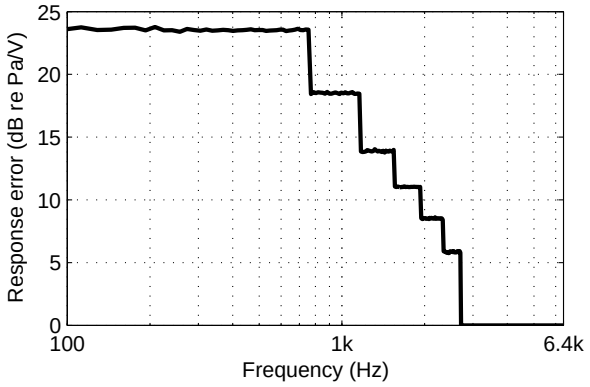


Figure 12: Response error caused by different beam widths.

Figure 13: An example of beamforming applications to a large object, i.e. crane in this case, from a distance. The picture was taken from Christensen and Hald (2004).

plications, such as hands-free mobile telephony (Ryan and Goubran, 2003), hearing aids (Kompis and Dillier, 2001; Kates, 1993), and speaker tracking and speech enhancement for a video conference (Valin *et al.*, 2006), are examples of beamforming used as an auralization method by suppressing background noise and reverberation, which cause a signal degradation and thereby may renders speech unintelligible.

In the area of noise source identification, beamforming is a relatively simple, yet robust, method compared to Near-field Acoustic Holography (NAH) (Maynard *et al.*, 1985), which requires more computational complexity as well as efforts of placing a microphone array close to the source plane in the presence of obstacles. For this reason, beamforming is widely used in the automotive industry (Marroquin *et al.*, 2007; Hald *et al.*, 2007) to identify noise sources at medium-to-high frequencies. Beamforming has an advantage of covering a large area by placing the array further away from the source at the cost of poor spatial resolution. Fig. 13 shows a measurement example of such a large object, in this case a crane (Christensen and Hald, 2004). A 42-channel microphone array with a diameter of 1 m was placed at 7 m from the crane hoisting at maximum load. It may be seen that a cover plate in the middle of the crane is the main noise source at around 2 kHz. Notice that to obtain a similar result using NAH a huge number of microphones is required, which may be almost impossible to place while operating the crane due to safety issues.

4 Synopsis of the thesis

The PhD study focused on applying beamforming techniques to psychoacoustics in order to discover new ways of investigating noise sources, which are very often encountered in different fields of industry. In the following, the main results of each investigation are reviewed, and are related each other to provide an overview of the thesis.

In Manuscript A, a procedure of generating sound quality metrics maps based on microphone array measurements is proposed and it is implemented both in a simulation program

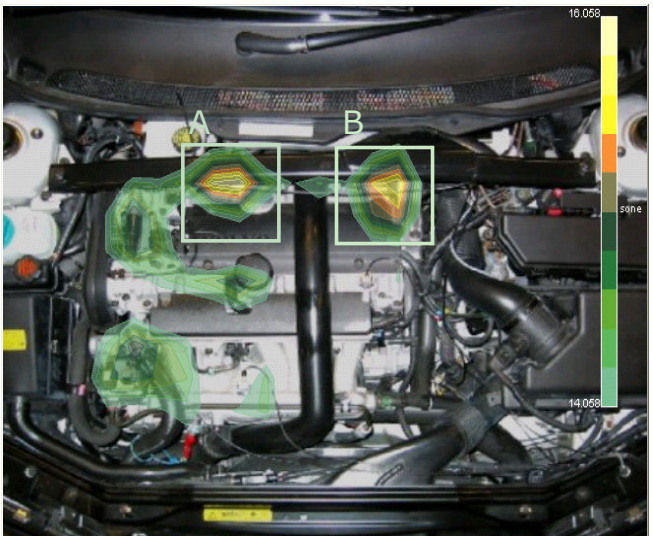


Figure 14: The loudness map of an engine compartment demonstrating that there are two noise sources having an almost identical size and peak loudness value.

as well as in a commercial measurement program. The procedure employs the standard loudness model (ISO 532, 1975) for diotic (same sound at the two ears) conditions and a 3-dB loudness summation rule (Sivonen and Ellermeier, 2006) for dichotic (different sounds at the two ears) conditions. Metrics other than loudness may be calculated based on the specific loudness spectra in the focused direction.

The procedure was validated through measurements on a simple loudspeaker setup in an anechoic chamber, and it was able to localize problematic sources more efficiently than traditional sound pressure mapping techniques. It was shown that loudness maps that take the listener's head rotation into account could be generated using binaural loudness mapping, and thereby provide the possibility of optimizing source locations in order to minimize the overall loudness of products. Furthermore, two methods, i.e. a new combined metric and an SPL dynamic range limit, were proposed to improve the mapping of sharpness, which might be corrupted by ghost images of beamforming.

The applicability of the proposed method was demonstrated by performing measurements on an engine compartment in a set of operating conditions with a 66-channel microphone wheel array. Sound pressure maps were directly compared with the corresponding loudness and sharpness maps, and the problematic sources were localized based on different metrics. The results suggest that sound quality metrics mapping provides a more efficient way of localizing problematic sources in sound fields.

Although sound quality metrics mapping may localize potential problematic sources, it lacks the ability of quantifying the psychoacoustic attribute contributed by an area covered by a noise source. Moreover, the contribution from multiple simultaneous sources needs to be determined to discover a mechanism of generating problematic noises. For example, Fig. 14 shows the loudness map of an engine compartment, and indicates clearly that the blank hole (source A) positioned in the opposite side of the oil refill cap and the engine mount (source B) are loud noise sources. However, the map does not show which of the two

sources is louder since they are almost identical in terms of size and peak loudness value. Furthermore, the loudness of the combined source A and B is hard to be determined in this example. Therefore, the loudness, i.e. the most fundamental metric, of simultaneous sources were investigated in Manuscript B and C.

Comprehensive loudness models for steady sound signals are well established (Zwicker and Fastl, 2006; Moore *et al.*, 1997), but they have mainly developed for diotic (the same sound at the two ears), or dichotic (different sounds at the two ears) headphone playback, and for sound presentation through a single loudspeaker placed in the frontal direction (Reynolds and Stevens, 1960; Scharf, 1969; Marks, 1978). Robinson and Whittle (1960) and more recently Sivonen and Ellermeier (2006) investigated loudness as a function of a sound incidence angle by presenting stimuli through a set of loudspeakers in the free field. These investigations, however, do not consider, what happens when two or more sources interact to produce an overall loudness percept. Thus, further investigations on loudness perception in sound fields with multiple sources are desirable to benefit from a combination of microphone array techniques and psychoacoustic measures, in order to find problematic sources in complex sound fields.

Therefore, a number of listening experiments, which are described in Manuscript B, were performed to investigate to which extent perceived loudness depends on the distribution of individual sound sources and how the loudness of individual components contributes to overall loudness. Three loudspeakers were positioned 1.5 m from the center of the listener's head, one straight ahead, two 10 degrees to the right and left in one condition, and two 30 degrees in the other. Listeners matched the loudness of either one or two simultaneous sounds (narrow-band noises with 1-kHz, and 3.15-kHz center frequencies) to a 2-kHz, either 45-dB or 60-dB SPL narrow-band noise placed in the frontal loudspeaker. The two simultaneous sounds were either originating from the central speaker, or from the two offset loudspeakers.

The results of the experiments revealed that the subjects perceived the noises to be equally loud independently of their distribution in space when the directional loudness sensitivity was equalized for individual sources. Furthermore, a 6-dB loudness summation rule was proposed to calculate the overall loudness of two simultaneous sounds. The 6-dB rule could predict the subjective data better than the traditional direct loudness summation of two simultaneous sounds. This result suggests that current loudness modeling will have to be extended to take the loudness summation of individual sources into account, and this may be achieved by combining beamforming techniques with the 6-dB rule suggested here.

Even though subjects were asked to judge the entire sound, not just a component of it during the listening experiments, there were a number of listeners who judged the overall loudness of simultaneous sounds based only on the loudest one. The average loudness matches of two simultaneous sounds from a subject are shown in Fig. 15 when two simultaneous noises, a 1-kHz and a 3.15-kHz noise, with the level of the latter being variable, were matched to the reference, a 2-kHz noise having either 45 or 60 dB SPL. Solid, dashed, and dash-dotted lines represent the data when the 3.15-kHz noise originated from the left, center, and right loudspeaker respectively, with the fixed 1-kHz noise being placed in the right, center, and left loudspeaker. The noise centered at 1 kHz had fixed sound pressure levels (L1, L2, L3) of 30, 35, or 40 dB SPL with the 45-dB reference and 40,

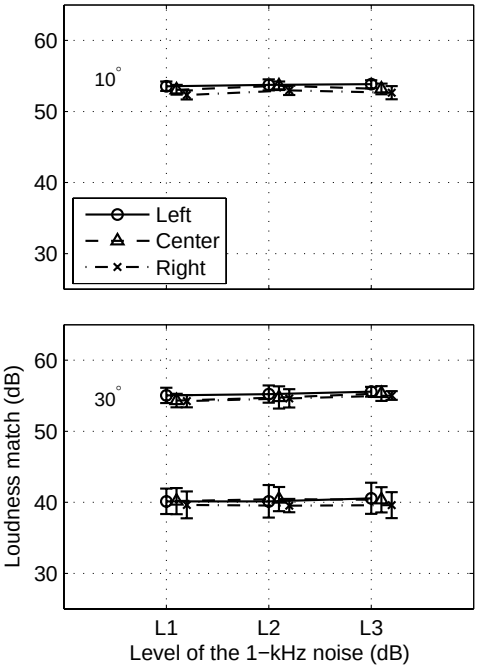


Figure 15: Loudness matches of the variable 3.15-kHz noise for different locations with the fixed 1-kHz noise in the opposite loudspeaker. Mean data with 95%-confidence intervals for a single listener. The lower curves in the lower panel indicate the results in the 45-dB reference condition, the others in the 60-dB reference condition. L1, L2, L3 were 30, 35, 40 dB SPL with the 45-dB reference and 40, 50, 55 dB SPL with the 60-dB reference.

50, or 55 dB with the 60-dB reference. The ordinate in Fig. 15 is the SPL of the variable 3.15-kHz noise. The details may be found in Manuscript B. In all conditions, the loudness matches in different levels of secondary sounds, i.e. L1, L2, and L3, were almost equal indicating that the secondary sound source did not influence overall loudness. This result reveals strongly that there must be a considerable loudness dominance of the primary sound in multiple-sound conditions, much like in the cocktail-party effect (Blauert, 2001).

Apart from the loudness dominance discussed previously, there is an important aspect to the role of a secondary source in relation to beamforming, namely sidelobes. Sidelobes in an array directivity pattern limit the dynamic range of mapping and may generate ghost images, i.e. false noise sources. A number of investigations have attempted to reduce the level of the maximum sidelobes, and either by applying weights, i.e. shadings, to measured signals at microphones depending on their position (Gallaudet and de Moustier, 2000) or by placing the microphones in an optimal way that minimizes the maximum sidelobes (Christensen and Hald, 2002). Despite these efforts, sidelobes are inevitable in the processing of beamforming measurements and for this reason sidelobes are clearly audible when auralizing a sound source in the focused direction using beamforming. Therefore, it is of important to investigate the threshold of perceived loudness, below which sidelobes, i.e. secondary sounds, do not contribute to overall loudness.

For this purpose, the loudness threshold for a secondary sound was measured in a series

of listening experiments and they are described in Manuscript C. A 1-kHz and 3.15-kHz narrow-band noise were used as stimuli, and they were auralized binaurally either in the frontal direction (0°) or in the two offset directions (±30°) using dummy-head HRTFs (Christensen *et al.*, 2000). The experiments were divided into a dual-frequency and a single-frequency condition. Two simultaneous noises were centered at different frequencies in the dual-frequency condition and at the same center frequency in the single-frequency condition. The influence of psychophysical method on the loudness threshold was also investigated by employing both an adaptive procedure (Jesteadt, 1980; Levitt, 1971) and free magnitude estimation (Stevens, 1975; Gescheider, 1997).

The results of a dual-frequency condition showed that the secondary sound contributes to a far lesser extent than expected given that the noise was clearly audible but did not contribute to the perceived loudness. On the other hand, the findings of the single-frequency condition agreed with the results from traditional Just-Noticeable Differences in Level (JNDL) studies (Zwicker and Fastl, 2006) in connection with level and frequency dependence. The two experimental procedures produced similar loudness thresholds for a secondary sound. The influence of spatial source separation (0° versus ±30°) was more obvious in the single-frequency condition than in the dual-frequency condition. The outcome of the study may be useful to design a microphone array that is more suitable for estimating the loudness of target sources in the presence of competing ones, and for calculating the overall loudness of simultaneous sounds.

In the first three manuscripts, problematic noise sources were localized based on their psychoacoustic attributes and the effects of combined sources on the overall percept were investigated. This was done by generating sound quality metrics maps of simple sound fields. However, some attributes, such as preference, cannot be derived in a similar fashion due to the lack of a metrics algorithm. Hence there is a need for auralizing a target sound devoid of background noise for further evaluation in listening experiments. Manuscript D provides the possibility of using beamforming for such a purpose.

In order to achieve a binaural auralization of sound fields, a beamformer should provide an almost equal directivity pattern in 3D space, and traditional planar microphone arrays are not suitable for such purpose due to their non-uniform directivity pattern. Recently, spherical microphone arrays have been investigated for the recording and analysis of a sound field (Rafaely, 2004, 2005; Meyer, 2001; Meyer and Agnello, 2003; Petersen, 2004) with the aim of overcoming the limitation of planar arrays. The major advantage of spherical microphone arrays where 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, independently of focused angle. Some studies (Duraiswami *et al.*, 2005; Li and Duraiswami, 2005) attempted to provide a theoretical derivation on how the free-field pressure obtained from spherical-harmonics beamforming (SHB) can be synthesized binaurally. However, the advantages of spherical-harmonics beamforming have not been demonstrated by means of psychoacoustic experiments.

The SHB technique was compared with traditional HRTF-based binaural synthesis (Møller, 1992; Hammershøi, 1995) for the auralization of target sound sources in the presence of background noise. In order to achieve this, the correct free-field pressure at the center of a spherical microphone array was estimated by deriving the theoretical pressure scaling of SHB. Six loudspeakers were positioned at 2.1 m away from the center of the setup in an

anechoic chamber. A setup of ten loudspeakers was simulated by flipping the position of four loudspeakers. The loudspeaker in the frontal direction was used as the target source through which the recorded sounds were synthesized and the rest of the loudspeakers served to create background noise. The procedure was verified physically by comparing simulated frequency response functions for each loudspeaker with directly measured ones both monaurally and binaurally. The results indicate that there is good agreement between simulated and measured responses in the frequency range of interest.

The proposed auralization method was evaluated subjectively by conducting a listening experiment. A set of 10 environmental and product sounds from a study by Ellermeier *et al.* (2004a) was 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. The influence of the background noise level was investigated by varying it in two steps (62, 72 dB SPL). Two independent groups of subjects (of N=14 each) evaluated either the loudness or the annoyance of the processed sounds during the experiment. The results indicate that SHB almost entirely restores the loudness (or annoyance) of the target sounds to unmasked levels, even when these are presented with background noise. Therefore the proposed auralization method may be a useful tool to psychoacoustically analyze target sources.

The psychoacoustic investigation of sound fields in a room, such as multi-channel audio setups and vehicle interior noise, ideally requires "blind" listening experiments in order not to bias subjects' responses due to expectations generated based on visual appearance. Furthermore, the independent variables selected will often have to be compared across experimental setups, e.g. when comparing the overall audio quality of multi-channel setups in a set of different cars. For this purpose, methods of measuring binaural room impulse responses (BRIRs), and convolving the input signals with them according to the listener's head movement measured by a head-tracking system was used in recent studies (Horbach *et al.*, 1999; Mackensen *et al.*, 2000; Spikofski and Fruhmann, 2001). The method has been used in multi-channel reproduced sound and in automotive applications to estimate the subjective effects of interior car sounds (Granier, 1996; Farina and Ugolotti, 1997; Christensen *et al.*, 2005; Bech *et al.*, 2005; Olive *et al.*, 2007).

Unfortunately, the traditional ways of recording sound fields binaurally are very time consuming since measurements have to be repeated for each head rotation angle. Christensen *et al.* (2005) introduced a new HATS with the possibility of head rotation through a motor controller (see Fig. 16), and demonstrated the method may reduce the measurement time significantly. Their method was also applied to a listening experiment to validate the system (Bech *et al.*, 2005). On the other hand, measuring BRIRs under nearly identical conditions with only head rotation varied may be unfeasible in some measurement scenarios, such as in on-road vehicle testing, and the measurement time still needs to be reduced further.

To this end, the binaural auralization of a 3D sound field using SHB was investigated and compared with the traditional one using a dummy head. This involves convolving individual beams with HRTFs and integrating them in 3D space. Six loudspeakers were positioned 2.1 m from the center of the setup in a listening room and their responses were measured with the three different methods using a microphone, a dummy head and a spherical microphone array. Simulated room impulse responses using SHB were compared

Figure 16: A head and torso simulator with the possibility of head rotation through a motor controller.

with directly measured ones both monaurally and binaurally and the results shows that there is good agreement in the frequency range between 0.1 to 6.4 kHz.

A listening experiment was performed to validate the procedure, i.e. to show that the SHB auralization produces similar results as does binaural synthesis based on measurements made with a head-and-torso simulator. Two musical excerpts, i.e. one pop and one classical, were processed for headphone presentation in two different ways: binaural synthesis using (1) dummy head measurements and (2) SHB. The influence of head rotation on subjective responses was investigated by having two head motility conditions, i.e. fixed and rotating, and six spatial processing modes, including phantom mono and stereo, were applied to obtain a wide range of spatial sensations. The outcome of the experiment indicates that the subjective scales of width, spaciousness and preference derived from SHB results were quite similar to the ones obtained from binaural synthesis using dummy head measurements, and in general results were not affected by head motility condition. This suggests that binaural auralization using SHB may be a useful tool to reproduce 3D sound fields based on a more efficient measurement, i.e. a single recording.

5 Discussion

5.1 Manuscript A: Sound quality metrics mapping using beamforming

The primary purpose of using beamforming is to localize and characterize noise sources in terms of sound pressure level (SPL) (Johnson and Dudgeon, 1993; Christensen and Hald, 2004, 2002). Recently, Washburn *et al.* (2005) applied beamforming techniques to determining the sound power level of large objects, e.g. earth-moving machinery. Beamforming simplifies such measurement by taking a measurement in the far field, and thereby avoiding the installation of microphones around large objects. Donavan (2007) identified noise sources in trucks that contribute most to the cruising passby noise levels. On the other hand, these methods do not provide a useful conversion from objective physical measures to perceptual quality of noise emitted from objects, e.g. how annoying noise sources are. In contrast, the current investigation provided the possibility of identifying noise sources based on perceptual quality rather than conventional measures such as SPL or sound power. Furthermore, sound quality metrics mapping could localize problematic sources more efficiently in a simple loudspeaker setup as well as on a personal-vehicle engine compartment.

The present study can be extended in a number of ways. One way could be utilizing other sound quality metrics, such as non-stationary loudness, roughness, and impulsiveness, in investigating the localization of problematic sources. Such metrics may relate to more specific noise problems in industries (Blommer *et al.*, 2005), and the limitation of beamforming, such as sidelobes and a limited frequency bandwidth, may need to be investigated for these metrics. Moreover, the findings of Manuscript B and C, i.e. a 6-dB loudness summation of multiple sources and a loudness threshold for a secondary sound, will have to be integrated in order to estimate loudness, potentially also other metrics, of sources in a sound field. Finally, the 6-dB loudness summation rule was developed by assuming the perfect sound pressure estimation of individual sources, but beamforming provides only an approximation. Therefore, it is of interest to investigate how much loudness estimation error may be caused by beamforming algorithms.

5.2 Manuscript B and C: Psychoacoustical analysis of multiple sound sources

Conventional loudness models (Zwicker and Fastl, 2006; Moore *et al.*, 1997; ISO 532, 1975) assume diotic (the same sound at the two ears) sound presentations either in the free or diffuse field, and binaural loudness models (Robinson and Whittle, 1960; Sivonen and Ellermeier, 2006) are developed based on dichotic (different sounds at the two ears) conditions when presenting stimuli through a single loudspeaker. When multiple sound sources are present in a sound field, loudness models based on diotic conditions assume a 3-dB loudness summation, i.e. power summation, for incoherent sources and a 6-dB loudness summation, i.e. pressure summation, for coherent sources since measurements are typically done using a microphone. In the case of binaural loudness models, the loudness summation

rule will be affected by the type of dummy head used for measurements since coherence between sources is dependent on the location of both ears relative to sound sources, and thus the gain of loudness summation ranges from 3 to 6 dB. In contrast, the current study revealed that the loudness summation of two incoherent narrow-band noises follows a 6-dB rule when measuring pressure contribution from individual sources in isolation, which may be achieved by performing beamforming measurements and by steering a beam toward a target source. The findings of the current investigation imply that the loudness of partial sound fields, which was not possible to measure using traditional microphone and dummy head measurements, may be estimated by combining beamforming measurements with the 6-dB loudness summation rule.

Most of Just-Noticeable Difference in Level (JNDL) experiments (Viemeister and Bacon, 1988; Hanna *et al.*, 1986; Jesteadt *et al.*, 1977; Zwicker and Fastl, 2006) employed pure tones that were presented diotically through a pair of headphones. Such studies determined the smallest audible level difference for otherwise identical stimuli. On the other hand, the loudness threshold for a secondary sound, which typically has different frequency content than a primary sound, cannot be explained by the results of the JNDL studies and needed to be determined in order to understand the interaction between sound sources in a multiple source environment. Jesteadt and Wier (1977); Stellmack *et al.* (2004) compared intensity discrimination in monaural and binaural listening, but binaural stimuli were presented either diotically or dichotically by simply introducing level difference between two ears. In contrast, the present study investigated the loudness threshold for a secondary sound by employing two simultaneous narrow-band noises and by convolving stimuli with the head-related transfer functions (HRTFs) in the corresponding direction. In this way, binaural signals with noticeable spatial images were generated, and therefore the effect of spatial configuration on the threshold could be investigated. Furthermore, this study revealed that a two-interval forced-choice adaptive procedure (Levitt, 1971), which was typically used in many JNDL studies (Viemeister and Bacon, 1988; Hanna *et al.*, 1986; Jesteadt *et al.*, 1977; Jesteadt and Wier, 1977; Stellmack *et al.*, 2004), failed to determine the loudness threshold for a secondary sound with the narrow-band noises used, whereas free (Stevens, 1975), or absolute (Gescheider, 1997) magnitude estimation was able to reliably obtain the thresholds.

The current investigation may be broadened by utilizing other than narrow-band noises, e.g. broad-band noise and speech, in investigating the loudness of multiple sources and the loudness threshold for a secondary sound. Such a real-life stimulus may draw listeners' attention toward a particular source in the presence of background noise, i.e. the phenomenon of the cocktail-party effect (Blauert, 2001), even more than what narrow-band noises did. The primary goal of such extended studies would be to investigate whether the 6-dB loudness summation rule and the thresholds obtained in this study remain unchanged for real-life stimuli.

The useful beamformer opening angle is in practice restricted to $\pm 30^\circ$. The maximum range of source separation angle investigated here was therefore $\pm 30^\circ$, at which the resolution becomes more than 50% greater than the on-axis resolution in beamforming processing using a planar microphone array (Christensen and Hald, 2004). On the other hand, spherical-harmonics beamforming (SHB) using a spherical microphone array (Rafaely, 2004, 2005; Meyer, 2001; Meyer and Agnello, 2003; Petersen, 2004) allows almost equal spa-

tial resolution independent of direction. For this reason, the same experimental paradigms may be investigated in a loudspeaker setup with wider angular separation between speakers. In such a setup, it may be possible to obtain larger directional loudness sensitivities (Sivonen, 2006), and thereby the role of directional loudness for individual sources may become more obvious.

Finally, the same experiments may be performed in a normal listening room to investigate the effect of sound field, e.g. including reflections in a room, on the loudness perception of simultaneous sounds. Such experiments will enable to apply the findings of the present study to more ecologically valid sound exposure since most sounds in our environment, e.g. in a room or in a car, include reverberation.

5.3 Manuscript D and E: Binaural auralization using beamforming

Traditional methods of binaural auralization, e.g. using a dummy head, have widely been used in a number of applications (Granier, 1996; Farina and Ugolotti, 1997; Christensen *et al.*, 2005; Bech *et al.*, 2005; Olive *et al.*, 2007). However, the method is not able to separate a target source from background noise or a partial sound field from the rest. In contrast, a binaural auralization based on spherical-harmonics beamforming (SHB) suggested here can perform spatial filtering of a given sound field, and thereby isolate a partial sound field of investigation. Moreover, recent studies on SHB (Rafaely, 2004, 2005; Meyer, 2001; Meyer and Agnello, 2003; Petersen, 2004) focused mainly on improving the signal-to-noise ratio of recordings, and still the output of beamformers is not scaled properly meaning that that it may not be used to generate stimuli for listening experiments. A novel scaling procedure proposed here enabled to obtain scaled stimuli in a focused direction, i.e. a target source, as well as of an entire sound field by removing the effect of different beam widths as a function of frequency.

Horbach *et al.* (1999); Mackensen *et al.* (2000); Spikofski and Fruhmann (2001) investigated the method of measuring binaural room impulse responses (BRIRs), and convolving the input signals with them according to the listener's head movement measured by a head-tracking system. The methods, however, require measuring BRIRs at different head rotation angles, and therefore is a very time-consuming process. A new method of measuring BRIRs as a function of head rotation angle was suggested in this study, and crucial parts of this method are convolving pressure contribution calculated by beamforming with the head-related transfer functions (HRTFs) in the corresponding direction, and subsequently integrating contributions from each direction in 3D space. This allowed to measure entire BRIRs in a sound field by a single recording.

A number of potential extensions may strengthen the findings of the present investigation. First of all, microphone array measurements may be performed in a real vehicle on a number of operating conditions, e.g. on-road vehicle testing, which was given as a example where operating conditions are not repeatable due to the variation of wind noise, tire noise, and vehicle speed. These measurements may validate the suggested procedure in direct recordings instead of transfer function measurements, e.g. BRIR measurements. Moreover, vehicle interiors are much smaller than the room utilized in this study and therefore

a number of consideration should be taken. These considerations include designing a smaller spherical array that has the similar size of an average human head to devoid the effect of array presence in a vehicle interior, and investigating the effect of obstacles, such as a head rest, closely mounted to the array on the SHB processing. In addition, the auditory attributes studied here were elicited by a multichannel loudspeaker array or simply assumed based on a literature survey, however vehicle interior noise problem may require a different set of attributes that are more relevant to vehicles. These attributes may be elicited by perceptual structure analysis (PSA) (Choisel and Wickelmaier, 2006) or by repertory grid technique (RGT) (Berg and Rumsey, 2006).

Recently, Independent Component Analysis (ICA) (Hyvarinen and Oja, 2000) has been suggested to separate target sounds from a set of mixed signals, without the aid of information (or with very little information) about the nature of the signals, namely blind source separation. ICA finds a linear representation of non-Gaussian data so that the components are statistically independent, or as independent as possible, and may be combined with beamforming (Saruwatari *et al.*, 2003). In general, the technique requires a smaller number of microphones, and may separate sources in the same direction (Ando *et al.*, 2005) in contrast to beamforming, which is based on spatial filtering. Therefore, it may be possible to perform a series of listening experiments to demonstrate the advantages of each source separation technique in terms of its psychoacoustical validity.

References

Ando, A., Iwaki, M., Ono, K., and Kurozumi, K. (**2005**). "Separation of sound sources propagated in the same direction", IEICE Transactions on Fundamentals of Electronics Communications and Computer Sciences **E88A**, 1665–1672.

ANSI S3.4 (**2005**). "Procedure for the Computation of Loudness of Steady Sounds", American National Standards Institute, New York, USA .

Bech, S., Gulbol, M.-A., Martin, G., Ghani, J., and Ellermeier, W. (**2005**). "A listening test system for automotive audio Part 2: Initial verification", in *Audio Engineering Society, 118th Convention*, preprint 6359 (Barcelona, Spain).

Berg, J. and Rumsey, F. (**2006**). "Identification of quality attributes of spatial audio by repertory grid technique", J. Audio Eng. Soc. **54**, 365–379.

Blauert, J. (**2001**). *Spatial Hearing - The Psychophysics of Human Sound Localization* (The MIT Press, London, England).

Blommer, M., Eden, A., and Amman, S. (**2005**). "Sound quality metric development and application for impulsive engine noise", in *SAE Noise and Vibration Conference and Exhibition*, preprint 2482 (Grand Traverse, MI, USA).

Choisel, S. and Wickelmaier, F. (**2006**). "Extraction of auditory features and elicitation of attributes for the assessment of multichannel reproduced sound", J. Audio Eng. Soc. **54**, 815–826.

Christensen, F., Jensen, C. B., and Møller, H. (**2000**). "The design of VALDEMAR - an artificial head for binaural recording purposes", in *Audio Engineering Society, 109th Convention*, preprint 5253 (Los Angeles, CA, USA).

Christensen, F., Martin, G., Minnaar, P., Song, W., Pedersen, B., and Lydolf, M. (**2005**). "A listening test system for automotive audio Part 1: System Description", in *Audio Engineering Society, 118th Convention*, preprint 6358 (Barcelona, Spain).

Christensen, J. J. and Hald, J. (**2002**). "A class of optimal broadband phased array geometries designed for easy construction", in *JSAE Annual Congress*, preprint 5335 (Yokohama, Japan).

Christensen, J. J. and Hald, J. (**2003**). "Improvements of cross spectral beamforming", in *Internoise*, preprint 356 (Seogwipo, Korea).

Christensen, J. J. and Hald, J. (**2004**). *Beamforming*, Technical Review number 1 (Brüel & Kjær, Nærum, Denmark).

Doclo, S. and Moonen, M. (**2003**). "Design of far-field and near-field broadband beamformers using eigenfilters", Signal Processing **83**, 2641–2673.

Donavan, P. R. (**2007**). "The influence of truck tire type and pavement on the emission of noise from trucks under highway operating conditions", in *SAE Noise and Vibration Conference and Exhibition*, preprint 2255 (St. Charles, Illinois, USA).

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 Headphones with HRTF Cues", in *Audio Engineering Society, 119th Convention*, preprint 6540 (New York, NY, USA).

Ellermeier, W., Zeitler, A., and Fastl, H. (**2004**a). "Impact of Source Identifiability on Perceived Loudness", in *ICA2004, 18th International Congress on Acoustics*, 1491–1494 (Kyoto, Japan).

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

Farina, A. and Ugolotti, E. (**1997**). "Subjective comparison of different car audio systems by the auralization technique", in *Audio Engineering Society, 103rd Convention*, preprint 4569 (New York, USA).

Gallaudet, T. C. and de Moustier, C. P. (**2000**). "On optimal shading for arrays of irregularly-spaced or noncoplanar elements", IEEE Journal of oceanic engineering **25**, 553–567.

Gescheider, G. A. (**1997**). *Psychophysics: The Fundamentals* (Lawrence Erlbaum Associates, New Jersey, USA).

Granier, E. (**1996**). "Comparing and Optimizing Audio Systems in Cars", in *Audio Engineering Society, 100th Convention*, preprint 4283 (Copenhagen, Denmark).

Hald, J. (**2004**). "Combined NAH and Beamforming using the same microphone array", Sound and Vibration **December issue**, 18–27.

Hald, J. (**2005**). "Estimation of partial area sound power data with beamforming", in *Internoise*, preprint 1511 (Rio de Janeiro, Brazil).

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*, preprint 334 (Grand Traverse, MI, USA).

Hammershøi, D. (**1995**). "Binaural technique – a method for true 3d sound reproduction", Ph.D. thesis, Aalborg University.

Hammershøi, D. and Møller, H. (**1996**). "Sound transmission to and within the human ear canal", J. Acoust. Soc. Am. **100**, 408–427.

Hanna, T., Vongierke, S. M., and Green, D. M. (**1986**). "Detection and intensity discrimination of a sinusoid", J. Acoust. Soc. Am. **80**, 1335–1340.

Horbach, U., Karamustafaoglu, A., Pellegrini, R., Mackensen, P., and Theile, G. (**1999**). "Design and Applications of a Data-based Auralization System for Surround Sound", in *Audio Engineering Society 106th Convention*, preprint 4976 (Munich, Germany).

Hyvarinen, A. and Oja, E. (**2000**). "Independent component analysis: algorithms and applications", Neural Networks **13**, 411–430.

ISO 532 (**1975**). "Acoustics – method for calculating loudness level", ISO, Geneva, Switzerland .

Jeffress, L. A. (**1948**). "A place theory of sound localization", J. Comp. Physiol. Psychol. **41**, 35–39.

Jesteadt, W. (**1980**). "An adaptive procedure for subjective judgments.", Perception & Psychophysics **28**, 85–88.

Jesteadt, W. and Wier, C. C. (**1977**). "Comparison of monaural and binaural discrimination of intensity and frequency", J. Acoust. Soc. Am. **61**, 1599–1603.

Jesteadt, W., Wier, C. C., and Green, D. M. (**1977**). "Intensity discrimination as a function of frequency and sensation level", J. Acoust. Soc. Am. **61**, 169–177.

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

Kates, J. M. (**1993**). "Superdirective arrays for hearing aids", J. Acoust. Soc. Am. **94**, 1930–1933.

Killion, M. C., Berger, E. H., and Nuss, R. A. (**1987**). "Diffuse field response of the ear", J. Acoust. Soc. Am. **81**, S75.

Kompis, M. and Dillier, N. (**2001**). "Performance of an adaptive beamforming noise reduction scheme for hearing aid applications. I. Prediction of the signal-to-noise-ratio improvement", J. Acoust. Soc. Am. **109**, 1123–1133.

Kuhn, G. F. (**1979**). "The pressure transformation from a diffuse sound field to the external ear and to the body and head surface", J. Acoust. Soc. Am. **65**, 991–1000.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics", J. Acoust. Soc. Am. **49**, 467–477.

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*, 106–109 (New York, NY, USA).

Liu, C., Wheeler, B. C., Jr., W. D. O., Bilger, R. C., Lansing, C. R., and Feng, A. S. (**2000**). "Localization of multiple sound sources with two microphones", J. Acoust. Soc. Am. **108**, 1888–1905.

Liu, C., Wheeler, B. C., Jr., W. D. O., Lansing, C. R., Bilger, R. C., Jones, D. L., and Feng, A. S. (**2001**). "A two-microphone dual delay-line approach for extraction of a speech sound in the presence of multiple interferers", J. Acoust. Soc. Am. **110**, 3218–3231.

Mackensen, P., Fruhmann, M., Thanner, M., Theile, G., Horbach, U., and Karamustafaoglu, A. (**2000**). "Head Tracker-Based Auralization Systems: Additional Consideration of Vertical Head Movements", in *Audio Engineering Society, 108th Convention*, preprint 5135 (Paris, France).

Marks, L. E. (**1978**). "Binaural summation of the loudness of pure tones", J. Acoust. Soc. Am. **64**, 107–113.

Marroquin, M., Frazer, T., and Jr., G. N. (**2007**). "In-vehicle panoramic noise source mapping", in *SAE Noise and Vibration Conference and Exhibition*, preprint 375 (Grand Traverse, MI, USA).

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*, preprint 5975 (New York, NY, USA).

Minnaar, P., Olesen, S. K., Christensen, F., and Møller, H. (**2001**a). "Localization with binaural recordings from artificial and human heads", J. Audio Eng. Soc. **49**, 323–336.

Minnaar, P., Olesen, S. K., Christensen, F., and Møller, H. (**2001**b). "The importance of head movements for binaural room synthesis", in *Proceedings of the 2001 International Conference on Auditory Display*, 21–25 (Espoo, Finland).

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

Møller, H., Sørensen, M. F., Hammershøi, D., and Jensen, C. B. (**1995**). "Head-related transfer functions of human subjects", J. Audio Eng. Soc. **43**, 300–321.

Møller, H., Sørensen, M. F., Jensen, C. B., and Hammershøi, D. (**1996**). "Binaural Technique: Do We Need Individual Recordings?", J. Audio Eng. Soc. **44**, 451–469.

Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). "A model for the prediction of thresholds, loudness, and partial loudness", J. Audio Eng. Soc. **45**, 224–240.

Olive, S., Welti, T., and Martens, W. L. (**2007**). "Listener loudspeaker preference ratings obtained in situ match those obtained via binaural room scanning measurement and playback system", in *Audio Engineering Society, 122nd Convention*, preprint 7034 (Vienna, Austria).

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

Perrett, S. and Noble, W. (**1997**). "The effect of head rotations on vertical plane sound localization", J. Acoust. Soc. Am. **102**, 2325–2332.

Petersen, S. O. (**2004**). "Localization of sound sources using 3D microphone array", Master's thesis, University of Southern 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. (**2005**). "Analysis and design of spherical microphone arrays", IEEE Transactions of Speech and Audio Processing **13**, 135–143.

Reynolds, G. S. and Stevens, S. S. (**1960**). "Binaural summation of loudness", J. Acoust. Soc. Am. **32**, 1337–1344.

Robinson, D. and Whittle, L. (**1960**). "The loudness of directional sound fields", Acustica **10**, 74–80.

Ryan, J. and Goubran, R. A. (**2003**). "Application of near-field optimum microphone arrays to hands-free mobile telephony", IEEE transactions on vehicular technology **52**, 390–400.

Saruwatari, H., Kurita, S., Takeda, K., Itakura, F., Nishikawa, T., and Shikano, K. (**2003**). "Blind source separation combining independent component analysis and beamforming", EURASIP Journal on Applied Signal Processing **2003**, 1135–1146.

Scharf, B. (**1969**). "Dichotic summation of loudness", J. Acoust. Soc. Am. **45**, 1193–1205.

Shaw, E. A. G. (**1974**). "Transformation of sound pressure level from the free field to the eardrum in the horizontal plane", J. Acoust. Soc. Am. **56**, 1848–1861.

Sivonen, V. (**2006**). "Directional loudness perception - the effect of sound incidence angle on loudness and the underlying binaural summation", Ph.D. thesis, Aalborg University.

Sivonen, V. P. (**2007**). "Directional loudness and binaural summation for wideband and reverberant sounds", J. Acoust. Soc. Am. **121**, 2852–2861.

Sivonen, V. P. and Ellermeier, W. (**2006**). "Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation", J. Acoust. Soc. Am. **119**, 2965–2980.

Song, W., Ellermeier, W., and Hald, J. (**2008**). "Using beamforming and binaural synthesis for the psychoacoustical evaluation of target sources in noise", J. Acoust. Soc. Am. **123**, 910–924.

Spikofski, G. and Fruhmann, M. (**2001**). "Optimisation of Binaural Room Scanning (BRS): Considering inter-individual HRTF-characteristics", in *19th Conference of the Audio Engineering Society*, 272–286 (Schloss Elmau, Germany).

Stellmack, M. A., Viemeister, N. F., and Byrne, A. J. (**2004**). "Monaural and interaural intensity discrimination: Level effects and the "binaural advantage"", J. Acoust. Soc. Am. **116**, 1149–1159.

Stevens, S. S. (**1975**). *Psychophysics: Introduction to its perceptual, neural and social prospects* (Wiley, New York, USA).

Thurlow, W. R. and Runge, P. S. (**1967**). "Effect of induced head movements on localization of direction of sounds", J. Acoust. Soc. Am. **42**, 480–&.

Valin, J. M., Michaud, F., and Rouat, J. (**2006**). "Robust 3D localization and tracking of sound sources using beamforming and particle filtering", in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing.*

Veen, B. V. and Buckley, K. (**1988**). "Beamforming: a versatile approach to spatial filtering", IEEE ASSP Magazine **5**, 4–24.

Viemeister, N. F. and Bacon, S. P. (**1988**). "Intensity discrimination, increment detection, and magnitude estimation for 1-kHz tones", J. Acoust. Soc. Am. **84**, 172–178.

Ward, D., Kennedy, R., and Williamson, R. (**1994**). "Design of frequency-invariant broadband far-field sensor arrays", Antennas and Propagation Society International Symposium **2**, 1274–1277.

Washburn, K. B., Frazer, T., and Kunio, J. (**2005**). "Correlating noise sources identified by beamforming with sound power measurements", in *SAE Noise and Vibration Conference and Exhibition*, preprint 2510 (Grand Traverse, MI, USA).

Williams, E. G. (**1999**). *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic Press, London, Great Britain).

Zwicker, E. and Fastl, H. (**2006**). *Psychoacoustics : Facts and Models* (Springer, Berlin, Germany).

Sound quality metrics mapping using beamforming *

Wookeun Song†

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

(Dated: March 2, 2008)

Sound quality metrics mapping is proposed as a methodology to identify sound sources in terms of their psychoacoustic attributes, such as loudness and sharpness. The theoretical description of the proposed method includes transient beamforming processing, convolution of HRTFs with the beamforming output, and applying either a diotic or binaural loudness model. The advantage of sound quality metrics mapping was demonstrated in simulations by deriving binaural loudness maps dependent on the listener's head rotation, and thereby it was possible to optimize the loudness contribution of a sound source in relation to the head rotation angle. Beamforming measurements were made in an anechoic chamber, in which multiple sound sources were simulated by a loudspeaker setup. The superiority of sound quality metrics mapping was demonstrated by comparing it with conventional sound pressure mapping. Practical measurements on an engine compartment could localize the loudest and sharpest sources in different RPM conditions, and illustrated that the location of major sources changes dependent on the metrics selected. Thus the proposed method may be useful to identify problematic sources in a more efficient manner than traditional pressure and intensity mapping.

I. INTRODUCTION

Conventional beamforming techniques calculate relative pressure contributions to the sound field at the array position and suppress the influence of background noise using a time alignment of the acoustic signals arriving at the array microphones (Johnson and Dudgeon, 1993). Beamforming sidelobes may be reduced by placing the microphones in the array irregularly (Hald and Christensen, 2002). These methods often perform well for the problem of identifying multiple sound sources emanating from a given test object. Correlations of relative sound pressure mappings with operational conditions, such as rpm, speed, or shaft angle, illustrate how noise sources interact when the given test object operates in specific conditions (Christensen and Hald, 2004).

Sound power is commonly used to identify problematic sources and thereby to reduce the noise level emitted by a given test object. But there are many situations, in which reducing the sound power of products by making costly design changes does not improve perceived sound quality. Practical noise problems are often related to specific psychoacoustic attributes, e.g. loudness, sharpness, or annoyance (Zwicker and Fastl, 2006). Therefore relative sound pressure maps sometimes lead to misinterpretations, which in turn result in irrelevant design changes. By contrast, sound quality metrics mapping is done based on pressure time data in the focused direction produced by beamforming. Thereby the mapping concentrates on more specific issues, and provides only relevant information, which may be used for the improvement of perceived sound quality.

In this paper, the advantage of sound quality metrics mapping will be demonstrated in simulations by deriving binaural loudness maps dependent on listener's head rotation and in loudspeaker measurements by comparing sound pressure maps with the corresponding loudness and sharpness maps. Methods of improving sharpness maps will also be illustrated with loudspeaker measurements. Furthermore, the application of sound quality metrics mapping will be outlined by generating loudness and sharpness maps of an engine compartment.

II. SOUND QUALITY METRICS MAPPING

Beamforming techniques may be categorized into fixed and adaptive beamforming (Doclo and Moonen, 2003), and they are widely applied in noise source identification and hearing aids (Hald, 2005a; Kompis and Dillier, 2001). Fixed beamformers provide a fixed spatial directivity pattern whereas adaptive beamformers are able to adapt to changing acoustic environments. In general, fixed beamformers require less computational power and are robust compared to adaptive beamformers. For this reason, delay-sum beamforming, i.e. the most popular fixed beamformer, is used for the current investigation.

In the delay-sum beamformer, a delay Δ_m and an amplitude weight w_m are applied to each microphone signal, then the resulting signals are summed. The beamforming output is shown in Eq. 1. The delay at each microphone position is calculated in the focused direction $\mathbf{k}$, and the amplitude weight, i.e. the array's shading, is selected to improve the beam's shape and to reduce sidelobe levels

*Portions of this work have been presented at the Internoise, Prague, Czech Republic, 2004 August 22-25

†Electronic address: wksong@bksv.com

(Johnson and Dudgeon, 1993).

$$b(\mathbf{k},t) = \sum_{m=1}^{M} w_m p_m(t - \Delta_m(\mathbf{k})) \quad (1)$$

where b is the beamforming output and p_m is the sound pressure at the microphone position m. Since the acquired signal at each microphone is sampled digitally, the resampling of time signals should be performed to apply individual delays. This is a very time consuming procedure and therefore the frequency domain version of Eq. 1 is often used and shown in Eq. 2.

$$B(\mathbf{k},\omega) = \sum_{m=1}^{M} w_m P_m(\omega) e^{-j\omega\Delta_m(\mathbf{k})} \quad (2)$$

where ω is the angular frequency, and B and P_m are the Fourier transformed counterpart of b and p_m.

The beamformer output B is the sound pressure contribution in the focused direction, and may be seen as free-field pressure at the center of the head with the head absent. Typically beamforming algorithms estimate how much of the pressure at the array position is incident from different directions, and no calibrated data are obtained. Recently, Hald (2005b) derived a scaling factor that provides good estimates of partial sound power by using area integration, and a similar procedure may be used to derive a pressure scaling, which results in calibrated free-field pressure assuming a monopole sound source in the focused direction.

Assuming one obtains a good approximation of free-field pressure using beamforming, there are two ways of calculating the loudness contribution in a focused direction. One is based on the free- and diffuse-field sound exposure, i.e. diotic, whereas the other is not dependent on the type of sound field and includes both diotic and dichotic (different sounds at the two ears) conditions. The loudness model for diotic sound presentations is standardized in ISO 532 (1975), and it has to be assumed that listeners turn their head toward a focused direction when employing this model. This type of loudness calculation may be useful to compare the loudness of individual sound sources independent of the listener's head rotation. On the other hand, a binaural loudness model based on dichotic sound presentations may be used in connection with beamforming to derive sound quality metrics maps dependent on the listener's head rotation, and it is useful to minimize the overall loudness at the listener's position by optimizing the location of sound sources relative to the head position.

Binaural loudness summation was modeled by Robinson and Whittle (1960), as well as Sivonen and Ellermeier (2006), specified as in Eq. 3.

$$L_{mon} = g \times log_2(2^{L_{left}/g} + 2^{L_{right}/g}) \quad (3)$$

where g is the maximal binaural gain (i.e. the loudness match between monotic and diotic stimulation), L_{mon} is the equivalent sound pressure needed for monotic stimulation to match any binaural combination of left-ear L_{left} and right-ear L_{right} input levels. Sivonen and Ellermeier (2006) proposed a "3-dB" binaural-summation rule ($g = 3$) and a binaural loudness model that estimates a single loudness value from binaural measurements, e.g. using a dummy head. Free-field pressure estimated by beamforming can be converted to binaural signals by convolving it with Head Related Transfer Functions (HRTFs) in each focused direction (Møller, 1992). Subsequently, the binaural loudness model can then be applied to the derived binaural signals. Binaural metrics may also be obtained by making use of the derived binaural loudness values.

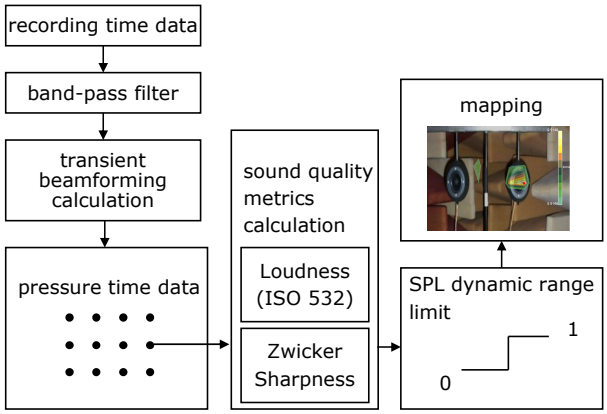


FIG. 1. Data flow to produce a sound quality metrics map.

A software for sound quality metrics mapping was developed as shown in Fig. 1. The measured time data were band-pass filtered in the frequency domain to avoid the influence of measurement noise on the sound quality metrics calculation and were subsequently passed on to the transient beamforming calculation. In the transient beamforming, the entire time signal at each microphone position was converted to the frequency domain without taking any average to reconstruct the output of beamformer in the time domain. As a result of that, pressure time data were generated for each focused direction. In case of applying binaural metrics, binaural signals were obtained by convolving HRTFs and the binaural loudness summation model was utilized to obtain the corresponding diotic pressure. Sound quality metrics calculation was applied to the pressure time data and the sound pressure in each direction was checked to find out whether the calculated values should be included in the map. An SPL dynamic range (see below) was introduced to avoid including sidelobes in the metrics calculation. Finally the processed data were transferred to the mapping software to show derived maps.

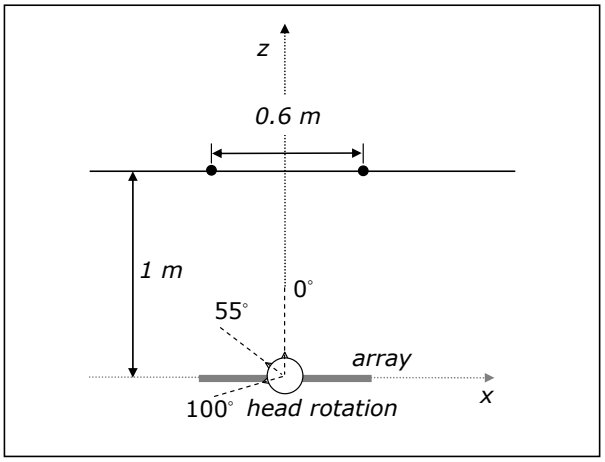


FIG. 2. Setup of monopoles and head rotation angles in the simulation.

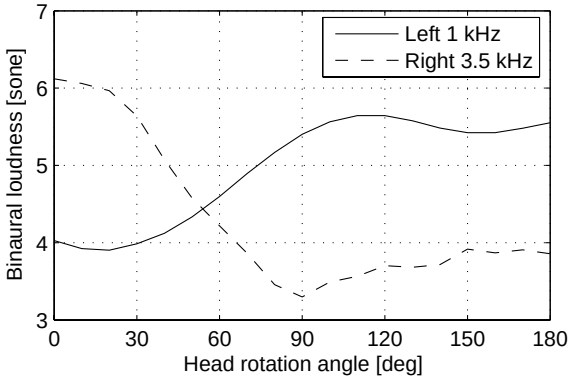


FIG. 3. Binaural loudness of each source as a function of head rotation angle. The legend indicates the location and the frequency of each source.

III. SIMULATION

The benefits of binaural loudness mapping were demonstrated in simulations using monopole sources with known SPL. For this purpose, the binaural loudness algorithm proposed by Sivonen and Ellermeier (2006) was implemented in Matlab together with transient delay-sum beamforming. Binaural signals in a focused direction were generated by convolving the corresponding HRTFs with the pressure contribution calculated by beamforming. The HRTFs employed in this study were taken from a database containing artificial-head HRTFs measured at 2° resolution (Bovbjerg *et al.*, 2000; Minnaar, 2001). HRTFs at the nearest direction were taken for the convolution rather than interpolating neighboring directions. Furthermore, the inverse HRTFs in the frontal direction were calculated using fast deconvolution with regularization (Kirkeby *et al.*, 1998), and used in the binaural loudness calculation.

In the simulation, two monopole sources were placed at 0.6 m apart from each other, and a 42-channel wheel microphone array was placed at 1 m distance from the source plane (see Fig. 2). A 1-kHz sine tone originated from the source in the left of the array center, and a 3.5-kHz sine tone from that in the right. Each sound source produced 60 dB SPL at the center of the array position. The map was generated in a grid of 1 m by 1 m with a spacing of 10 cm corresponding to 121 directions.

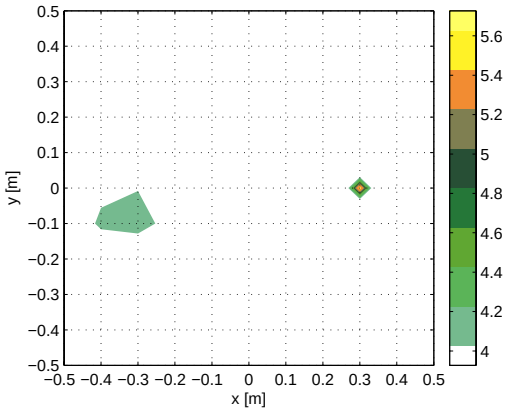
Fig. 3 shows the binaural loudness of each source as a function of head rotation angle. This does not involve the beamforming processing, but rather the direct convolution of HRTFs and source signals, and the subsequent binaural loudness calculation. Three interesting head angles, 0°, 55° and 100°, were selected from Fig. 3, and they are indicated in Fig. 2. When a listener's head points toward the frontal direction (0°), the source with the 3.5-kHz pure tone in the right hand side is louder than the left one by approximately 2 sones. At 55°, both sources are almost equally loud, and the left source with the 1-kHz pure tone is louder than the right one at 100°.

The binaural loudness maps for the three head rotation angles are shown in Fig. 4 with a dynamic range of 2 sones. Notice that the peak value of each map is adjusted to the maximum loudness value in each grid. The same conclusions as derived from Fig. 3 could be drawn in that the right source is louder than the left one at 0°, both sources are equally loud at 55°, and the left one is louder than the right at 100°. One may notice that the levels in Fig. 4 are slightly different from those in Fig. 3, and this may be due to the fact that the beamforming processing did not employ pressure scaling and thereby calculated relative sound pressures. The size of the two sources is different since the beam width of the delay-sum beamforming is inversely proportional to frequency. The results of the simulation indicate that the loudness contribution of each sound source is affected by the listener's head rotation angle, and that binaural loudness mapping is a useful tool to optimize the loudness contribution of individual sources as well as overall loudness in terms of the listener's head position.

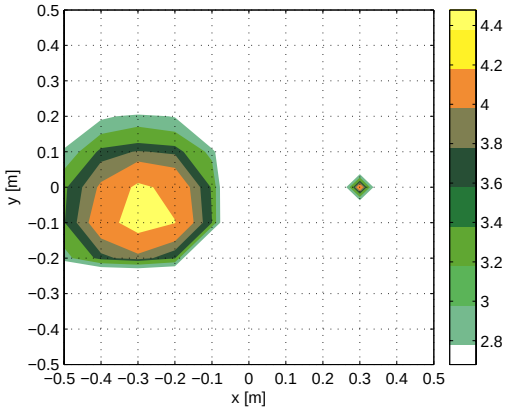
IV. LOUDSPEAKER MEASUREMENT

A. Measurement Setup

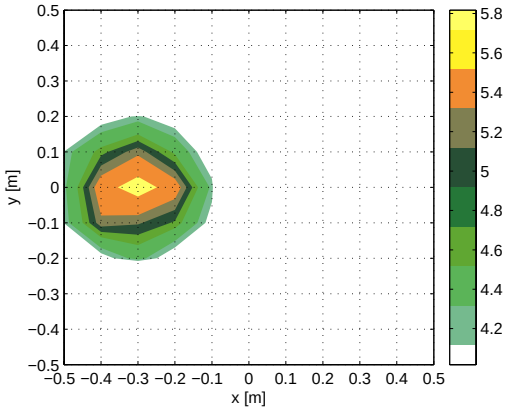
Two loudspeakers were positioned at a distance of 1.7 m from the 42-channel Brüel & Kjær (type WA 0890) beamforming wheel array in an anechoic room as show in Fig. 5. This array configuration gives approximately 10 dB side lobe suppression up to 6.4 kHz. The distance between the two loudspeakers was kept to 0.4 m and the diameter of the microphone array was approximately 1m. The height of the two loudspeakers was adjusted to the center of the microphone array. A digital camera was installed in the center of the microphone array to take pictures, which were superimposed on the contour



(a) 0°



(b) 55°



(c) 100°

FIG. 4. Binaural loudness maps at different head rotation angles. Each map shows the range of loudness values from the maximum with the dynamic range of 2 sones.

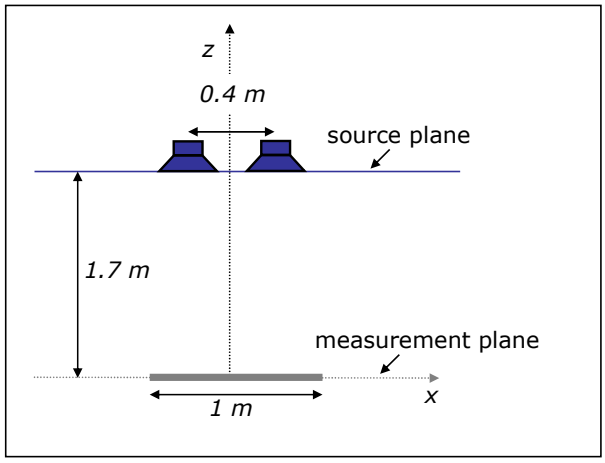


FIG. 5. Measurement setup for simulating two simultaneous sources in the anechoic chamber.

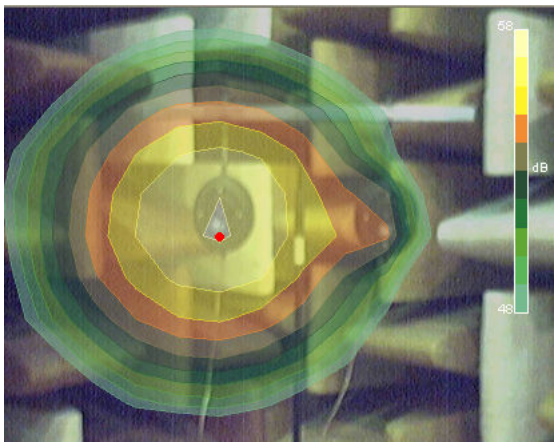
to align the sound sources with their physical location. Two microphones were excluded after the measurement due to cable breaks. The removal of these two channels resulted in a slightly lower dynamic range, but produced no changes in the localization of the two loudspeakers.

B. Stimuli

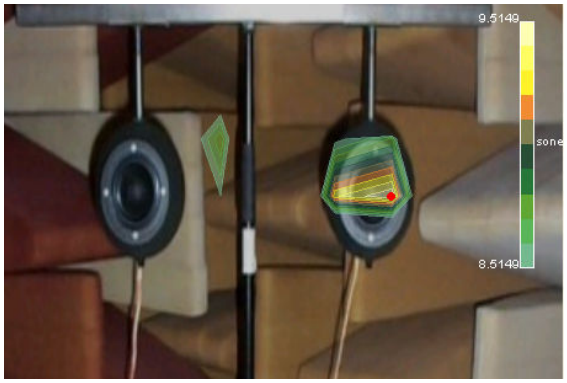
A 1-kHz reference narrow band noise was fed to the left loudspeaker and 3.5-kHz and 5-kHz noises from the right loudspeaker were compared with the 1-kHz reference. The center frequency of 3.5 kHz was chosen due to the fact that the equal loudness contour contained a dip around that frequency, which made it possible to produce low sound pressure with relatively high loudness. Since the higher pitch contents produce higher sharpness, the center frequency of 5 kHz was selected to have significant differences between sound pressure and sharpness in the given stimuli. A VXPocket 440 notebook sound card was used for the sound playback and the output of the sound card was connected to a Rotel RB-976 MKII power amplifier. The amplifier output passed through the wall between the control room and the anechoic room and was linked to both loudspeakers. During the measurement, sounds were played through both loudspeakers at the same time.

C. Results

In order to demonstrate the potential of sound quality metrics mapping, the beamforming measurement was done using the stimuli described in IV.B. Fig. 6 shows a comparison between the sound pressure and loudness mappings. The loudness mapping was created based on the loudness model for the diotic condition (ISO 532, 1975). The left loudspeaker played a 1 kHz narrow band

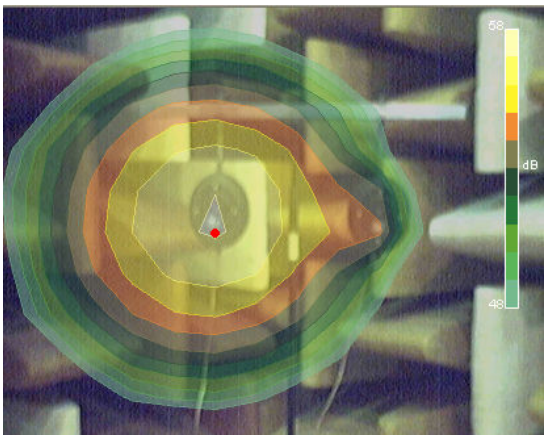


(a) pressure map



(b) loudness map

FIG. 6. Comparison of sound pressure and loudness map. For explanation, see the text.



(a) pressure map

(b) sharpness map

FIG. 7. Comparison of sound pressure and sharpness map. For explanation, see the text.

noise and the right one generated a 3.5 kHz noise. To make the advantages of a loudness map clear, the stimulus on the right loudspeaker was attenuated by 9 dB. Consequently, there is almost no sound pressure contribution from the right loudspeaker as shown in Fig. 6(a). By contrast, the loudness map in Fig. 6(b) reveals the right loudspeaker to be louder than the left. We conclude that the noise from right loudspeaker should be reduced in order to decrease total loudness while the sound pressure map prioritizes decreasing the level on the left loudspeaker.

In the next measurement, the stimulus of the right loudspeaker was a 5-kHz narrow band noise attenuated by 10 dB relative to the left, which was set to 1 kHz, as before. No sound pressure contributions are observed from the right loudspeaker as shown in Fig. 7(a). One may conclude that there is only one source, the left loudspeaker, in this situation. Sharpness was calculated at each of the focused points and the resulting map with 1 acum dynamic range is displayed in Fig. 7(b). There appear to be a number of sharp sources around the edge of the map where in fact no sources should be.

Zwicker's sharpness model is defined in Eq. 4 (Zwicker and Fastl, 2006).

$$S = 0.11 \frac{\int_0^{24} N'(z) \cdot f(z) \cdot z \cdot dz}{\int_0^{24} N'(z) \cdot dz} acum$$

$$f(z) = \left\{ \begin{array}{ll} 1 & for \quad z \leq 16 \\ 0.066 \cdot e^{0.171 \cdot z} & for \quad z > 16 \end{array} \right\} \qquad (4)$$

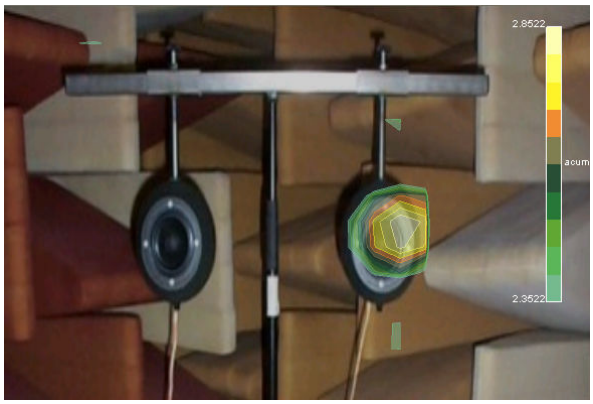
where S is the sharpness to be calculated, N' is specific loudness, $f(z)$ is a weight factor dependent on critical-band rate, and the denominator gives the total loudness N. The sharpness calculation is normalized with respect to loudness as shown in Eq. 4. Therefore, the shape of the specific loudness spectrum plays an important role in the sharpness calculation, not the magnitude.

It is known that the side lobes of the beamforming calculation produce ghost images in the calculation plane. These ghost images typically have a 10 dB lower SPL than the maximum level in the map, but the shape of their loudness spectra will be affected by the frequency contents of all sources present in the sound field. This could be the main reason why the ghost images appeared

around the edge of the frame on the sharpness map since the loudness ratio of the 1 and 3.5 kHz noise determines the sharpness values. In this study, two methods are proposed to overcome the limitation of the sharpness map caused by side lobes.

1. Making use of a combined metric, i.e. loudness × sharpness, instead of sharpness.
2. Applying a dynamic range limit by setting sharpness to "0" at the points with sound pressure below the beamforming dynamic range.

The combined metric in 1) puts weights, i.e. loudness, on each point after the sharpness calculation. Applying weights to the sharpness calculation will allow removing the ghost images, which are lower in loudness. On the other hand, using the combined metric will make it difficult to localize sharp sound sources when the loudness is dominant and there are no significant sharpness differences at the points in a given sound field. Applying a sound pressure dynamic range as in 2) is utilized to remove ghost images in a sound pressure map before sharpness calculation. This method has a drawback, since we practically have to find the beamforming dynamic range before the calculation. Fig. 8 shows that both methods improve the sharpness map such that the sharper sound source is correctly localized from the same measurement as Fig. 7.



(a) Applying sound pressure dynamic range

(b) Combined metric

FIG. 8. The improvement of sharpness mapping.

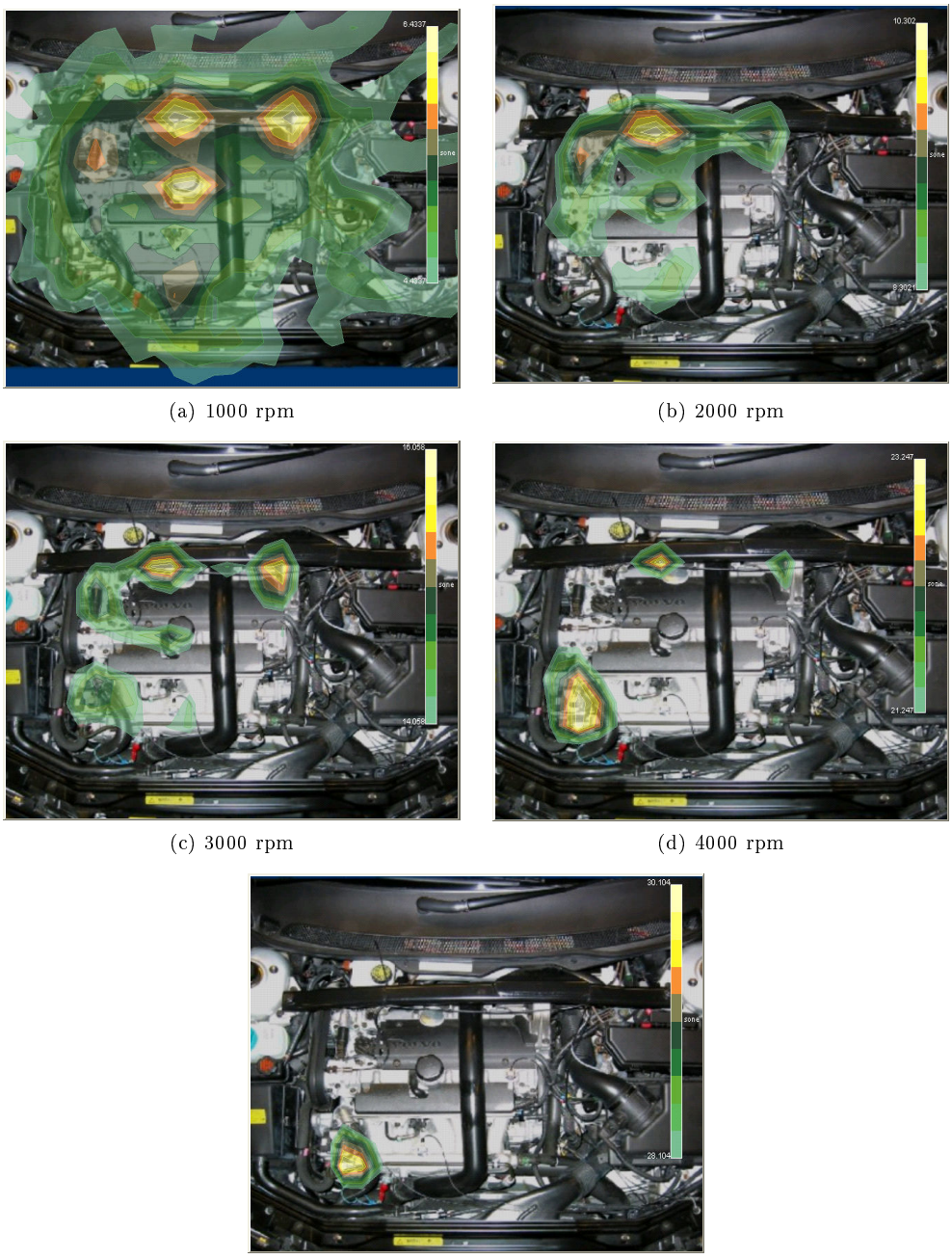
V. VEHICLE ENGINE MEASUREMENT

In order to apply the loudness and sharpness mapping to more practical situations, a measurement on an engine compartment of a personal vehicle with a 5-cylinder 4-stroke engine was performed. The car was installed in a normal exhibition room and the engine was manually controlled during the measurement. Stationary sound fields were recorded for 5 sec at 1000, 2000, 3000, 4000, 5000 rpm without any loading on the wheels. One second of time data was selected out of the 5 sec recorded samples to minimize the non-stationarity of operating condition. The maximum RPM change at 1000 rpm was 30 rpm and the rest of operating conditions kept within 15 rpm variation over the 1 sec of selected time record. A 66-channel wheel array of 1 m diameter was mounted parallel to the car engine compartment at a distance of 0.75 m. The sound field was measured by the PULSE Acoustic Test Consultant (Type 7761) with a 25.6 kHz frequency range. The transient beamforming calculation was performed by the PULSE Beamforming application (Type 7768). Due to strong reflections from the ceiling and the floor to the array and the background noise, such as air conditioning, the dynamic range of the beamforming calculation above 6.3 kHz was below 4 dB. Therefore the loudness and sharpness calculation was conducted up to 5 kHz only.

Fig. 9 shows the loudness maps for the specified operating conditions in the frequency range between 15 and 18 bark. Since we are looking at a constant frequency range, the dominant order contents are changed according to the engine RPM. Up to 3000 rpm the blank hole opposite the oil refill cap, which is located in the center of the picture, is the major source of the engine compartment. But it is clear that the power steering pump starts to be active above 4000 rpm and is dominant at 5000 rpm.

Fig. 10 compares the location of major noise sources as represented by the sound pressure and loudness maps. The engine was running at constant 5000 rpm, and the maps display the results in the frequency range between 15 to 18 bark. The loudness map is the same as that of Fig. 9(e). The pressure map identifies a major source close to the engine mount whereas the loudness map points to the source in the power steering pump. This indicates that investigating loudness maps is a more efficient way of localizing the loudest noise in the engine compartment, and the sound pressure map may lead to irrelevant design changes.

Fig. 11(a) shows the sound pressure map with a 4 dB dynamic range at 5000 rpm and Fig. 11(b) is the corresponding sharpness map with the same dynamic range. The proposed combined metric was also calculated on



(a) 1000 rpm

(b) 2000 rpm

(c) 3000 rpm

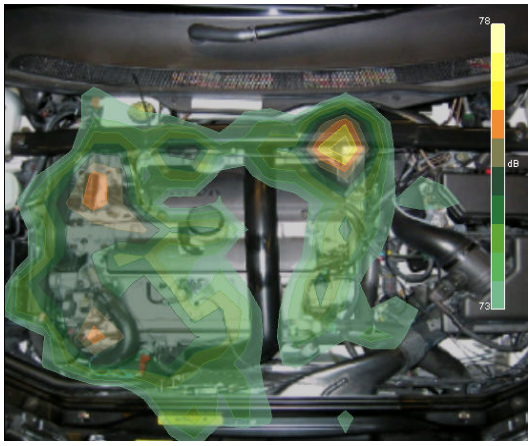
(d) 4000 rpm

(e) 5000 rpm

FIG. 9. Loudness maps of the engine compartment from 1000 to 5000 rpm. Frequency range between 15 to 18 bark.

1 to 5 kHz pre-filtered signals as shown in Fig. 11(c). The sound pressure map localized two sources, one in the power steering pump and the other in the engine top, and the shape of sources is smeared in that it covers a quite large area. On the other hand, both sharpness maps (Fig. 11(b) and Fig. 11(c)) contain the major source at the gap between the engine block and the pipes. The map 11(c) also shows there are two other sharp sound sources close to the engine mount and the lower right corner of the engine block, which might be caused by the air flow from the gap between the engine top and its cover.

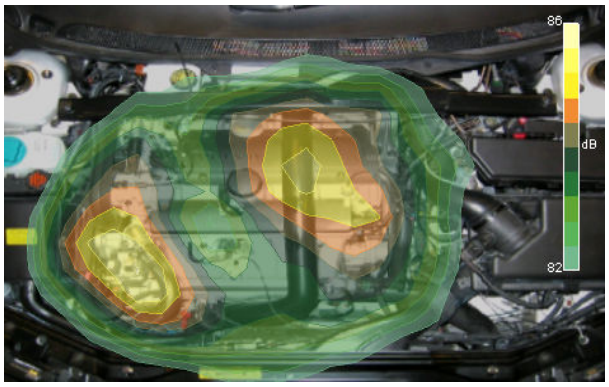
As we may see from the loudness and sharpness maps, the location of the most prominent individual sources changes dramatically depending on the metrics we are looking at. Especially since the sharpness map is affected only by the specific loudness spectrum shape of the individual sources, in general there are large differences between the loudness and sharpness maps in terms of the location of the most problematic sources.



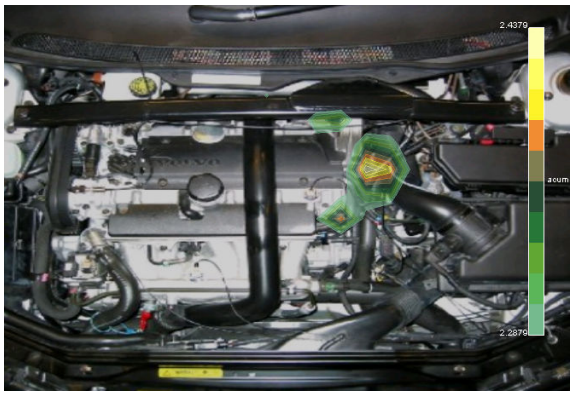
(a) pressure map

(b) loudness map

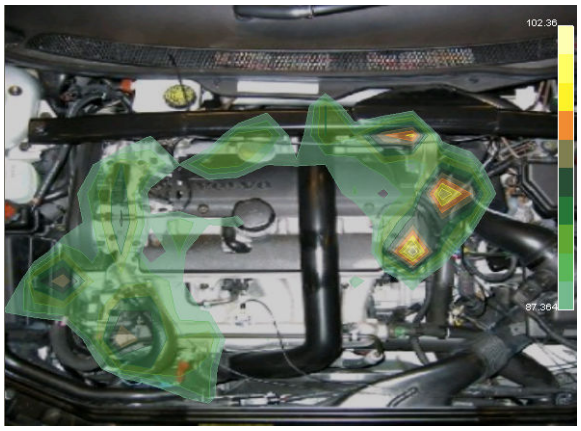
FIG. 10. The comparison of sound pressure and loudness map on the engine compartment at 5000 rpm. Frequency range between 15 to 18 bark.



(a) pressure map



(b) Applying sound pressure dynamic range



(c) Combined metric

FIG. 11. The comparison of sound pressure and sharpness map on the engine compartment at 5000 rpm. Frequency range between 1 to 5 kHz.

VI. CONCLUSION

1. A proposal of creating sound quality metrics maps was outlined and implemented both in Matlab for simulation and in a commercial software program for measurement. The standard loudness model (ISO 532, 1975) as well as a binaural loudness model based on the 3-dB loudness summation rule were utilized for loudness mapping. It was found that binaural loudness mapping provides the possibility of showing loudness maps dependent on listener's head rotation, and thereby optimizing the location of noise sources to reduce overall loudness or the loudness of partial sound fields.

2. The superiority of sound quality metrics mapping was demonstrated based on measurements on a simple loudspeaker setup in an anechoic chamber. The proposed sound quality metrics mapping localized problematic sources more efficiently compared to traditional sound pressure mapping.

3. A new combined metric and an SPL dynamic range was introduced to remove ghost images from sharpness maps both in the loudspeaker setup and on an engine compartment.

4. Practical measurements were performed on an engine compartment with a 66-channel microphone wheel array and showed that the mapping of the

sound quality metrics was applicable to automotive measurements. Traditional sound pressure maps were compared with the corresponding loudness and sharpness maps, and this suggested that relevant sound quality metrics should be selected for efficient noise source identification.

Acknowledgments

This research was carried out as part of the "Centerkontrakt on Sound Quality" which establishes participation in and funding of the "Sound Quality Research Unit" (SQRU) at Aalborg University. The participating companies are Bang & Olufsen, Brüel & Kjær, and Delta Acoustics & Vibration. Further financial support comes from the Ministry for Science, Technology, and Development (VTU), and from the Danish Research Council for Technology and Production (FTP).

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*, preprint 5264 (Los Angeles, CA, USA).

Christensen, J. J. and Hald, J. (**2004**). *Beamforming*, Technical Review number 1 (Brüel & Kjær, Nærum, Denmark).

Doclo, S. and Moonen, M. (**2003**). "Design of far-field and near-field broadband beamformers using eigenfilters", Signal Processing **83**, 2641–2673.

Hald, J. (**2005**a). "An Integrated NAH/Beamforming Solution for Efficient Broad-Band Noise Source Location", in *SAE Noise and Vibration Conference and Exhibition*, preprint 2537 (Grand Traverse, MI, USA).

Hald, J. (**2005**b). "Estimation of partial area sound power data with beamforming", in *Internoise*, preprint 1511 (Rio de Janeiro, Brazil).

Hald, J. and Christensen, J. J. (**2002**). "A class of optimal broadband phased array geometries designed for easy construction", in *Internoise* (Dearborn, MI, USA.).

ISO 532 (**1975**). "Acoustics – method for calculating loudness level", ISO, Geneva, Switzerland .

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

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

Kompis, M. and Dillier, N. (**2001**). "Performance of an adaptive beamforming noise reduction scheme for hearing aid applications. I. Prediction of the signal-to-noise-ratio improvement", J. Acoust. Soc. Am. **109**, 1123–1133.

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

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

Robinson, D. and Whittle, L. (**1960**). "The loudness of directional sound fields", Acustica **10**, 74–80.

Sivonen, V. P. and Ellermeier, W. (**2006**). "Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation", J. Acoust. Soc. Am. **119**, 2965–2980.

Zwicker, E. and Fastl, H. (**2006**). *Psychoacoustics : Facts and Models* (Springer, Berlin, Germany).

Loudness assessment of simultaneous sounds using beamforming *

Wookeun Song†

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

Sound Quality Research Unit, Department of Acoustics, Aalborg University, Fredrik Bajers Vej 7B, DK-9220 Aalborg East, Denmark

(Dated: March 11, 2008)

Listening experiments were conducted to investigate to which extent perceived loudness depends on the distribution of individual sound sources in space. Three loudspeakers were positioned 1.5 m from the center of the listener's head, one straight ahead, two 10 degrees, and two 30 degrees to the right and left, respectively. Listeners matched the loudness of either one, or two simultaneous sounds (narrow-band noises with 1-kHz, and 3.15-kHz center frequencies) to a 2-kHz, either 45-dB or 60-dB SPL narrow-band noise placed in the frontal loudspeaker. The two sounds either originated from the central speaker, or from the two offset loudspeakers. It turned out that the subjects perceived the noises to be equally loud independently of their distribution in space when the directional loudness sensitivity was equalized for individual sources. A 6-dB (pressure) loudness summation rule was suggested to calculate the overall loudness of two simultaneous sounds, and it predicted the subjective data better than did a direct loudness summation of the two sounds. This suggests that current loudness modeling will have to be extended to take the loudness summation of individual sources into account, and this may be achieved by combining beamforming techniques with the 6-dB loudness summation of sources.

I. INTRODUCTION

Identification of noise sources in a complex sound field is an important step for optimizing the noise emission from products. Typically, this has been achieved by means of array measurement techniques, such as near-field acoustic holography (Hald, 1989; Maynard *et al.*, 1985) or beamforming (Christensen and Hald, 2004; Johnson and Dudgeon, 1993). Array measurement techniques derive sound pressure (or intensity) maps, and the peak level or sound power level of individual noise sources is compared to determine the most problematic noise source.

This approach has been criticized for not taking into account psychoacoustic attributes, such as loudness. An earlier study presented a method for deriving loudness and sharpness maps from beamforming measurements (Song, 2004). Even though it is possible to localize noise sources in terms of their loudness, in some cases the loudness map may not be able to detect the loudest sound source in a sound field especially when the individual sources are similar in terms of their spatial extent and peak loudness. Also, it is desirable to predict the perceived loudness of combined sources in order to compare the loudness of sources formed over a relatively large area. Therefore, understanding the role of the spatial source distribution in loudness perception is of great importance to optimize sound fields in terms of their perceived loudness.

Comprehensive loudness models have been developed for steady sound signals (Moore *et al.*, 1997; Zwicker and Fastl, 2006), and are mostly based on diotic (the same sound at the two ears) or dichotic (different sounds at the two ears) headphone playback, or the presentation through a single loudspeaker placed in the frontal direction (Marks, 1978; Reynolds and Stevens, 1960; Scharf, 1969). Recent work revealed, however, that perceived loudness varies as a function of a sound incidence angle, measured as a directional sensitivity, when presenting a sound through a single speaker in the free field (Robinson and Whittle, 1960; Sivonen and Ellermeier, 2006). By contrast, the loudness perception of sound fields with *multiple* components and sources has not been studied from a basic-research perspective. It is desirable to take up this topic in order to combine psychoacoustic metrics with the array measurement techniques that are used for finding noise sources in complex sound fields.

To this end, a series of listening experiments was conducted based on a simple loudspeaker setup in an anechoic chamber. These experiments investigated whether loudness judgments of distributed sounds differ from judgments obtained when sounds are focused at a single location. Directional sensitivities were measured and equalized individually to get rid of the effect of the incidence angle for each of the physical sources. The cru-

*Portions of this work have been presented at the Forum Acusticum Congress, Budapest, Hungary, 2005 August 29 - September 2 and the JSAE Annual Congress, Yokohama, Japan, 2006 May 24-26

†Electronic address: wksong@bksv.com

cial comparison consists of presenting two simultaneous sounds through the center loudspeaker as opposed to two offset loudspeakers ("focused" vs. "distributed" sources).

The results are related (a) to the loudness measured in the center position of listener's head using a single microphone and (b) to measurements using a dummy head to show how well the traditional loudness model can predict the loudness judgments obtained. Subsequently, an algorithm predicting the loudness of simultaneous sounds is suggested and verified by predicting the subjective responses obtained in this study. Beamforming isolates sources of interest from competing ones by controlling the directivity pattern of a microphone array. Thus, a method of estimating the loudness of individual as well as combined sources in the presence of undesired competing sounds is outlined by utilizing beamforming and the summation algorithm found in this study.

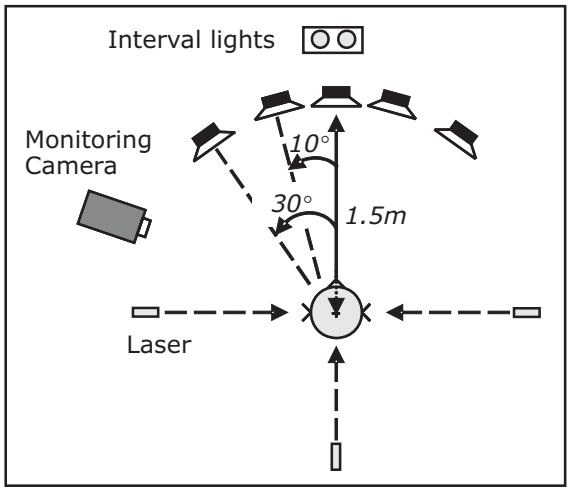


FIG. 1. Experimental setup in the anechoic chamber.

II. GENERAL METHOD

Two listening experiments were carried out using the same setup and procedure. The two experiments only differed in the angular separation of the loudspeakers employed which was 10° in the first experiment and 30° in the second.

A. Subjects

Six normal hearing subjects (5 male, 1 female), between 23 and 31 years of age, completed the first experiment, and ten (7 male, 3 female), between 23 and 32 years of age, the second experiment. All of them were students at Aalborg University, Denmark. 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 389-1, 1998) at more than one frequency. None of the thresholds exceeded 25 dB hearing level. The subjects were also screened for known hearing problems and were all paid for their participation.

B. Apparatus and materials

Third-octave band noises having center frequencies of 1, 2, and 3.15 kHz were used in the experiment. The sounds had a total duration of 1 s. To produce a flat response in the frequency range of interest, the loudspeaker impulse response functions (IRF) were measured using dual-channel FFT and their inverse filters were calculated from the average IRF using fast deconvolution with regularization (Kirkeby *et al.*, 1998). The inverse filters and third octave filters were applied to the random signals during the experiment using an FFT-based convolution. A 10-ms ramp was applied in the beginning and at the end of the stimuli in order not to generate impulsive sounds.

A computer with a sound card (RME DIGI96) was used to send digital sound signals to an external D/A converter (RME ADI-8 DI). All stimuli were played with a sampling rate of 48 kHz and delivered via loudspeakers using a power amplifier (Rotel RB-976 Mark II). A customized program written in C# controlled the experimental procedure, generated the stimuli, and collected subjects' responses.

The experiment was carried out in an anechoic chamber. Three loudspeakers with 15.5 cm diameter were positioned at 1.5 m distance from the center position of listener's head (see Fig. 1). The loudspeakers were spaced at 10° angular separations in the first experiment and 30° in the second. The listeners were seated in a height-adjustable chair with a headrest. Their head was positioned in the center of the set-up with the help of three laser beams, which were mounted to the sides, and behind the listener. A camera was mounted to monitor the head movements of the listener during the experiment. Furthermore, two lights were used to indicate observation intervals, i.e. to assist the listeners in mapping the sound sequence to response buttons. The listeners gave their responses using a two-button box connected to a parallel port in the computer.

C. Procedure

Two types of experiments were performed in sequence. They were distinguished by the number of narrow-band noises playing simultaneously. Only one narrow-band noise was presented at a time in the single-sound condition, and two narrow-band noises having different center frequencies were played simultaneously during the dual-sound condition. Both conditions shared the same reference stimuli, which were 2-kHz narrow-band noises hav-

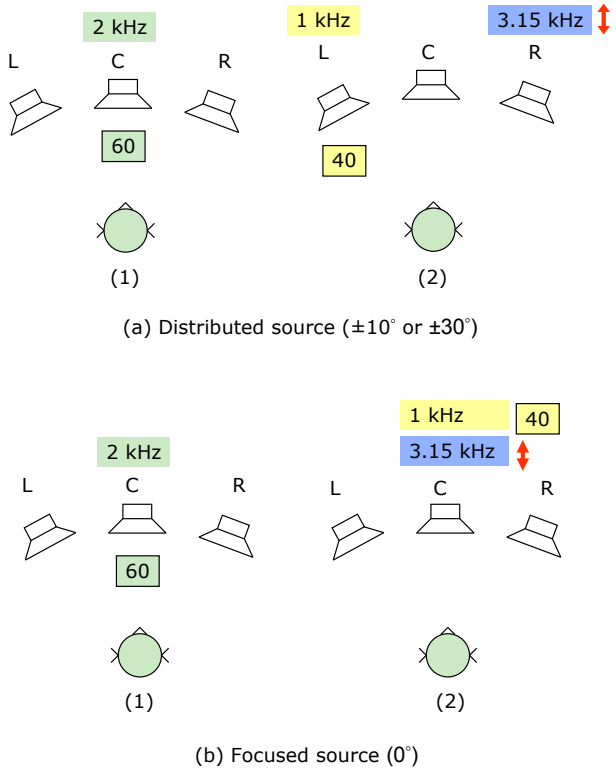


FIG. 2. Sound presentation examples in the dual-sound condition. The left pictures indicate a reference stimulus, and the right ones test stimuli. The values in the rectangular boxes show level in dB SPL.

ing either 45 or 60 dB SPL. The two sounds, the 1-kHz and 3.15-kHz noise, were presented simultaneously from the two offset loudspeakers or at the center in the dual-sound condition. Two narrow-band noises of different center frequencies were utilized in order to produce two distinct sound sources and in this way the relationship between perceived loudness and the distribution of sources could be investigated for independent sound sources. The noise centered at 1 kHz had fixed sound pressure levels of 30, 35, or 40 dB SPL with the 45-dB reference and 40, 50, or 55 dB with the 60-dB reference. The level of the 3.15-kHz noise was varied during the experiment.

Fig. 2 shows two examples of sound presentations in the dual-sound condition. The reference signal, i.e. a 2-kHz noise having 60 dB SPL, played at the center in both examples (see the left pictures of Fig. 2). In the first example (a), the 1-kHz noise originated from the left loudspeaker, and its level was fixed to 40 dB SPL. The variable 3.15-kHz noise played at the same time from the right loudspeaker. In the half of the experimental conditions, the variable 3.15-kHz noise was placed in the left loudspeaker, whereas in the other half in the right. The fixed 1-kHz noise then played from the opposite side of the 3.15-kHz noise. In the second example (b), both the fixed 1-kHz and the variable 3.15-kHz noise were presented simultaneously from the frontal loudspeaker.

An adaptive two-interval, two-alternative forced choice procedure (Jesteadt, 1980; Levitt, 1971) using a one-up, one-down rule was employed. The starting level of the variable 3.15-kHz noise was randomized for each track in the range of 60±10 dB in the first experiment, and was selected either 10 dB above or 10 dB below the loudness match of the reference in the selected direction in the second experiment. There was a 500-ms pause between the two sound presentations on each trial. In total, eight reversals were collected in each track and the last four reversals were averaged to calculate the loudness match for each track.

The reference signal played from the frontal loudspeaker, and a test signal (detailed below) were played in randomized order within a trial, and the subject's task was to say whether the first or the second sound was louder by pressing the respective button on the response box. The level of next presentation in a track was changed according to the subject's previous response. Subjects were asked to judge perceived loudness, not any other changes in sounds occurring during the experiment. Also, they were asked to judge the entire sound, not just a component of it. One block of trials consisted of 6 tracks in the single-sound condition, or 9 in the dual-sound condition, and lasted about 15 minutes on average. Both the order of tracks and the succession of trials in each track were randomized separately for each subject. One session consisted of 4 blocks. The subjects took a 30-s break between blocks, and a 5-minute break after two blocks were finished. In the beginning of the experiments, four practice blocks were completed prior to the data collection proper.

The adaptive procedure controlled the SPL of the test stimuli until they had the same perceived loudness as the reference. In both experiments, the same reference signal presented through the center loudspeaker was used. The 60-dB reference was used with test signals of 0° and 10° angular separation in the first experiment, and a 45 and 60-dB reference was used with the 0° and 30° angular separation in the second. The 45 and 60-dB reference conditions in the second experiment were counterbalanced according to the ABBA scheme (Montgomery, 2001).

III. EXPERIMENTS

A. Directional effects: single-sound condition

1. Rationale

The direction of incidence is one factor affecting loudness perception of sounds. A direct measure of this is the difference in perceived loudness of a sound arriving from different directions, and defined here as "directional sensitivity". This directional sensitivity should be equalized for each loudspeaker to investigate the separate effect of spatial distribution on perceived loudness. The main objective of the single-sound condition was to measure the

directional sensitivity as a function of the center frequencies and the loudspeaker positions for each subject, and to adjust the gain for each channel accordingly prior to the dual-sound condition.

2. Results

Loudness matches for each subject and condition were determined as a function of loudspeaker position, i.e. $-30°$, $-10°$, $0°$, $10°$, and $30°$. The directional sensitivity of each individual listener was calculated by subtracting his or her loudness match for the offset position from the match made for the center. A positive directional sensitivity indicates that the sound from that direction is being perceived as being louder than that from the center. The loudness matches and directional sensitivities of subject BJ at different center frequencies are shown in the upper and the lower panel of Fig. 3 respectively. The level of the reference was 60 dB SPL. The averages are based on 12 runs in the first experiment and 8 runs in the second shown with 95%-confidence intervals.

The discrepancy between the two curves illustrates the frequency dependence of loudness, and the difference between loudspeaker positions marked along the abscissa shows the directional effect on loudness. Subject BJ perceived the 3.15-kHz noise to be louder than the 1-kHz noise requiring approximately 7 dB less to obtain a match. Furthermore, he perceived the 1-kHz noise to be louder by approximately 2 dB when presented from the offset loudspeaker. It can be seen that there is no significant difference on the directional sensitivity of the two separation angles, and the average directional sensitivity for the 1-kHz noise was approximately 1.5 dB higher than that of the 3.15-kHz noise. The directional sensitivities were compensated individually for each listener prior to participating in the dual-sound condition.

The directional sensitivity outcomes of each experimental condition were averaged across six subjects in the first experiment and ten subjects in the second, and are shown in Fig. 4. The directional sensitivity patterns were not level dependent on average (see the lower versus the upper panel of Fig. 4). The 3.15-kHz noise was perceived to be equally loud independent of sound incident angle (dashed line) whereas the subjects perceived the 1-kHz noise from the offset loudspeakers to be approximately 1 dB louder than that from the frontal direction.

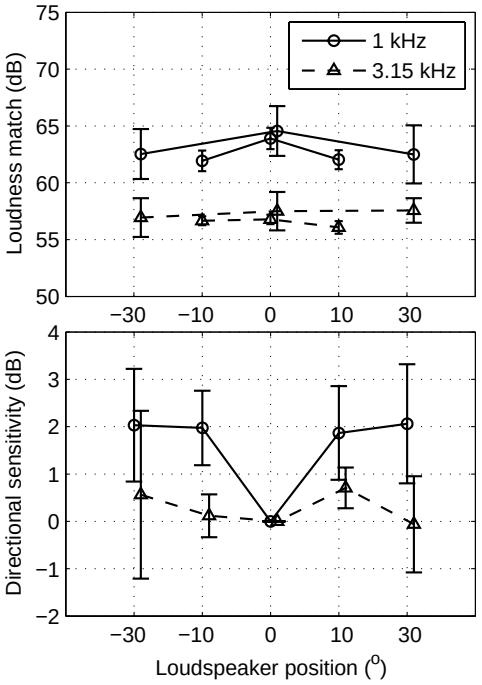


FIG. 3. The result of subject BJ for different loudspeaker locations and two center frequencies. Negative angle indicates the left-hand side of the subjects. The SPL of the reference was 60 dB. The upper panel indicates the loudness matches and the lower the directional sensitivities.

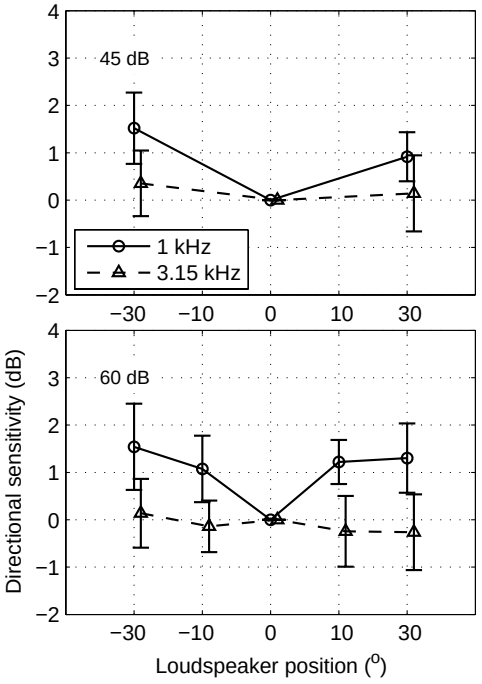


FIG. 4. Average directional sensitivity for different loudspeaker locations and two center frequencies. Negative angle indicates the left-hand side of the subjects.

B. Source distribution: dual-sound condition

1. Rationale

The directional sensitivities for each subject and condition found in the single-sound condition were employed to equalize the sources used in the dual-sound condition. This implies that the independent directional effect of a narrow-band noise was removed, so the interaction

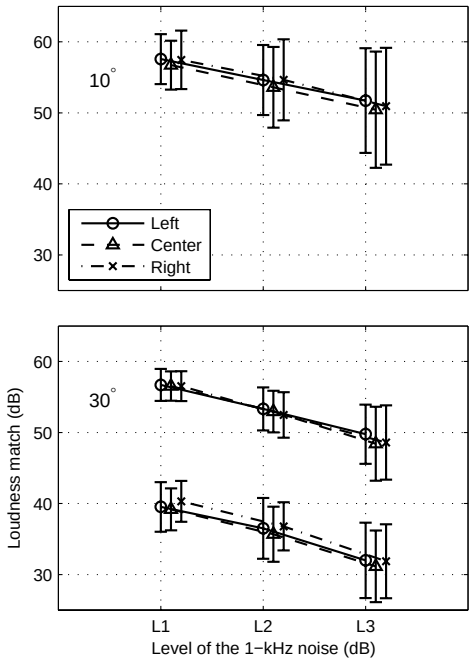


FIG. 5. Average loudness matches of the variable 3.15-kHz noise for different locations with the fixed 1-kHz noise in the opposite loudspeaker. The lower curves in the lower panel indicate the results in the 45-dB reference condition, the others in the 60-dB reference condition. L1, L2, L3 were 30, 35, 40 dB SPL with the 45-dB reference and 40, 50, 55 dB SPL with the 60-dB reference.

between the two simultaneous noises and their physical source locations could be investigated in the dual-sound condition. The goal of this investigation was to understand the role of the spatial arrangement of sources in perceived loudness, to test whether the loudness of a single sound differed from that of two simultaneous sounds, and to derive the loudness summation rule for simultaneous sound sources.

2. Results

The average loudness matches of the dual-sound condition are shown in Fig. 5 when two simultaneous noises, the 1-kHz and 3.15-kHz noise, with the level of the latter being variable, were matched to the reference, a 2-kHz noise having either 45 or 60 dB SPL. The averages are based on 12 runs in the first experiment and 8 runs in the second shown with 95%-confidence intervals. Solid, dashed, and dash-dotted lines represent the data when the 3.15-kHz noise originated from the left, center, and right loudspeaker respectively, with the fixed 1-kHz noise being placed in the right, center, and left loudspeaker. The noise centered at 1 kHz had fixed sound pressure levels (L1, L2, L3) of 30, 35, or 40 dB SPL with the 45-dB reference and 40, 50, or 55 dB with the 60-dB reference. The ordinate in Fig. 5 is the SPL of the variable 3.15-kHz noise.

It can be seen that the loudness matches decreases monotonically when the level of the 1-kHz noise is increased. If the slope of the curves is steep, the subjects take the loudness of the 1-kHz noise more into account when judging overall loudness. The slope seems to be independent of the location of the 3.15-kHz noise and reference level, and specifies the influence of the 1-kHz noise when the subjects judge the overall loudness of the two simultaneous sounds. The present result implies that overall loudness depends on the SPLs of both component sources, and is largely independent of their spatial distribution.

From Fig. 5, the difference in loudness matches between the center and the offset speaker(s), defined here as "distribution sensitivity", was calculated and listed in Table I. Negative distribution sensitivity indicates that the two simultaneous sounds were perceived to be softer when presented through the two offset loudspeakers (distributed) rather than mixed in the center loudspeaker (focused). Most of mean distribution sensitivities had small negative values (see in Table I). To determine whether they are statistically different from zero, one-sample t-tests (two-tailed, $\alpha = 0.05$) were performed. Table I summarizes the outcome: three of the 18 tests are statistically significant and the rest of them are not. This shows that the distribution of physical sources does not influence the perceived loudness of simultaneous sounds with multiple components when the directional sensitivity of each sound source was equalized.

IV. PHYSICAL MEASUREMENTS

After the experiments were completed, the stimulus conditions resulting in subjective loudness matches were recorded in the same set-up. This was done by reproducing the sound and recording it with a microphone placed at the center of the listener's head, as well as a dummy head in the center of the set-up. A conventional loudness metric (assuming diotic presentation) and binaural loudness (dichotic presentation, i.e. using the two input levels at the artificial ears of the dummy head) were calculated from the recordings of a microphone or a head-and-torso simulator (Brüel & Kjær Type 4100), respectively, using Brüel & Kjær Sound Quality (Type 7698) software.

Fig. 6 shows the measured loudness based on a single microphone and Fig. 7 the binaural loudness of the sound fields in the single-sound (upper panel) and dual-sound (lower panel) conditions. The loudspeakers were spaced at 10°, and the reference level was 60 dB. Binaural loudness was obtained based on the 3-dB summation rule proposed by Sivonen and Ellermeier (2006). The dotted line is the measured loudness of the reference. The analogous measurements made for all mean matches generated

TABLE I. Mean distribution sensitivity: the difference in loudness matches between the focused and the distributed sounds (boldface p-values indicate significant deviations from zero based on one-sample t-tests)

Separation angle	Ref. level	Location of the 3.15-kHz noise	Level of the 1-kHz noise		
			L1	L2	L3
10°	60 dB	left	-0.84	-1.04	-1.28
			p=0.011	p=0.071	**p=0.030**
10°	60 dB	right	-0.74	-1.06	-0.48
			p=0.147	p=0.058	p=0.234
30°	45 dB	left	-0.34	-0.83	-0.84
			p=0.535	p=0.265	p=0.240
30°	45 dB	right	-1.13	-1.12	-0.72
			p=0.039	p=0.090	p=0.293
30°	60 dB	left	-0.17	-0.38	-1.35
			p=0.552	p=0.291	p=0.110
30°	60 dB	right	0.02	0.48	-0.18
			p=0.940	p=0.091	p=0.395

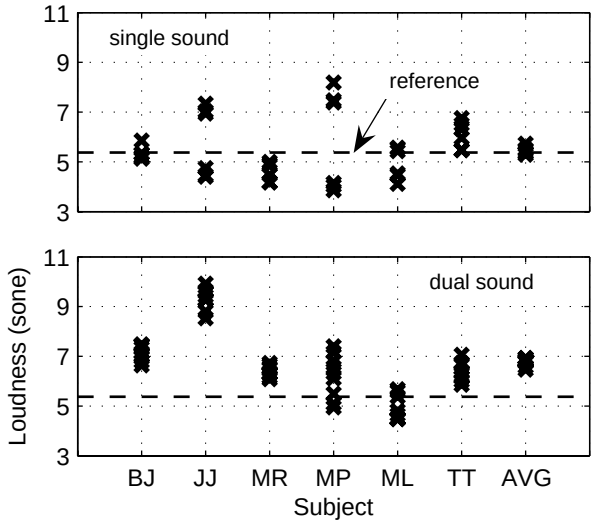


FIG. 6. Conventional loudness metrics (diotic presentation) for the sound fields produced by each subject when making a match: single-sound condition (upper panel), and dual-sound condition (lower panel). Six (upper panel) or nine (lower panel) measured loudness values corresponding to the experimental conditions are marked. The loudspeakers were spaced at 10°, and the reference level was 60 dB.

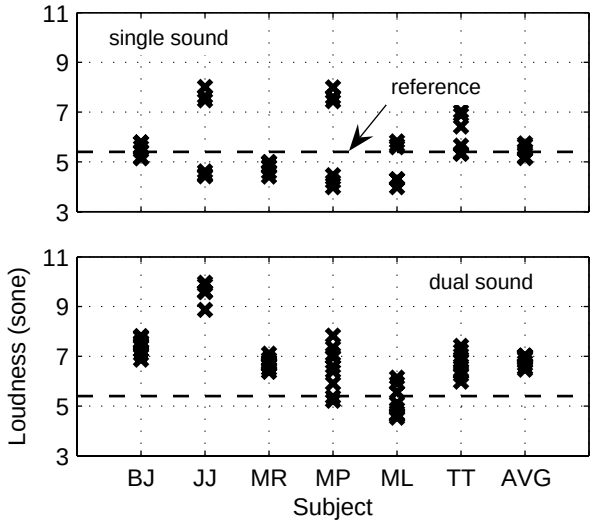


FIG. 7. Binaural loudness (dichotic presentation) of the sound fields produced by each subject when making a match: single-sound condition (upper panel), and dual-sound condition (lower panel). Six (upper panel) or nine (lower panel) measured loudness values corresponding to the experimental conditions are marked. The loudspeakers were spaced at 10°, and the reference level was 60 dB.

by each of the listeners in the 6 (single-sound condition) and 9 (dual-sound condition) experimental conditions are marked by crosses. If the loudness model worked perfectly, then all data points should coincide with the dotted line depicting the measured loudness of the reference stimulus.

In general, the measured loudness values are scattered around the reference for the single-sound condition (see the upper panels of Fig. 6 and Fig. 7), and tend to lie slightly above the reference for the dual-sound condition (see the lower panels of Fig. 6 and Fig. 7). Subject JJ adjusted the SPL to be much higher (resulting in higher

measured loudness) when the sound consisted of more than one narrow-band noise, whereas ML created no measurable difference between the two conditions. Generally, the single-microphone and dummy-head measurements are in close agreement. This is because the two offset loudspeakers were positioned close to each other so that the interaural level difference of the offset positions was quite small.

It was expected that the conventional loudness metric would accurately predict the subjective data on average as the two offset loudspeakers were positioned close to the center (10°, and 30°). From the average result (see AVG on the x-axis of Fig. 6 and Fig. 7), it is noticeable that both conventional (diotic presentation) and binaural (dichotic presentation) loudness calculations predict the matches well for one narrow-band noise, but overestimate them by about 1 sone when two noises are presented at the same time. This also means that the subjects will perceive two simultaneous noises to be softer by approximately 1 sone than one narrow-band noise if the sound fields for the given conditions are equalized by a conventional loudness meter.

The results of the dual-sound condition suggest that current loudness modeling will have to be extended to perform spatial loudness summation of individual sound sources identified in a sound field. This shall be done in the next section.

V. MODELING

The loudness matches, in which one component of the two simultaneous sounds was varied in level, revealed that both components were relevant, and contributed to overall loudness in a strictly monotonic fashion. Therefore, the goal of this section is to develop an algorithm, which predicts the loudness matches that the subjects made, and to explore the possibility of applying it to the loudness evaluation of multiple noise sources.

A. Loudness summation of sound sources

In order to predict the overall loudness of simultaneous sounds having different center frequencies, the SPL of the 1-kHz and 3.15-kHz noises was converted to the equivalent SPL of the 2-kHz noise. Equivalent SPL is defined in this study as the sound pressure level of the 2-kHz narrow-band noise that produces the same perceived loudness as the corresponding 1-kHz or 3.15-kHz noise. The individual as well as the average loudness matches are displayed as a function of equivalent SPL of the fixed 1-kHz component in Fig. 8. The data points in Fig. 8 indicate equal loudness of the two simultaneous sounds. In general, the individual data scatter around the average data marked by solid triangles and the equivalent SPL of the 3.15-kHz noise decreases when the level of the 1-kHz noise is increased.

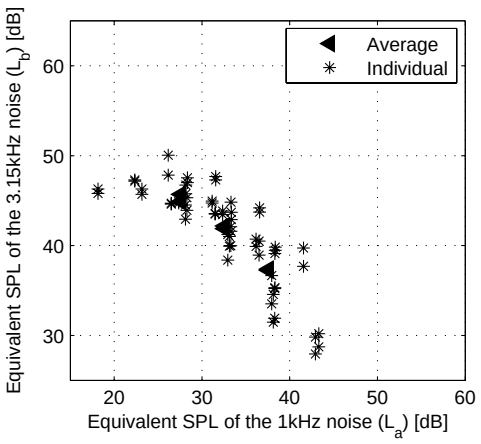


FIG. 8. Loudness matches of the individual subjects as a function of equivalent SPL (Reference: 45 dB, Incident angle: 30°)

A loudness summation rule, Eq. 1, is proposed to calculate the overall loudness of two simultaneous sounds:

$$L_{ref} = \beta log_2(2^{L_a/\beta} + 2^{L_b/\beta}) \quad (1)$$

Here, β is a gain constant characterizing the degree of summation, L_a and L_b are the equivalent SPLs of the two simultaneous sounds, and L_{ref} is the SPL of the reference. The equation has the same structure as the one proposed for binaural loudness summation by Robinson and Whittle (1960).

The average data for each source separation angle and reference SPL were fitted with Eq. 1 using the nlinfit function in Matlab, which estimates the coefficients of a nonlinear regression function using least squares estimation. The results are summarized in Table II. The fitted gains are close to a 6-dB loudness summation rule, i.e. Eq. 1 with $\beta = 6$, in most conditions. The estimated loudness summation was slightly higher in the 45-dB condition than in the 60-dB condition. The 6-dB rule implies that if there are two simultaneous sounds having the same SPL, then the "equal-loudness" SPL of the combined sound is 6 dB higher than that of each component.

In Fig. 9, the 6-dB rule is plotted together with all average data, and it appears that the 6-dB rule fits the average data fairly well both for the 45 and 60-dB references. This shows that our subjects behaved as if they were summing the SPL of individual sources according to the 6-dB rule when judging overall loudness.

B. Verification of the algorithm

A verification of the 6-dB rule was performed by generating noises for each average loudness match, calcu-

TABLE II. Loudness summation in dB (β in Eq. 1) for each angle and reference SPL

Reference SPL	0°	10°	30°
45 dB	7.2	-	6.6
60 dB	6.4	6.0	5.9

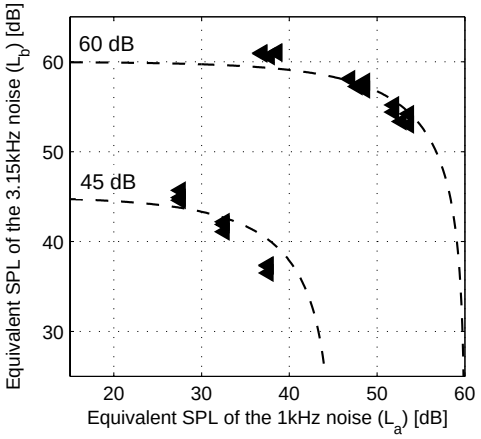


FIG. 9. Average loudness matches (Reference: 45 and 60 dB, Incident angle: 0°, 10°, and 30°) across the subjects as a function of equivalent SPL with the 6-dB loudness summation rule (dashed line)

lating their loudness from the signals, and subsequently comparing the calculated loudness of the two component signal with that of the reference. For the 6-dB rule, the calculation consisted of the 1/3 octave analysis of individual sources, the 6-dB loudness summation based on Eq. 1, and loudness calculation of the combined spectrum according to ISO 532 (1975). As an alternative prediction, the loudness of each component was calculated, and subsequently summed to compare with the 6-dB rule. Fig. 10 illustrates the outcome of the simulation with the reference loudnesses marked as dashed lines. Note that for an optimal prediction, the computed loudness of the combined source denoted by a cross or asterisk should coincide with its reference to which it was actually matched in the experiments. Both algorithms do fairly well for the 45-dB reference conditions, but the direct loudness summation rule overestimates total loudness by approximately 1 sone in the 60-dB reference conditions. Thus, the 6-dB rule proposed in this study appears to make a better prediction than the direct loudness summation of individual sources.

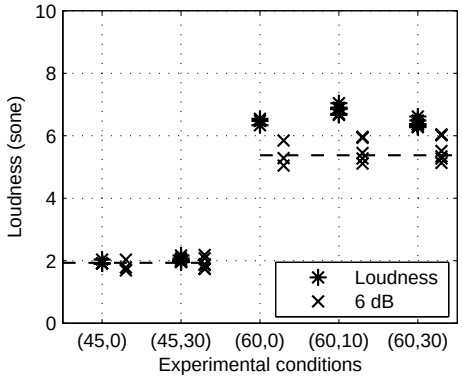


FIG. 10. Loudness prediction of the average data using a direct loudness summation (marked as an asterisk) and the 6-dB rule (marked as a cross). The first number in the parentheses specifies the SPL of the reference in dB and the second the incident angle in degrees.

C. Loudness calculation using beamforming

Typically, in sound quality applications, a number of sound sources are under investigation, and competing sources in a sound field have to be suppressed for the further calculation. Beamforming calculates the sound pressure contribution from sound sources of interest and minimizes the effect of competing sources by controlling the directivity pattern of a microphone array. In an earlier study (Song, 2004), a method of loudness mapping was suggested to identify those sound sources that would appear particularly loud to human listeners. It was emphasized that the loudness of sources including combined ones should be calculated in order to rank them in terms of psychoacoustical salience rather than according to some physical metric such as sound power. For that reason, a procedure for calculating the loudness of sources identified by loudness mapping will be described here.

First, the inspection of a loudness map is of importance to identify sources of interest and define the area of individual sources on a source plane for further processing. Secondly, the sound pressure contribution of individual sources should be calculated using beamforming to generate the 1/3-octave spectrum of each source. The output of the beamforming process should be scaled in order to estimate the free-field sound pressure in the absence of the microphone array, and this can be achieved by utilizing an intensity scaling of the beamforming output (Hald, 2005). Subsequently the 1/3-octave spectra of sources should be merged into a combined spectrum using the 6-dB loudness summation rule, see Eq. 1. The final step is to calculate the loudness of the combined spectrum based on the free-field loudness model described in ISO 532 (1975).

This procedure provides a method of calculating the loudness of combined sources as well as a single sound

source with minimal interference from competing sources. This may be distinguished from traditional methods, such as a single microphone or artificial head measurements where it is generally very difficult to separate the sources of interest from competing ones.

VI. CONCLUSION

(1) Loudness matches of simultaneous narrow-band noises were performed in an anechoic environment, and the results show that the distribution of individual sources does not influence overall loudness judgment provided the directional loudness sensitivity of hearing compensated for. The outcome does not seem to be dependent on the overall sound-pressure level of the stimuli and the separation angle of sources in the range investigated.

(2) Microphone as well as dummy head measurements were performed for the stimulus conditions resulting in subjective loudness matches. They showed that conventional loudness measurements overestimated the dual-sound conditions by approximately 1 sone compared to the single-sound conditions.

(3) A 6-dB loudness summation rule was proposed to estimate the overall loudness of two simultaneous sound sources, and it predicted the subjective data better than a direct summation of each source's loudness metric.

(4) A procedure for the loudness estimation of multiple sound sources using beamforming was outlined and the benefits of this approach were illustrated. The procedure can be used for evaluating the loudness of individual as well as combined sources in the presence of background noise or competing sources.

Acknowledgments

This research was carried out as part of the "Centerkontrakt on Sound Quality" which establishes participation in and funding of the "Sound Quality Research Unit" (SQRU) at Aalborg University. The participating companies are Bang & Olufsen, Brüel & Kjær, and Delta Acoustics & Vibration. Further financial support comes from the Ministry for Science, Technology, and Development (VTU), and from the Danish Research Council for Technology and Production (FTP). The authors would like to thank Ville Pekka Sivonen for valuable discussions of binaural loudness.

Christensen, J. J. and Hald, J. (**2004**). *Beamforming*, Technical Review number 1 (Brüel & Kjær, Nærum, Denmark).
Hald, J. (**1989**). *STSF – a unique technique for scan-based Near-field Acoustic Holography without restrictions on coherence*, Technical Review number 1 (Brüel & Kjær, Nærum, Denmark).
Hald, J. (**2005**). "Estimation of partial area sound power data with beamforming", in *Internoise*, preprint 1511 (Rio de Janeiro, Brazil).
ISO 389-1 (**1998**). "Reference zero for the calibration of audiometric equipment – part 1: Reference equivalent threshold sound pressure levels for pure tones and supra-aural earphones", ISO, Geneva, Switzerland .
ISO 532 (**1975**). "Acoustics – method for calculating loudness level", ISO, Geneva, Switzerland .
Jesteadt, W. (**1980**). "An adaptive procedure for subjective judgments.", Perception & Psychophysics **28**, 85–88.
Johnson, D. H. and Dudgeon, D. E. (**1993**). *Array Signal Processing: Concepts and Techniques* (Prentice Hall, London, Great Britain).
Kirkeby, O., Nelson, P. A., Hamada, H., and Orduna-Bustmante, F. (**1998**). "Fast deconvolution of multichannel systems using regularization", IEEE Transactions of Speech and Audio Processing **6**, 189–194.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics", J. Acoust. Soc. Am. **49**, 467–477.
Marks, L. E. (**1978**). "Binaural summation of the loudness of pure tones", J. Acoust. Soc. Am. **64**, 107–113.
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.
Montgomery, D. C. (**2001**). *Design and Analysis of Experiments, 5th edition* (Wiley, New York, USA).
Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). "A model for the prediction of thresholds, loudness, and partial loudness", J. Audio Eng. Soc. **45**, 224–240.
Reynolds, G. S. and Stevens, S. S. (**1960**). "Binaural summation of loudness", J. Acoust. Soc. Am. **32**, 1337–1344.
Robinson, D. and Whittle, L. (**1960**). "The loudness of directional sound fields", Acustica **10**, 74–80.
Scharf, B. (**1969**). "Dichotic summation of loudness", J. Acoust. Soc. Am. **45**, 1193–1205.
Sivonen, V. P. and Ellermeier, W. (**2006**). "Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation", J. Acoust. Soc. Am. **119**, 2965–2980.
Song, W. (**2004**). "Sound quality metrics mapping using beamforming", in *Internoise*, preprint 271 (Prague, Czech Republic).
Zwicker, E. and Fastl, H. (**2006**). *Psychoacoustics : Facts and Models* (Springer, Berlin, Germany).

Loudness threshold for a secondary sound source

Wookeun Song*

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

Sound Quality Research Unit, Department of Acoustics, Aalborg University, Fredrik Bajers Vej 7B, DK-9220 Aalborg East, Denmark

(Dated: March 13, 2008)

The threshold below which a secondary sound does not contribute to overall loudness was measured in a series of listening experiments. 1-kHz and 3.15-kHz narrow-band noises were used as stimuli, and they were auralized binaurally either in the frontal direction (0°) or in two offset directions (±30°) using dummy-head HRTFs. The influence of psychophysical method on loudness threshold was investigated by employing both an adaptive procedure and free magnitude estimation. The experiments also investigated dual-frequency and single-frequency conditions. In the dual-frequency condition, the two simultaneous noises were centered at different frequencies, whereas they shared the same center frequency in the single-frequency condition. The results of the dual-frequency condition showed that the loudness threshold for a secondary sound was much higher than the threshold of hearing, indicating that the secondary noise was clearly audible but did not contribute to overall loudness. Results from the single-frequency condition generally agree with traditional JNDL studies. The two psychophysical methods produced similar loudness thresholds for the secondary sound. The influence of spatial source separation (0° and ±30°) was more obvious in the single-frequency condition than in the dual-frequency condition. The outcome of this study may be useful to design a microphone array that is more suitable for estimating the loudness of individual sources in the presence of competing ones, and for calculating the loudness of simultaneous sounds.

I. INTRODUCTION

Beamforming has been widely employed in the field of sound source identification and in speech communication applications (Christensen and Hald, 2004; Johnson and Dudgeon, 1993; Li, 2005). The method is particularly interesting when a given test object, such as an automobile engine, is connected to complicated physical obstacles, and thereby Near-field Acoustic Holography (NAH) (Hald, 1989; Maynard *et al.*, 1985) can not easily be implemented due to its requirement of placing a microphone array close to the source. Moreover, beamforming is suited to generate sound pressure or intensity maps at relatively high frequencies with the help of employing irregular microphone positions. On the other hand, the spatial resolution of beamforming is relatively poor compared to NAH especially at low frequencies and that is why the combination of NAH and beamforming has been investigated to localize noise sources in a wide range of frequencies (Hald, 2005).

In an earlier study, a method for producing loudness maps using beamforming was presented and its superiority was demonstrated by comparing it with conventional sound pressure mapping (Song, 2004). Loudness maps can be derived by collecting pressure time data in a focused direction and subsequently computing a specific loudness spectrum based on the assumption of diotic sound presentation (ISO 532, 1975). Assuming that monopole sound sources are distributed in a source plane and that beamforming steers its directivity toward a particular monopole, this implies that the listeners turn their heads toward that monopole and judge its loudness. A loudness mapping may then be derived by scanning the beam over the source plane.

One of the main issues in beamforming is the creation of sidelobes in the array directivity pattern. A number of investigations have been conducted to reduce the level of the maximum sidelobes, and they either work by applying shadings to the signals measured at the microphones depending on their position (Gallaudet and de Moustier, 2000), or by placing the microphones in the array (Christensen and Hald, 2002) in a way that minimizes the maximum sidelobes. Despite such efforts, sidelobes are inevitable in the processing of beamforming measurements. Sidelobes are clearly audible when auralizing a sound source in the focused direction using beamforming and may influence the perceived loudness. Therefore, it is of interest to investigate the threshold of perceived loudness, below which sidelobes, i.e. secondary sounds, do not contribute to overall loudness.

In beamforming, competing sources generating sound from other directions contribute to the estimation of

*Electronic address: wksong@bksv.com

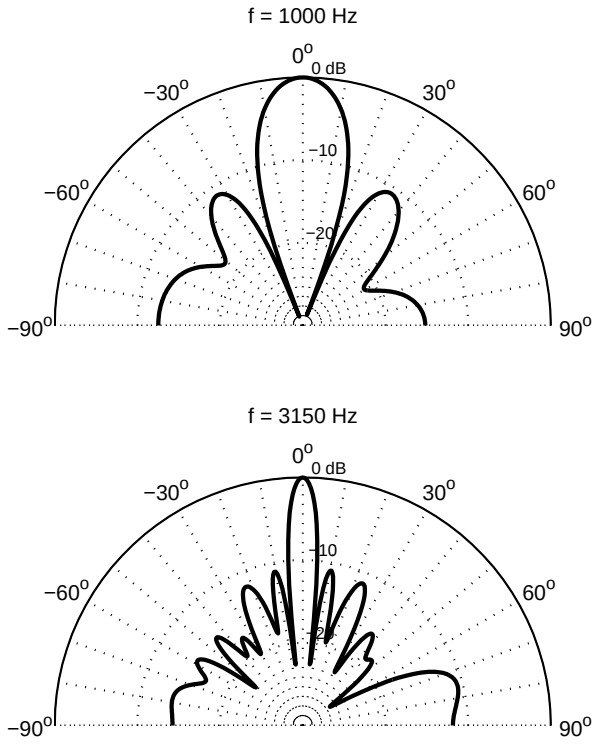


FIG. 1. The array directivity pattern of a 42-channel wheel array at 1 and 3.15 kHz.

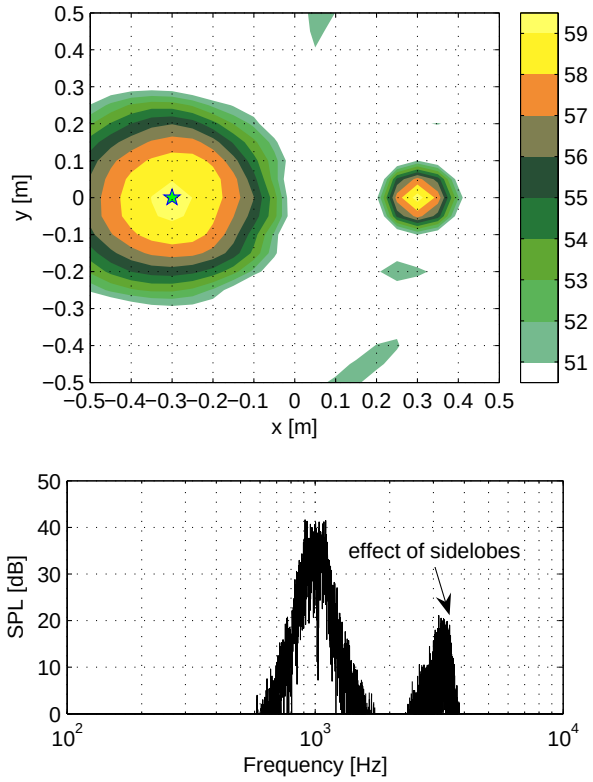


FIG. 2. The array directivity pattern of a 42-channel wheel array at 1 and 3.15 kHz.

sound pressure toward a focused direction. This implies that one may not get a sufficiently precise estimate of loudness using beamforming when there are strong competing sources in other directions than the focused one. Furthermore, the dynamic range of a loudness mapping is limited due to sidelobes, which may depend on the shape of the array and on shadings applied during the beamforming calculation (Song, 2004). In the following, a few fundamental concepts needed are reviewed and the goal of the current investigation is specified.

A. Sidelobe effect in beamforming

In beamforming, contributions in the focused direction add up coherently, but other directions do not. On the other hand, the contributions from other directions cannot be completely removed and the artifacts thus generated are termed 'sidelobes'. Sidelobes can be illustrated by the array directivity pattern shown in Fig 1. A 42-channel wheel array was used to estimate this directivity pattern, and the mainlobe direction is 0° in this case. The upper panel displays the directivity pattern at 1 kHz, and the lower panel at 3.15 kHz. The width of the mainlobe at 1 kHz is greater compared to that at 3.15 kHz. The maximum sidelobe level (MSL) is around -10 dB at both frequencies. It is also noticeable that the shape of the sidelobes in the directivity pattern depends on frequency. If there was a competing sound sources in the direction of 25° at 3.15 kHz, its SPL would be reduced in the focused direction by approximately 11 dB. This indicates that sounds from other directions will still be influential when the beamformer is focused in the frontal direction.

The effects of sidelobes can be interpreted as ghost images in the sound pressure map derived. Ghost images are sources that are not present physically, but appear in a beamforming map due to leaks caused by the sidelobes of the microphone array employed. Fig. 2 demonstrates the effects of sidelobes in the generated sound pressure map and in the spectrum for the selected point. The selected point is marked as a star in the map. These simulations were obtained by assuming two monopole sources playing third-octave band noise and being placed 1 m away from a 42-channel wheel array. The two monopole sources were placed 0.6 m apart. The left source generated a noise centered at 1 kHz, and the right one a noise centered at 3.15 kHz. Delay-sum beamforming (Christensen and Hald, 2004; Johnson and Dudgeon, 1993) was employed in the simulation using Matlab.

It is obvious (see Fig. 2) that the left sound source appears to be much larger than the right one due to its lower center frequency, and that ghost images appear in

some places in the map (see the upper panel in Fig. 2). The dynamic range of the map was adjusted to be 10 dB. The map demonstrates that the results of beamforming calculations may be contaminated by the sidelobes of the array, and that the dynamic range of the map should be determined dependent on the array used. The effect of sidelobes can be also demonstrated in the FFT spectrum obtained in the focused direction (see the lower panel in Fig. 2). Even though in the focused direction there is only a 1-kHz noise source, the 3.15-kHz noise which is off to the side appears in the frequency spectrum as well. If the left source were auralized using beamforming, the influence of the competing source, i.e. the 3.15-kHz noise, should be audible as well. Thus, the loudness in the focused direction may be affected by competing sound sources in a given sound field. In order to properly auralize sound sources using beamforming, a strategy to simulate the effects of sidelobes generated by the processing might be needed to investigate the more general effect of secondary sources on perceived loudness. Based on such research, the criteria for designing a beamformer for estimating the loudness of sources more accurately might be specified.

B. Loudness assessment of simultaneous sound sources

If the loudness of a sound field is to be determined objectively, i.e. without performing subjective judgments, current methodology either employs a single microphone or a dummy head. Single microphone measurements provide loudness values by assuming measured sound fields to approximate either free-field field or diffuse field conditions (ISO 532, 1975). Binaural loudness can be obtained from dummy head measurements by combining the acoustic signals measured at the two ears of the artificial head (Sivonen, 2006). In addition, in the field of noise source identification, it is necessary to identify problematic sources, and to calculate the combined loudness of noise sources in a sound field. Song *et al.* (2006) suggested that the loudness of individual or combined sources may be obtained by measuring the sound field using a microphone array, and subsequently combining the loudness of the sources in terms of a 6-dB rule. The method created a possibility of assessing the loudness of a partial sound field.

When estimating the loudness of multiple sound sources, it is interesting to investigate whether there might be a loudness threshold for secondary sources below which they do not contribute to overall loudness any longer. The use of such thresholds might simplify calculations when integrating the loudness of individual sources. For example, sound sources below the loudness threshold determined might be ignored when estimating the combined loudness, and this may be a more accurate approach compared to a simple integration without such a threshold.

C. Just-Noticeable Level Difference

In Just-Noticeable Difference in Level (JNDL) experiments (Hanna *et al.*, 1986; Viemeister and Bacon, 1988; Zwicker and Fastl, 2006), subjects are asked to detect small differences in level in order to determine the smallest audible level difference for otherwise identical stimuli. Zwicker and Fastl (2006) performed a series of listening experiment on the JNDL using pure tones as well as broad-band noises. It was shown that the JNDL decreases as the sound pressure level (SPL) and the duration of a sound increases, and the JNDL for a broadband noise remained constant, at about 0.5 dB, for SPLs higher than 40 dB. Jesteadt *et al.* (1977); Wier *et al.* (1975) investigated intensity discrimination as a function of frequency and sensation level, and revealed no effect of signal frequency on the size of intensity discrimination. This implies that the change in discrimination with level may be represented by a single frequency.

In recent studies (Green, 1988), profile analysis is proposed to measure auditory intensity discrimination, which is relatively immune to changes in the overall level of the sounds and to time between intervals within a trial. The hypothesis of profile analysis is that the auditory system is able to detecting small intensity increments based upon analysis of spectral profiles, and thus the experiment involves simultaneous comparison within each interval. Green *et al.* (1983) revealed that profile analysis is not based entirely on local comparisons and in fact improves when more frequency components are provided, assuming the additional components are inserted to the edges of the spectral pattern. Dai and Green (1992) showed that the intensities of two stimuli with different frequencies are compared with higher precision when the two stimuli are presented simultaneously than when they are presented successively, and thus support the hypothesis, in which comparison between different frequencies effectively cancels the common noise in stimuli.

Jesteadt and Wier (1977) compared monaural and binaural discrimination of intensity and frequency using a two-interval forced-choice adaptive procedure (Levitt, 1971), and revealed that binaural difference limens are uniformly smaller than the monaural for both intensity and frequency discrimination at all frequencies of investigation. In another study (Stellmack *et al.*, 2004), the level effects in monaural intensity discrimination in a two-interval task was compared with discrimination of interaural intensity differences in a single-interval task. It is suggested that the interaural thresholds showed a small (about 2 dB) advantage over monoaural thresholds only in the broadband noise conditions, and the basic mechanisms of the level effects on intensity discrimination are common to monaural and interaural processing.

Assuming a power summation of individual noises, the loudness threshold for a secondary sound can be compared with the JNDL, if the same type of noises is used for both primary and secondary sound sources, and they are presented in the frontal direction. Moreover, assum-

ing small directional loudness variations, the comparison will still be valid if the sources only cover a small angular separation. One of the goals in this study is to investigate whether the loudness threshold determined in a multiple-source environment is different from a JNDL by comparing it with results from the literature. Since JNDL data using exactly the same type of stimuli were not available, general trends of the results from the current study will be compared with the literature.

D. Goals of the current investigation

1. To measure the threshold below which a secondary source does not contribute to overall loudness any longer, and to do this as a function of SPL and source separation angle. The type of stimuli shall be controlled in two ways. In one condition, two narrow-band noises having different center frequencies are used, and two incoherent noises with the same center frequency in another condition. This enables to investigate whether the outcome is affected by the frequency content of the two simultaneous noises.

2. To check the validity of the psychophysical method employed. An adaptive procedure and free magnitude estimation are used. In the adaptive procedure, the subjects are instructed to compare the loudness of two simultaneous noises with that of a single noise, and the procedure attempts to find a loudness match. This may cause some bias due to the expectation that the two simultaneous sounds should always be louder than the single one since they share the same primary sound in all comparisons. To eliminate this kind of bias, a direct scaling procedure in which one sound was assessed at a time was used for comparison.

II. GENERAL METHOD

A. Subjects

Ten normal hearing subjects (6 male, 4 female), between 22 and 28 years of age, completed the experiment. All of them were students at the University of Aalborg, Denmark. 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 389-1, 1998) at more than one frequency. None of the thresholds exceeded 25 dB hearing level. Participants were all paid for their participation. In a post-experimental questionnaire, none of them reported having any hearing disorders.

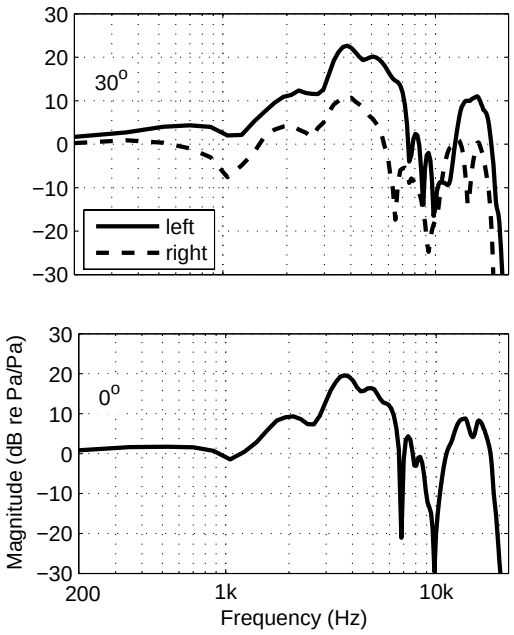


FIG. 3. Dummy-head HRTFs used in the experiments.

B. Apparatus

The experiment was carried out in a small listening room with sound-isolating walls, floors, and ceiling, which conforms with the ISO 8253-2 (1992) standard. Listeners were seated in a height-adjustable chair with a headrest. They were asked to fix their head position to the frontal direction during the experiments. Their head movement was checked through a camera positioned in the listening room. Two PC monitors, one in the control room and the other in the listening room, displayed the experimental procedure simultaneously with the help of a VGA splitter. A small loudspeaker played the sound that the subjects listened to back to the experimenter in order to check the stimuli and to monitor the listeners' behavior during the experiment.

A personal computer with a 16-bit sound card (RME DIGI96) was installed in a control room attached to the listening chamber and controlled the experimental procedure. The sound was produced with a sampling rate of 48-kHz and played through an electrostatic headphone (Sennheiser HE60) connected to an amplifier (Sennheiser HEV70) with a fixed volume control to assure constant gain. An external amplifier (Finalizer, t.c. Electronic) was installed between the headphone amplifier and the sound card to facilitate control of the playback level.

C. Stimuli

Third-octave band noises having center frequencies of 1 and 3.15 kHz were used in the investigation. Their center frequencies were chosen so that at equal level there is

reasonable loudness difference and no frequency masking between them. Each sound had a total duration of 1 s. A 20-ms ramp was applied in the beginning and at the end of the stimuli in order to avoid abrupt changes.

With the aim of generating the perception of two simultaneous sound sources in space, two loudspeakers spaced at 0° (i.e. frontal direction) and ±30° angular separations in free-field condition were simulated by convolving these signals with the head-related transfer functions (HRTFs) in the corresponding direction. The HRTFs used in this study were taken from measurements made on the artificial head VALDEMAR at the department of Acoustics, Aalborg University (Christensen and Møller, 2000). The measurement of the HRTFs was performed with a customized maximum length sequence (MLS) analyzer (Olesen *et al.*, 2000), and each HRTF consisted of 256 taps. Examples of the HRTFs are displayed in Fig. 3. Symmetry of the HRTFs was assumed by using only the HRTFs for the left hemisphere. For example, the HRTFs at 30° were used to simulate responses at −30° by swapping the transfer functions of the left and right ear, and the left-ear HRTF was used for the right ear in the frontal direction.

Headphone transfer functions (PTFs) were measured using a customized maximum length sequence (MLS) analyzer (Olesen *et al.*, 2000), and the artificial head was placed in a quite listening room during the measurements. Between measurements, the headphone was repositioned. Fig. 4 displays the overlay of five repetitive PTF measurements, and shows that the five PTFs are similar to each other in the frequency range of the investigation. It can be seen that the difference between repetitions becomes larger above 7 kHz. An average was taken from these five repetitive measurements and its inverse filter was calculated using fast deconvolution with regularization (Kirkeby *et al.*, 1998). The calculated inverse filter was then band-pass filtered between 25 and 8000 Hz in order to minimize the effect of the high-frequency peaks. These inverse PTFs were applied to the stimuli prior to the experiment.

A random signal of 1 s was generated in Matlab, and subsequently third octave filters were applied to this signal. The level of the output signals was then adjusted to the predefined ones and the inverse gain of the playback system applied afterward. The PTF of each ear was then applied to the generated signals for the 0° angular separation, and it was applied after the convolution with the HRTFs for the ±30° virtual loudspeaker locations. When the stimuli were generated to evaluate from the two offset positions, the location of the narrow-band noises was balanced in that the 1-kHz noise was generated at −30° in half of the conditions, and at 30° in the other half. The 3.15-kHz noise was placed at the opposite location. For the adaptive procedure (see II.D.1), the HRTFs were convolved with the narrow-band noises during the experiment using an FFT-based convolution, since the level of each narrow-band noise had to be changed according to the subject's response on the previous trial.

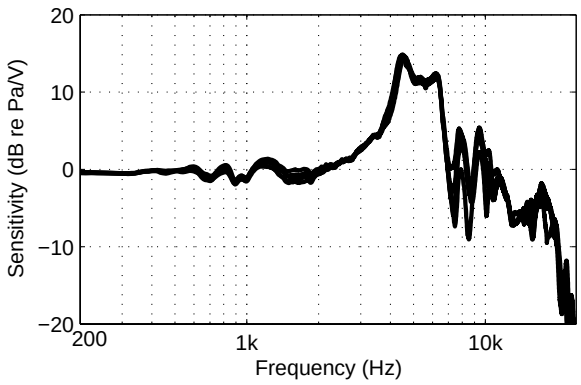


FIG. 4. Five headphone transfer functions at the left ear of the dummy head.

The output calibration was performed so as to produce the desired sound pressure level (SPL) at the blocked entrance of the ear canal of the dummy head as if it was measured in the free field. To this end, the sensitivity of the sound card and the external amplifier were measured. The measured sensitivities were combined with the PTFs to determine the output sensitivity of the entire sound playback system. The playback levels of the two narrow-band noises centered at 1 and 3.15 kHz having 60-dB SPL originating from two offset loudspeaker positions (±30°) were checked by comparing the SPLs measured using the dummy head with the simulated ones. The differences between the levels of measured and simulated signals were on the order of ±0.5 dB. The output of the external amplifier was checked before each session to make sure that the setup remained constant.

D. Procedure

Two types of psychophysical method, i.e. an adaptive procedure and free magnitude estimation, were employed in this study. The magnitude experiment consisted of two parts: two simultaneous noises having different center frequencies were presented in one part of the experiment (dual-frequency condition), and noises having the same center frequency in the other (single-frequency condition). The experiment based on an adaptive procedure had only the dual-frequency condition. In the dual-frequency condition, the adaptive procedure and free magnitude estimation were balanced in that half of the subjects started the adaptive procedure first, and the rest of them free magnitude estimation.

1. Adaptive procedure

An adaptive two-interval, two-alternative forced choice procedure with a one-up, two-down rule (Jesteadt, 1980; Levitt, 1971) was used in this experiment. The subjects

were asked to compare the loudness of a single sound with that of a dual sound on each trial. There was a 500-ms pause between the two intervals on each trial. The standard was a single sound, which contained only the primary sound (i.e. noise with a fixed level), and the comparison was a dual sound consisting of both the primary and the secondary sound (i.e. noise with a variable level). The level of the secondary sound varied in the adaptive track. The primary sound had a fixed SPL of either 45 or 60 dB, and the starting level of a secondary sound was the same as the primary. The level of the secondary sound in the next trial of the same adaptive track was adjusted according to the subjects' previous response. The initial step size was 10 dB, it was decreased to 3 dB after two reversals, i.e. changes in the direction of the adaptive track, and to 1 dB after four reversals. If the subjects judged the dual sound to be louder in two consecutive responses, this led to a level decrease of the secondary sound and one response to be softer led to an increase. This procedure made the dual sound converge toward the level at which it was judged louder than the single sound in 71 % of the trials. Subjects were asked to judge perceived loudness, not any other changes in sounds occurring during the experiment. In total, eight reversals were collected in each track and the last four reversals were averaged to calculate the loudness threshold for a secondary sound in each condition.

One block of trials consisted of 8 tracks (2 center frequencies $\times$ 2 levels $\times$ 2 source separation angles), and they were divided into two parts of 10 minutes duration with a 30-s break in between. Each block of the experiment corresponded to one repetition of all experimental conditions. The subjects were instructed to stay in the listening room during a block of trials. Both the order of tracks and the succession of trials in each track were randomized separately for each subject. One session consisted of 4 blocks, and each subject completed one session. The subjects took a 5-minute break after each block was finished. In the beginning of the experiments, a practice block with only two selected tracks was completed prior to the actual experiment.

2. Free magnitude estimation

Magnitude estimation is a wide-spread methodology to scale psychoacoustic attributes, such as loudness, auditory sharpness, and the likes (Stevens, 1955). It is extremely efficient and allows quick acquisition of responses using large amounts of stimuli. In a magnitude estimation experiment, listeners are required to provide direct numerical responses that correspond to their perception of the degree to which an attribute is present in a set of stimuli. Stevens proposed two methods of magnitude estimation. In one version, listeners are presented with a standard stimulus, which by default has an assigned number, e.g. 10. Subsequently they are presented by a test stimulus and asked to provide a number relative to the standard. The other version is called free (Stevens, 1975), or absolute (Gescheider, 1997) magnitude estimation and no standard is presented or defined in the experiment. In this method, listeners assign any numbers they like to their loudness sensations in proportion to their magnitude.

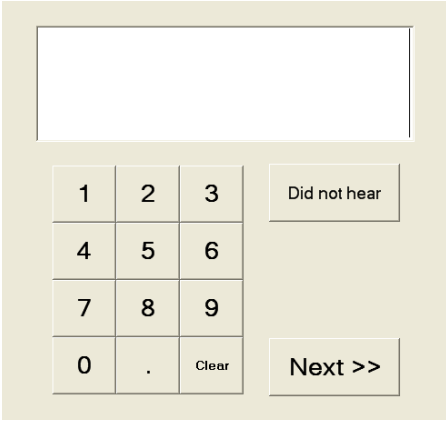


FIG. 5. The user-interface for the free magnitude estimation experiments.

Zwislocki and Goodman (1980) found that listeners judged a particular stimulus independent of other stimuli used in the experiment. Furthermore, Gescheider (1997) showed that the use of a standard in magnitude estimation could result in biasing effects. For example, in the case of the present experiment, if a narrow-band noise with a specific center frequency is used as a standard, listeners might utilize different strategies to judge the loudness of a sound with the same center frequency than others. Therefore, free magnitude estimation was selected to collect the present subjective loudness judgments.

The instructions were modelled on suggestions made by Stevens (Stevens, 1975):

You will be presented with a series of stimuli in irregular order. Your task is to tell how loud they seem by assigning numbers to them. Call the first stimulus any number that seems appropriate to you. Then assign successive numbers in such a way that they reflect your subjective impression. You can use any positive number that you want to. There is no limit to the range of numbers that you may use. You may use whole numbers or decimals. Try to make each number match the loudness as you perceive it.

Fig. 5 shows a screen shot of the user-interface used for the magnitude estimation experiments. The size of the input box was adjusted so that listeners could focus their attention on the number they typed in. The numeric key pad was provided at the lower left corner of the input box, and it contained a "Clear" button to allow the listeners to correct their inputs. In addition, a "Did not hear" button was available to be pressed when they did not hear the

stimuli. After listeners gave their response on a trial, they were allowed to proceed to the next one by pressing the "Next" button. Participants were instructed to inform the experimenter of any typing errors they made during the procedure. The typing errors they could identify were corrected immediately after each block of the experiment.

Two listening experiments were performed using this procedure. One is the dual-frequency condition where two narrow-band noises were played at the same time either in the frontal direction or in the two offset directions. The other is the single-frequency condition, in which two incoherent noises having the same center frequency were presented in the two spatial configurations. The SPLs of the stimuli were: mute, 25, 35, 45, 50, 55, 60 dB. The stimuli were 1 and 3.15-kHz narrow-band noises, and each noise had the predefined levels in each experimental condition.

In accord with a fully crossed matrix design, all levels of the 1-kHz noise were combined with all levels of the 3.15-kHz noise in the dual-frequency condition. This resulted in 96 stimuli, which excludes combining two mute conditions from 7 levels of the 1-kHz noise $\times$ 7 levels of the 3.15-kHz noise $\times$ 2 spatial configurations (0° and $\pm$30°). In the single-frequency condition, the combinations of all levels formed 54 stimuli for each center frequency by avoiding duplicate stimuli conditions with the same level combination, e.g. a level combination of 35 dB and 60 dB is the same as that of 60 dB and 35 dB. Two repetitions of 54 stimuli were presented in a block of the experiment. When the stimuli were played through the two offset directions, the location of narrow-band noises was balanced in that a noise with higher SPL was generated from the left position on half of the trials, and with lower SPL in the other half. In both conditions, the order of trials within a block of the experiment was randomized separately for each subject.

A stimulus was presented on each trial, and there was a 1-s pause between trials after the subjects provided their response. Either all 96 (dual) or 108 (single) stimuli were presented in a block of the experiment. One session consisted of 4 blocks in the dual-frequency condition and 2 blocks in the single-frequency condition, corresponding to 4 repetitions. The subjects took a 30-s break between blocks, and a 5-minute break after two blocks were finished. In the beginning of the experiment, one practice block containing only eight stimuli was completed prior to the data collection proper. The practice block consisted of the stimuli with a wide range of SPLs to help the listeners to establish their own subjective scale prior to the experiment. The experimenter stayed inside the listening room during the practice block to instruct the listeners if they had questions about the procedure. In total, the experiment took 1.5 hour in the dual-frequency condition and 45 minutes in the single-frequency condition during which each subject accumulated 4 repetitions.

TABLE I. Loudness threshold of a secondary sound source using the adaptive procedure. (unit: dB SPL)

variable noise	level of fixed noise	angle	loudness threshold
1 kHz	45	0°	30.8
1 kHz	45	30°	30.3
1 kHz	60	0°	54.1
1 kHz	60	30°	46.6
3.15 kHz	45	0°	19.5
3.15 kHz	45	30°	19.4
3.15 kHz	60	0°	36.5
3.15 kHz	60	30°	34.2

III. RESULTS

A. Adaptive procedure

If the subjects compared the loudness of two simultaneous sounds (the comparison) with that of a single sound (the standard), the level of the variable component of the comparison should eventually decrease during the experiment since the two intervals shared the same primary sound (i.e. noise with a fixed level). Unfortunately, some subjects perceived the loudness of two simultaneous sounds to be softer than a single one (i.e. playing only the primary sound) from the beginning, and this prevented the adaptive procedure from converging. When this happened, the maximum level of the variable noise, which is determined by the dynamic range of the playback system, was recorded as the threshold. This primarily happened when the 1-kHz noise was the variable component, and may be due to the fact that some subjects judged the annoyance of the stimuli rather than their loudness even though they were instructed to judge loudness. Increasing the level of the 1-kHz noise results in decreasing the overall annoyance of the two simultaneous noises because it will decrease the sharpness of the stimuli and make the sound more comfortable to listen. This supports the idea that the perception of loudness and annoyance are sometimes confounded (Ellermeier *et al.*, 2007), and we concluded that magnitude estimation may be a better psychophysical method to measure the loudness threshold of a secondary sound source since the procedure does not require direct comparisons between two simultaneous noises and a single noise sharing the same primary sound.

The loudness thresholds of a secondary sound source averaged across all ten subjects are listed in Table I. They were calculated from the loudness threshold of a secondary sound source when the component with a fixed level served as a standard. Since the same primary component with a fixed SPL was presented both in the comparison and the standard and the two simultaneous noises

are separated in critical bands, loudness thresholds close to the threshold of hearing were expected when a dual sound, i.e. a comparison stimulus, was directly compared with a single sound, i.e. the standard.

Unexpectedly, the thresholds found in this experiment were much higher than the threshold of hearing (see Table I). The results does not seem to depend on the angle of incident, but on the level of the fixed noise and on center frequency. If there was no interaction between the simultaneous sources, then the loudness threshold should not depend on the level of the fixed noise since the two narrow-band noises were sufficiently separated in the specific loudness spectrum. This indicates that the subjects may have ignored a secondary sound in their loudness judgments even though it was clearly audible.

B. Free magnitude estimation

1. Dual-frequency condition

In contrast to the adaptive procedure, the magnitude estimation does not involve a direct comparison between two simultaneous noises and a single noise. It is assumed that the magnitude estimation is better suited for determining the contribution of a secondary sound source to overall loudness.

Fig. 6 and Fig. 7 show the geometric means in the dual-frequency condition. The upper panel shows the results as a function of the level of the 1-kHz noise, and the lower panel as a function of the level of the 3.15-kHz noise. Note that these are the same results presented in a different display format. Fig. 6 shows the results when the stimuli were generated from the frontal direction, and Fig. 7 from the two offset positions.

In general, loudness judgments increase when the level of one of the components increases, and loudness judgments tend to asymptote as the level of the secondary sound displayed on the x-axis is decreased. It can be seen that loudness decreases faster when the level of the 3.15-kHz noise is changed (compare between the upper and lower panel). The higher the level of the fixed component (see the legend of the figures) the sooner do the responses asymptote. Furthermore, the results from the offset positions (see Fig. 7) reveal that there seems no effect of angular separation on the determination of a loudness threshold since the two figures are very similar.

In the experiment, the subjects were allowed to assign any positive number corresponding to the loudness they perceived, and there was no limit to the range of numbers. This resulted in a large discrepancy in the range of numbers each subject assigned. For example, a subject used the range from 0.5 to 6.2 whereas another subject employed a number range from 2 to 90. This increases the probability of failing to find significant differences between average data, and leads to large loudness thresholds. To avoid this situation, z-scores (see Eq. 1) were

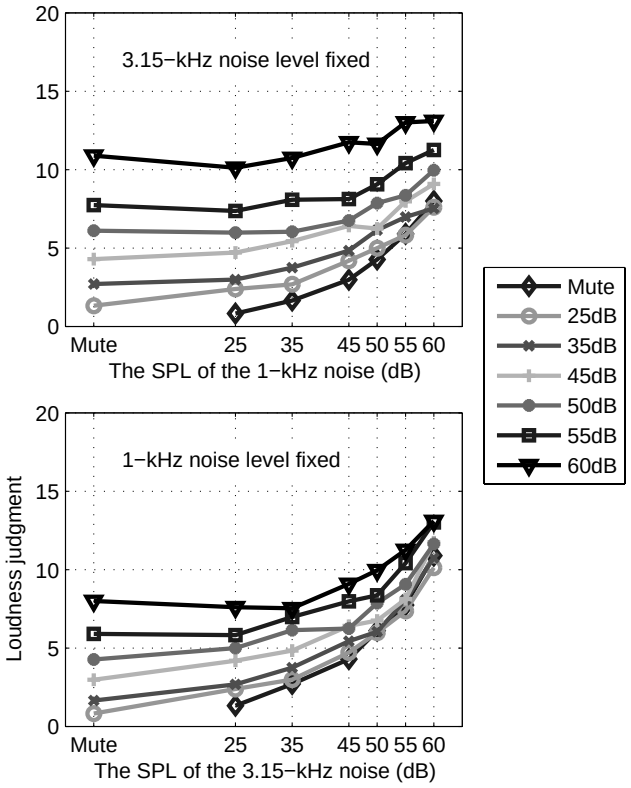


FIG. 6. Loudness judgments in the dual-frequency condition. The upper panel shows the results as a function of the SPL of the 1-kHz noise, and the lower as a function of the level of the 3.15-kHz noise. The stimuli were presented in the frontal direction, 0°. The legend indicates the level of the fixed noise.

calculated from the subjective data.

$$z = \frac{x - \mu_x}{\sigma_x} \tag{1}$$

where x are the subjective data, μ_x is the mean, and σ_x is the standard deviation. The z-score of an item indicates how far and in what direction, the item deviates from its mean, expresses in units of its standard deviation. After the z-score transformation, the transformed data will have a mean of zero and a standard deviation of one.

The means in the mute condition (indicated as "Mute" on the x-axis) have to be compared with those in the other conditions to determine the level at which a significant contribution is made by the secondary source. This requires simultaneous comparisons between the control ("Mute") condition, and the experimental (the other) conditions. Dunnett's test (Geoffrey Keppel, 2004) was utilized to check whether the mean of the mute condition is different from that of the other conditions, and it was applied to the z-score transformed data. If all p-values of the Dunnett's test are larger than 5%, the loudness threshold should be higher than 60 dB, and if all p-values are smaller than 5%, the threshold has to be lower than 25 dB (marked as ↓). Otherwise, the first

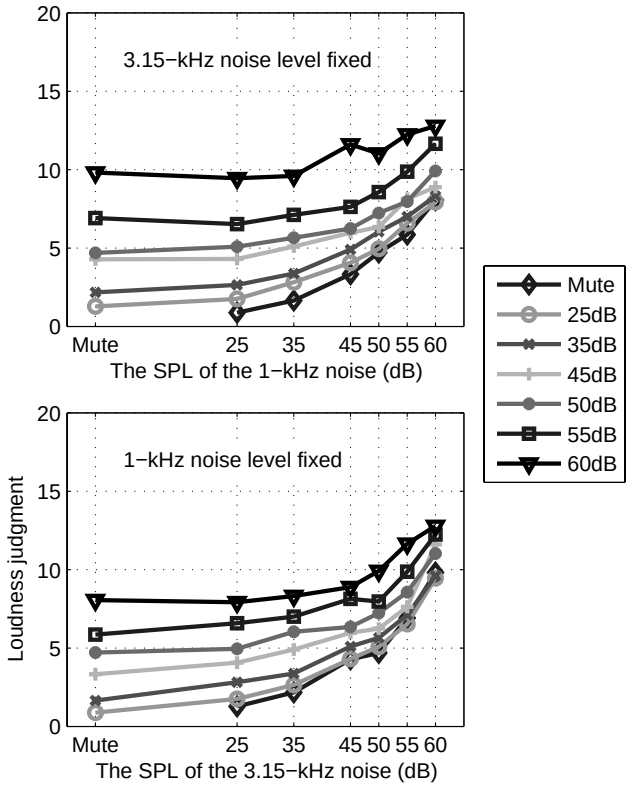


FIG. 7. Loudness judgments of the dual-frequency condition. The upper panel shows the results as a function of the SPL of the 1-kHz noise, and the lower as a function of the level of the 3.15-kHz noise. The stimuli were presented in the offset positions, ±30°. The legend indicates the level of the fixed noise.

data point with a p-value larger than 5% is treated as the loudness threshold of the secondary sound.

Table II and Table III summarize the results of the Dunnett tests performed. In general, the calculated loudness thresholds are quite large, and indicate that the subjects could hear the secondary sound clearly but it did not contribute to overall loudness. This proves that there is a considerable loudness dominance of the primary sound in multiple-sound conditions. The higher the level of the fixed component was the higher the loudness threshold. There are slight differences between the two spatial configurations (0° and ±30°) observed in some conditions, though most of the conditions had quite similar thresholds. Higher loudness thresholds were obtained when the level of the 1-kHz noise was varied (compare Tables II and III). The loudness thresholds could not be determined at the lower levels of the fixed noise, and more steps of SPL at low levels are required to determine the loudness thresholds especially when the level of the 3.15-kHz is varied.

The results of the adaptive procedure are displayed in parentheses in Table II and III. Similar loudness thresholds were obtained in both type of experiments. The advantage of the adaptive procedure may be the fact that the loudness thresholds are always obtained in all experimental condition whereas in the free magnitude estimation the results of some conditions could not be obtained. On the other hand, the direct comparison between a single sound and a dual sound may lead some subjects to judge annoyance rather than loudness due to that the perception of annoyance and loudness is confounded.

TABLE II. Loudness threshold of a secondary sound for the dual-frequency condition when the level of the 1-kHz noise was varied. The results from the adaptive procedure are displayed in parentheses. ↓ indicates that the threshold is below 25 dB. (unit: dB SPL)

level of the 3.15-kHz noise	Spatial configuration	
	0°	±30°
25	↓	↓
35	25	25
45	35 (30.8)	25 (30.3)
50	45	25
55	45	45
60	45 (54.1)	50 (46.6)

TABLE III. Loudness threshold of a secondary sound for the dual-frequency condition when the level of the 3.15-kHz noise was varied. The results from the adaptive procedure are displayed in parentheses. ↓ indicates that the threshold is below 25 dB. (unit: dB)

level of the 1-kHz noise	Spatial configuration	
	0°	±30°
25	↓	↓
35	↓	↓
45	↓ (19.5)	↓ (19.4)
50	↓	45
55	35	35
60	45 (36.5)	50 (34.2)

2. Single-frequency condition

In the single-frequency condition, the stimuli consisted of narrow-band noises having the same center frequency, and they were presented either in the frontal direction or in the two offset directions. The perception of spatial configuration in this condition was different from that

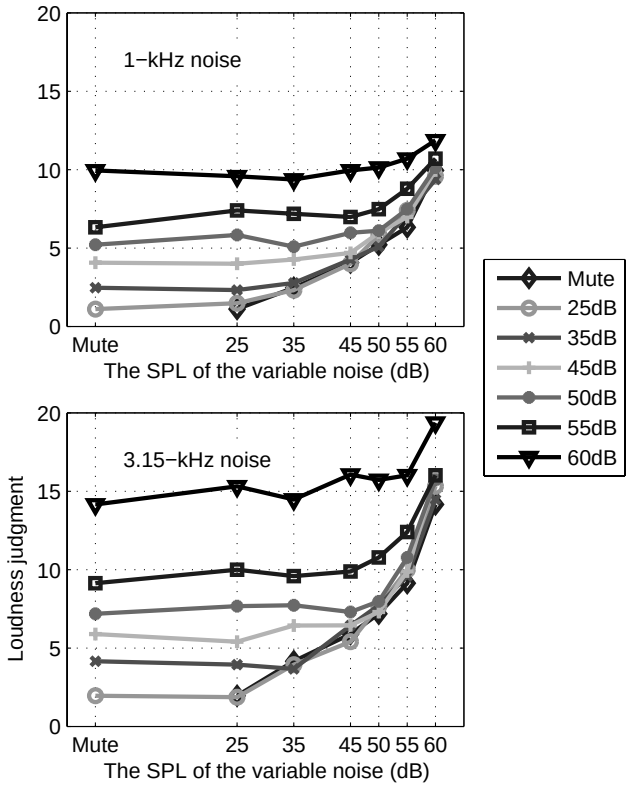


FIG. 8. Loudness judgments of the single-frequency condition. The upper panel shows the results of the 1-kHz noise, and the lower that of the 3.15-kHz noise. The stimuli were presented in the frontal direction, 0°. The legend indicates the level of the fixed noise.

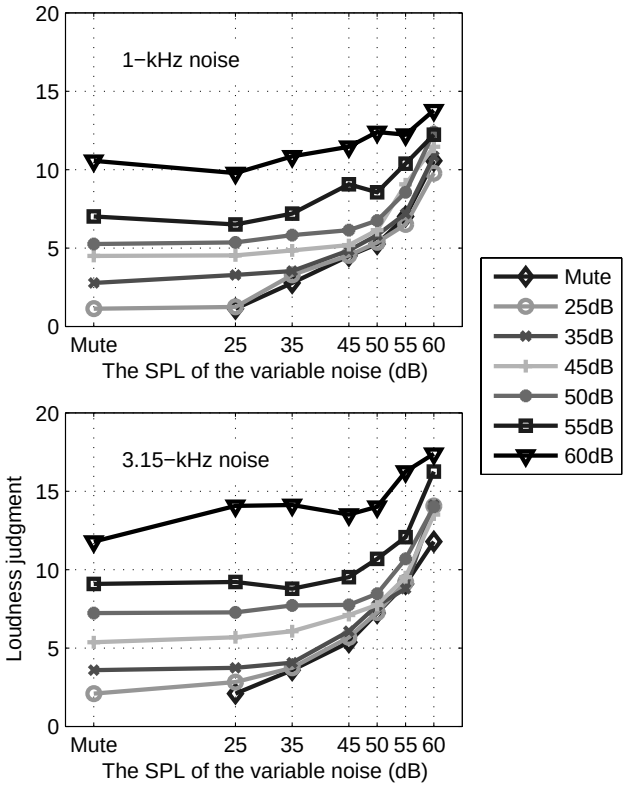


FIG. 9. Loudness judgments of the single-frequency condition. The upper panel shows the results of the 1-kHz noise, and the lower that of the 3.15-kHz noise. The stimuli were presented in the offset positions, ±30°. The legend indicates the level of the fixed noise.

of the dual-frequency condition in that the subject perceived a single sound event when the noises were presented in the frontal direction, and they perceived the location of two noises more distinct in the dual-frequency condition than in the single-frequency one when the stimuli were played from the two offset positions.

The results of this condition are displayed in Fig. 8 and Fig. 9 in the same format as the previous section. The upper panel illustrates the results for the 1-kHz noise, and the lower for the 3.15-kHz. The same behavior may be observed that the listeners' loudness judgments converges when the level of the secondary sound is decreased. With the higher level of variable components the subjects produced similar loudness judgments between different levels of fixed components. Higher loudness judgments were obtained in the 3.15-kHz noise compared to the 1-kHz noise. Loudness judgments in the frontal direction seem to converge faster than those in the offset positions. The results of the 3.15-kHz noise spread across a wider range of loudness compared to that of the 1-kHz noise. This result is contrary to Stevens' power law (Stevens, 1957) according to which the loudness of a given sound is proportional to its RMS pressure raised to the power 0.6.

Dunnett's test was performed on the z-score transformed data in the same way as described in the previous section, and the results are summarized in Table IV for the 1-kHz noise condition and Table V for the 3.15-kHz noise condition. In general, the loudness thresholds obtained in the single-frequency condition were higher than those of the dual-frequency condition, and the thresholds increased by increasing the level of the fixed noise. Since the two noises had the same center frequency, one may notice that the subjects could distinguish the maximum level difference of approximately 3 dB, the fixed and variable noise having the same SPL, according to the power summation rule. There are slight differences in loudness threshold between the two spatial configurations (0° and ±30°) in that the loudness threshold in the frontal direction is slightly higher than that of the offset positions. This may reveal that the spatial separation of the two components helped the subjects to focus on the loudness changes in each component of the simultaneous sounds.

TABLE IV. Loudness threshold of a secondary sound for the single-frequency condition when the 1-kHz was played. (unit: dB)

level of the fixed noise	Spatial configuration	
	0°	±30°
25	25	25
35	35	25
45	45	35
50	45	35
55	50	35
60	55	45

TABLE V. Loudness threshold of a secondary sound for the single-frequency condition when the 3.15-kHz was played. ↓ indicates that the threshold is below 25 dB. (unit: dB)

level of the fixed noise	Spatial configuration	
	0°	±30°
25	↓	25
35	35	35
45	50	35
50	35	45
55	25	50
60	55	↓

IV. DISCUSSION

One of the goals of this study was to provide a guideline for designing a beamforming array that may be better suited for estimating the loudness of sound sources in a sound field. This requires to investigate the relationship between maximum sidelobe levels (MSLs) and the findings of the present study. The loudness thresholds in the dual-frequency condition were lower than those obtained in the single-frequency condition. This reveals that competing sources having different center frequencies (dual-frequency condition) may influence the estimation of loudness more than source having the same center frequency (single-frequency condition). For the single-frequency condition, the minimum loudness threshold was approximately 10 to 15 dB lower than the level of the primary sound. For the dual-frequency condition, the minimum loudness threshold was about 25 dB lower than the level of the primary sound. These results indicate that a microphone array needs to be designed to fulfill the requirements of the dynamic range, otherwise the influence of sidelobes on loudness estimation may be inevitable. The dynamic-range requirements are dependent on whether sound sources in a given sound field share similar frequency characteristics.

In an earlier study (Song *et al.*, 2006), a proposal was made to calculate the loudness of simultaneous sounds using beamforming, and this proposal may be extended based on the findings of the present study. In the algorithm (Song *et al.*, 2006), the 1/3-octave spectra of simultaneous sources were calculated using beamforming, combined in terms of a 6-dB loudness summation rule, and finally the loudness of the combined spectrum was obtained according to ISO 532 (1975). The present results suggest that a loudness threshold should be applied when combining the 1/3-octave spectra of sources. Since the dual-frequency condition resulted in very low loudness thresholds, the loudness thresholds may only be applied within the same 1/3-octave band. That is, for each 1/3 octave band, secondary sources having 10 to 15 dB lower SPL than the primary source should be ignored when calculating the combined spectrum.

The JNDLs of the narrow-band noises used in the current investigation can be derived by performing a power summation (3-dB summation) of simultaneous noises in the single-frequency condition, and the calculated JNDLs are displayed as a function of SPL in Fig. 10. It can be seen that the JNDLs decreases with increasing SPL, and agrees with results from the literature (e.g. Zwicker and Fastl, 2006). Ozimek and Zwislocki (1996) investigated the effect of frequency on JNDL. The same results were observed in that the effect of frequency on JNDLs was larger at low SPL, and it decreased as SPLs increased. This also agrees with the outcome suggested by Johnson *et al.* (1993), according to which JNDLs are coupled directly to loudness, since the narrow-band noises with different center frequency resulted in different loudness values. Furthermore, the JNDLs were affected by the angular separation between sources in such a way that the subjects were able to distinguish level differences better when the two sound sources were spatially separated. It may be due to the fact that the primary sound had the same SPL both in a single sound condition and the corresponding dual sound conditions, and the level difference was always introduced in the secondary source, which was spatially separated from the primary one. In this way, the spatial separation of sources might help the subjects to notice the level difference caused by a secondary sound when judging overall loudness.

V. CONCLUSION

(1) The threshold below which a secondary sound source does not contribute to overall loudness was investigated in a series of listening experiments. A 1-kHz and 3.15-kHz narrow-band noise were used as stimuli and they were auralized binaurally using dummy-head HRTFs. The results of a dual-frequency condition in which the secondary sound has different frequency content than the primary one, showed that the secondary sound contribute to a far lesser extent than expected. This may be

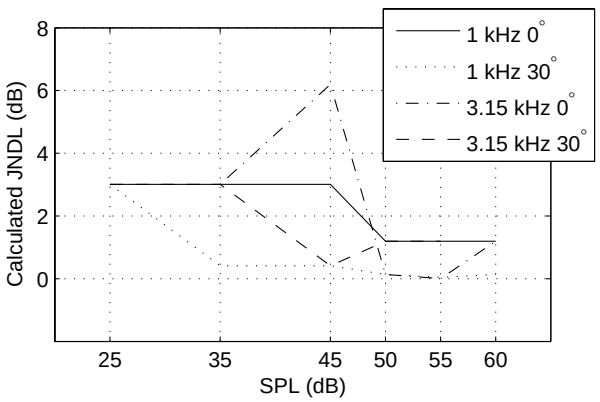


FIG. 10. Calculated JNDL from the loudness thresholds of the single-frequency condition.

explained by the concept of loudness dominance, meaning that subjects focus on the loudest component and ignore the others in a multiple-sound condition. The loudness thresholds obtained in the single-frequency condition were higher than in the dual-frequency condition. But, the actual loudness contribution caused by a secondary sound is much smaller compared to the dual-frequency condition due to summation across critical bands.

(2) The two experimental procedures, i.e. an adaptive procedure and free magnitude estimation, produced similar loudness thresholds for a secondary sound. With the adaptive procedure, however, some subjects perceived two simultaneous sounds to be softer than a single sound from the beginning, and this prevented the adaptive track to converge on a "threshold" value. This happened when the 1-kHz component was variable noise, and may be due to those subjects judging annoyance rather than loudness even though they were instructed to judge loudness. Increasing the level of the 1-kHz noise decreases the overall annoyance of the two simultaneous noises because it will reduce sharpness and thereby make the sounds more comfortable to listen. It appears that loudness and annoyance were confounded for some subjects.

(3) The influence of spatial source separation (0° versus ±30°) was more obvious in the single-frequency condition than in the dual-frequency condition. It may be that in the dual-frequency condition the subjects focused more on the difference in frequency of the two noises than on the spatial configuration.

(4) The outcome of the listening experiments can be used for designing a microphone array that is more suitable for estimating the loudness of target sources in the presence of competing ones in that the sidelobes of the array should be limited according to the findings of this study. The current investigation suggests that a microphone array with approximately 10 to 15 dB maximum sidelobe level (MSL) should be designed for assessing the loudness of sources. Furthermore, sound sources with sound pressure levels lower than the loudness thresholds obtained in this study, e.g. 10 to 15 dB, do not have to be integrated when calculating the loudness of simultaneous sounds.

(5) The results of the single-frequency condition were converted to JNDLs by assuming a power summation of simultaneous noises, and the results agree with traditional JNDL findings concerning level and frequency dependence. This indicates that the procedure employed in this study is suitable to estimate loudness thresholds, if the levels of the component noises are selected properly.

Acknowledgments

This research was carried out as part of the "Centerkontrakt on Sound Quality" which establishes participation in and funding of the "Sound Quality Research Unit" (SQRU) at Aalborg University. The participating companies are Bang & Olufsen, Brüel & Kjær, and Delta Acoustics & Vibration. Further financial support comes from the Ministry for Science, Technology, and Development (VTU), and from the Danish Research Council for Technology and Production (FTP).

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

Christensen, J. J. and Hald, J. (**2002**). "A class of optimal broadband phased array geometries designed for easy construction", in *JSAE Annual Congress*, preprint 5335 (Yokohama, Japan).

Christensen, J. J. and Hald, J. (**2004**). *Beamforming*, Technical Review number 1 (Brüel & Kjær, Nærum, Denmark).

Dai, H. P. and Green, D. M. (**1992**). "Auditory intensity perception successive versus simultaneous, across-channel discriminations", J. Acoust. Soc. Am. **91**, 2845–2854.

Ellermeier, W., Zeitler, A., Zimmer, K., and Fastl, H. (**2007**). "The role of source identifiability in loudness and annoyance judgments", submitted for publication .

Gallaudet, T. C. and de Moustier, C. P. (**2000**). "On optimal shading for arrays of irregularly-spaced or noncoplanar elements", IEEE Journal of oceanic engineering **25**, 553–567.

Geoffrey Keppel, T. D. W. (**2004**). *Design and Analysis : A Researcher's Handbook* (Prentice Hall).

Gescheider, G. A. (**1997**). *Psychophysics: The Fundamentals* (Lawrence Erlbaum Associates, New Jersey, USA).

Green, D. M. (**1988**). *Profile Analysis: Auditory Intensity Discrimination* (Oxford U.P., Oxford, England).

Green, D. M., Kidd, G., and Picardi, M. C. (**1983**). "Successive versus simultaneous comparison in auditory intensity discrimination", J. Acoust. Soc. Am. **73**, 639–643.

Hald, J. (**1989**). *STSF – a unique technique for scan-based Near-field Acoustic Holography without restrictions on coherence*, Technical Review number 1 (Brüel & Kjær, Nærum, Denmark).

Hald, J. (**2005**). "An Integrated NAH/Beamforming Solution for Efficient Broad-Band Noise Source Location", in *SAE Noise and Vibration Conference and Exhibition*, preprint 2537 (Grand Traverse, MI, USA).

Hanna, T., Vongierke, S. M., and Green, D. M. (**1986**). "Detection and intensity discrimination of a sinusoid", J. Acoust. Soc. Am. **80**, 1335–1340.

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

ISO 532 (**1975**). "Acoustics – method for calculating loudness level", ISO, Geneva, Switzerland .

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

Jesteadt, W. (**1980**). "An adaptive procedure for subjective judgments.", Perception & Psychophysics **28**, 85–88.

Jesteadt, W. and Wier, C. C. (**1977**). "Comparison of monaural and binaural discrimination of intensity and frequency", J. Acoust. Soc. Am. **61**, 1599–1603.

Jesteadt, W., Wier, C. C., and Green, D. M. (**1977**). "Intensity discrimination as a function of frequency and sensation level", J. Acoust. Soc. Am. **61**, 169–177.

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

Johnson, J. H., Turner, C. W., Zwislocki, J. J., and Margolis, R. H. (**1993**). "Just noticeable differences for intensity and their relation to loudness", J. Acoust. Soc. Am. **93**, 983–991.

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

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics", J. Acoust. Soc. Am. **49**, 467–477.

Li, Z. (**2005**). "The capture and recreation of 3d auditory scenes", Ph.D. thesis, University of Maryland.

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.

Olesen, S. K., Plogsties, J., Minnaar, P., Christensen, F., and Møller, H. (**2000**). "An improved MLS measurement system for acquiring room impulse responses", in *Proceedings of NORSIG 2000, IEEE Nordic Signal Processing Symposium*, pp. 117–120 (Kolmården, Sweden).

Ozimek, E. and Zwislocki, J. J. (**1996**). "Relationships of intensity discrimination to sensation and loudness levels: Dependence on sound frequency", J. Acoust. Soc. Am. **100**, 3304–3320.

Sivonen, V. (**2006**). "Directional loudness perception - the effect of sound incidence angle on loudness and the underlying binaural summation", Ph.D. thesis, Aalborg University.

Song, W. (**2004**). "Sound quality metrics mapping using beamforming", in *Internoise*, preprint 271 (Prague, Czech Republic).

Song, W., Ellermeier, W., and Minnaar, P. (**2006**). "Loudness estimation of simultaneous sources using beamforming", in *JSAE Annual Congress*, preprint 404 (Yokohama, Japan).

Stellmack, M. A., Viemeister, N. F., and Byrne, A. J. (**2004**). "Monaural and interaural intensity discrimination: Level effects and the "binaural advantage"", J. Acoust. Soc. Am. **116**, 1149–1159.

Stevens, S. S. (**1955**). "The Measurement of Loudness", J. Acoust. Soc. Am. **27**, 815–829.

Stevens, S. S. (**1957**). "Concerning the form of the loudness function", J. Acoust. Soc. Am. **29**, 603–606.

Stevens, S. S. (**1975**). *Psychophysics: Introduction to its perceptual, neural and social prospects* (Wiley, New York, USA).

Viemeister, N. F. and Bacon, S. P. (**1988**). "Intensity discrimination, increment detection, and magnitude estimation for 1-khz tones", J. Acoust. Soc. Am. **84**, 172–178.

Wier, C. C., Jesteadt, W., and Green, D. M. (**1975**). "Pure-tone intensity and frequency discrimination as a function of frequency and sensation level", in *The 90th Meeting of the Acoustical Society of America.*

Zwicker, E. and Fastl, H. (**2006**). *Psychoacoustics : Facts and Models* (Springer, Berlin, Germany).

Zwislocki, J. J. and Goodman, D. A. (**1980**). "Absolute scaling of sensory magnitudes - A validation", Perception & Psychophysics **28**, 28–38.

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:

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

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

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

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:

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

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

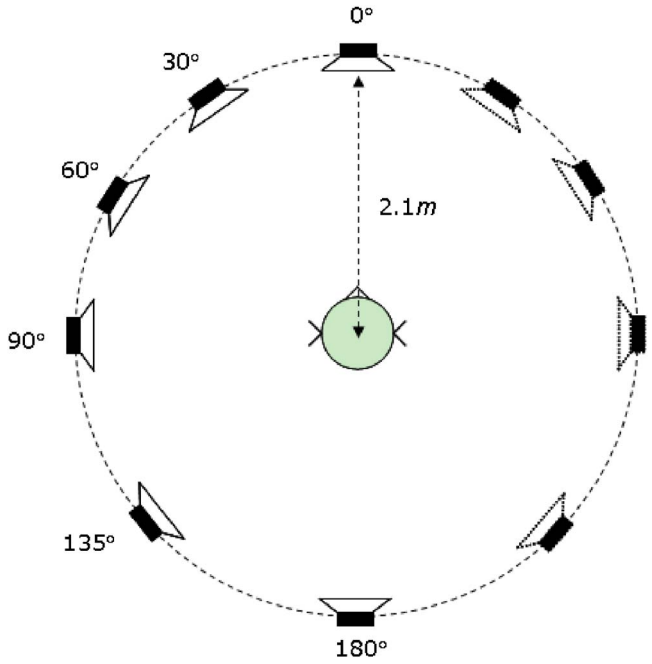


FIG. 1. (Color online) The loudspeaker setup in the anechoic chamber.

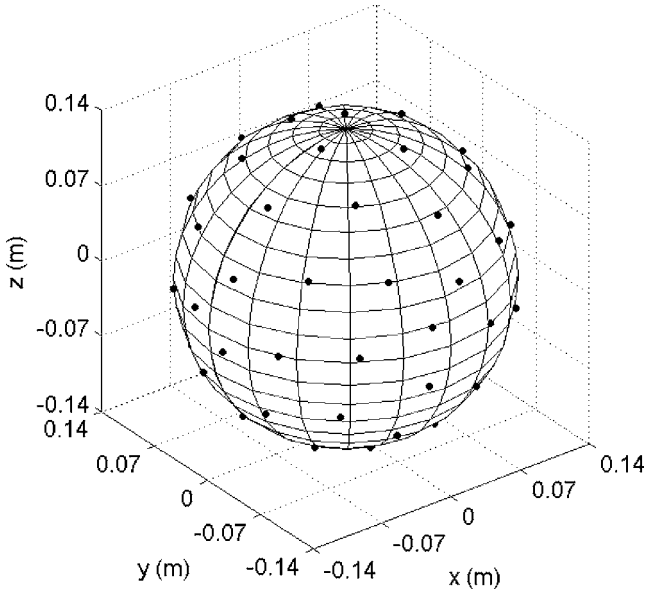


FIG. 2. The array consisting of 64 microphones placed on the hard surface of a sphere having a 14-cm radius. The dots on the sphere indicate the microphone positions.

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).

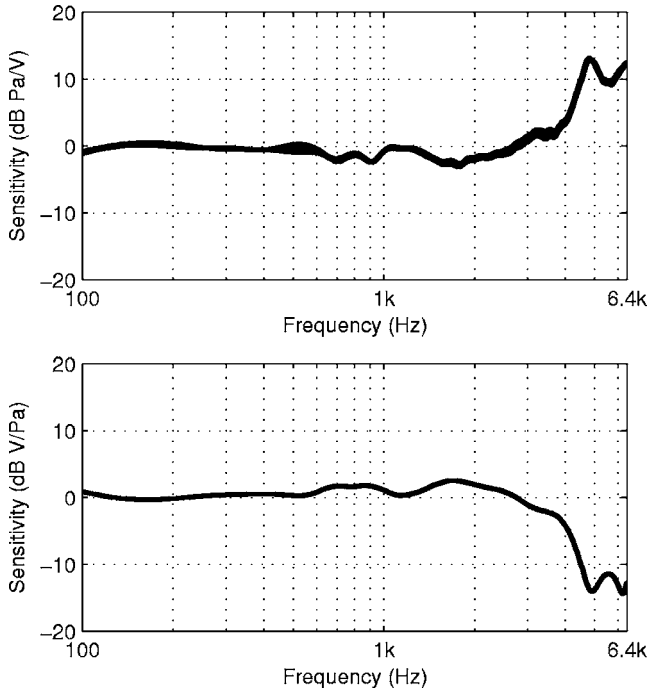


FIG. 3. Five headphone transfer functions (upper panel) measured at the left ear of the dummy head and the inverse filter derived (lower panel).

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.

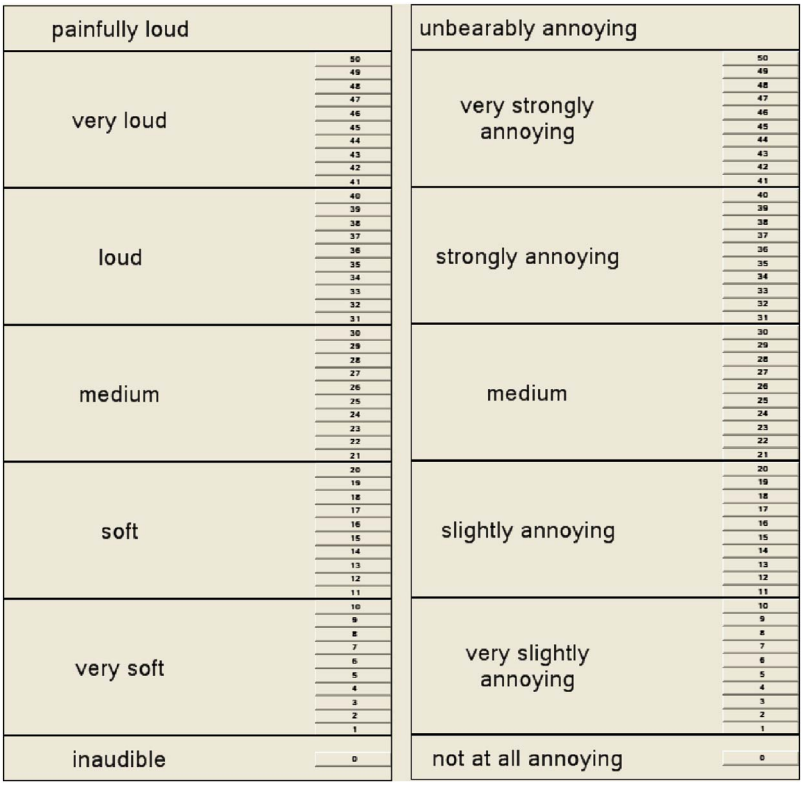


FIG. 4. (Color online) Category subdivision scales for loudness (left) and annoyance (right).

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

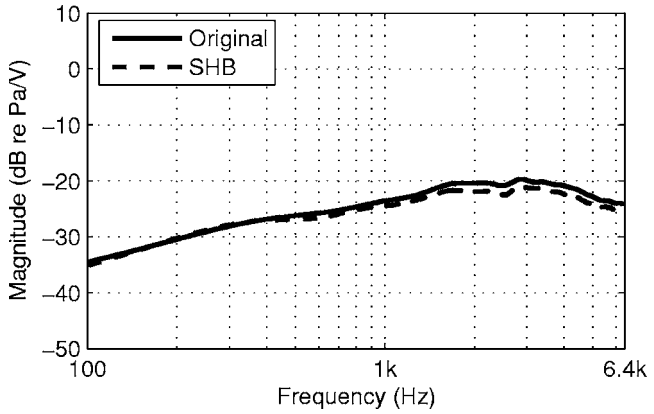


FIG. 5. Free-field loudspeaker response (30°): Measured (solid line) and synthesized (dashed line) using SHB.

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

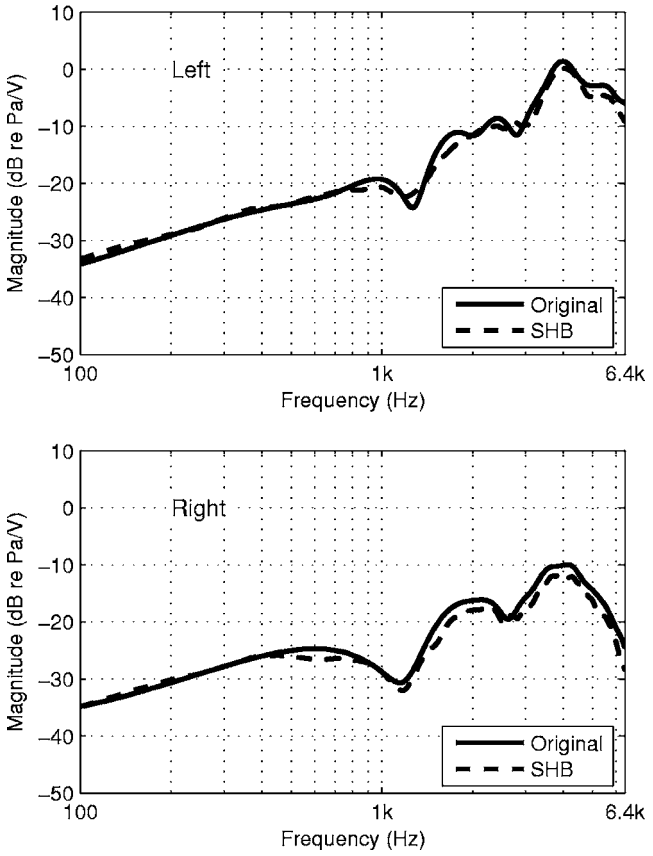


FIG. 6. Loudspeaker response (30°) at both ears: Measured (solid line) and synthesized (dashed line) using SHB.

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-

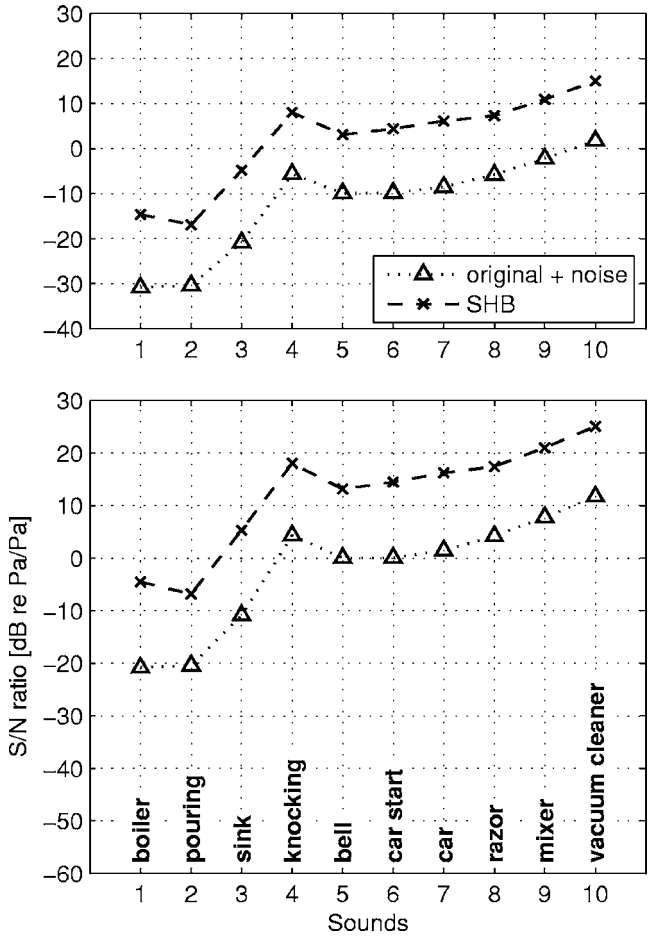


FIG. 7. Monophonic S/N ratio of dummy head (original+noise) and SHB measurements in the low (lower panel) and high (upper panel) background noise conditions.

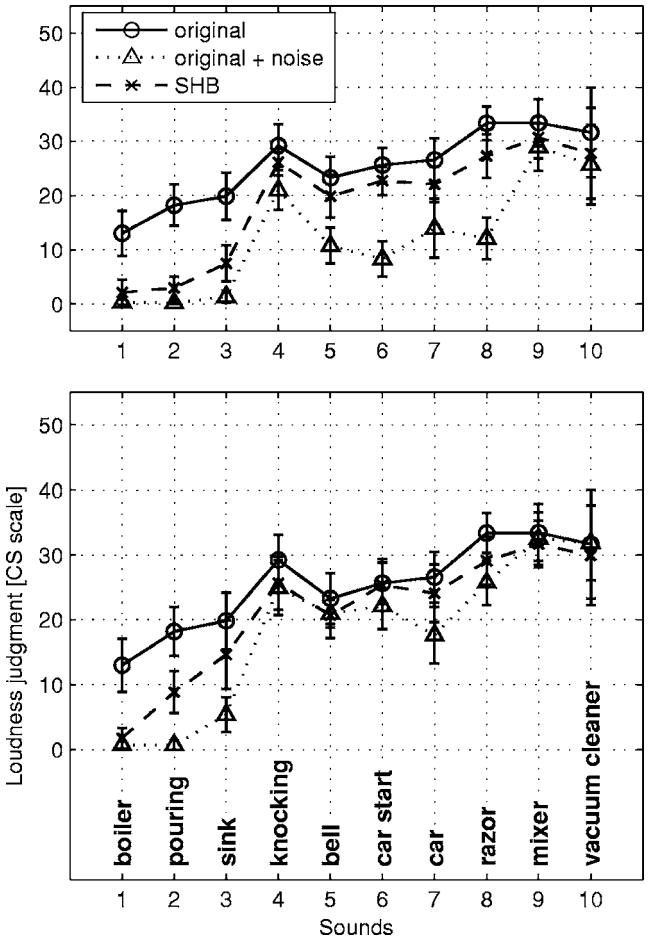


FIG. 8. Loudness judgments of the ten test sounds in the low (lower panel) and high (upper panel) background noise condition. The target sounds are labeled along the abscissa and the error bars indicate 95%-confidence intervals. Listeners focused on the target sound only.

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-

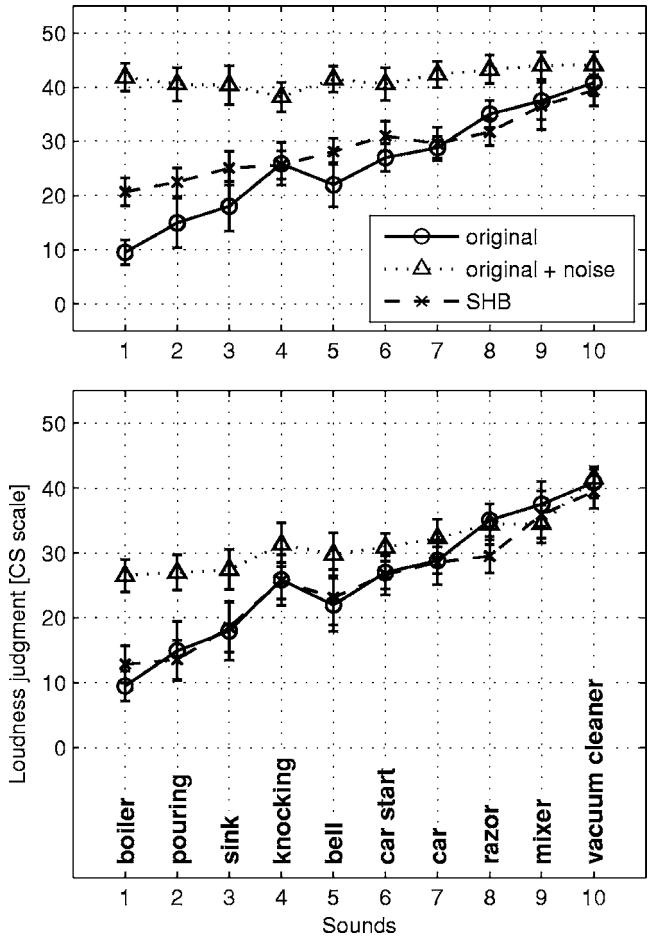


FIG. 9. Loudness judgments of the ten test sounds in the low (lower panel) and high (upper panel) background noise condition. The target sounds are labeled along the abscissa and the error bars indicate 95%-confidence intervals. Listeners judged the entire sound event.

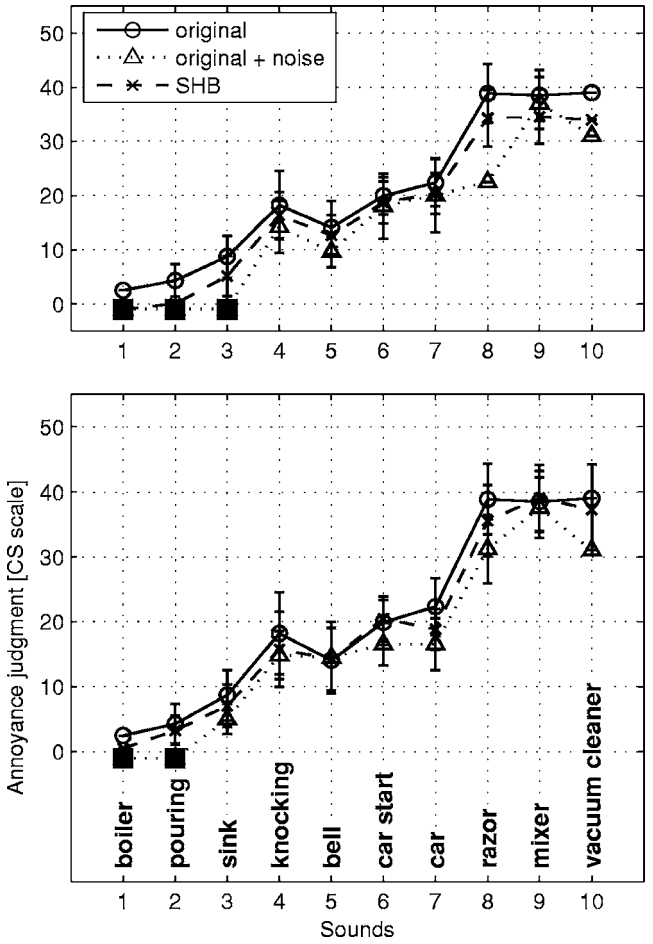


FIG. 10. Annoyance judgments of the ten test sounds in the low (lower panel) and high (upper panel) background noise condition. The target sounds are labeled along the abscissa and the error bars indicate 95%-confidence intervals. Listeners focused on the target sound only. If the majority of the participants 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

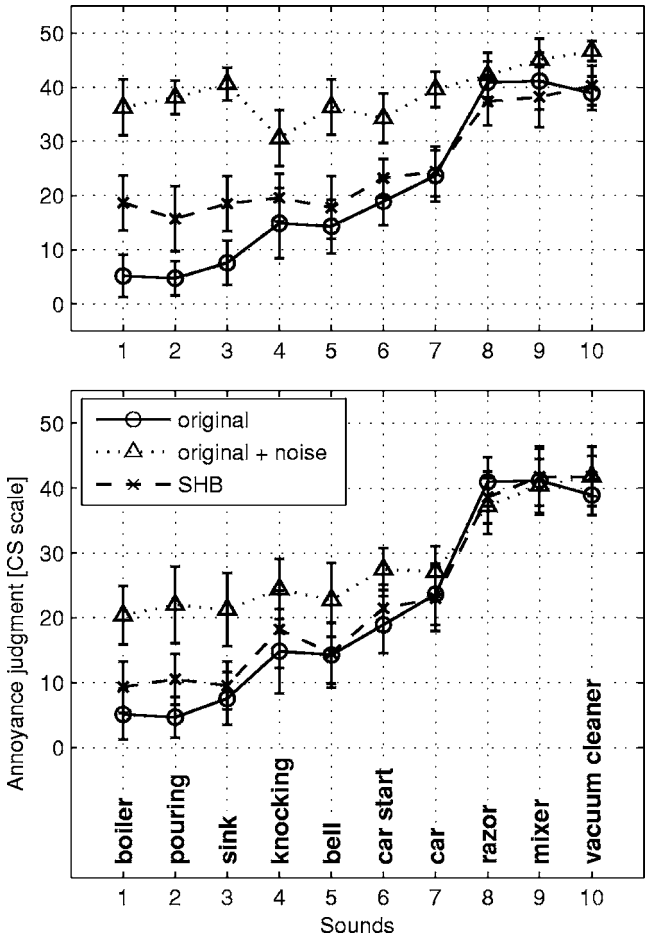


FIG. 11. Annoyance judgments of the ten test sounds in the low (lower panel) and high (upper panel) background noise condition. The target sounds are labeled along the abscissa and the error bars indicate 95%-confidence intervals. Listeners judged the entire sound event.

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]

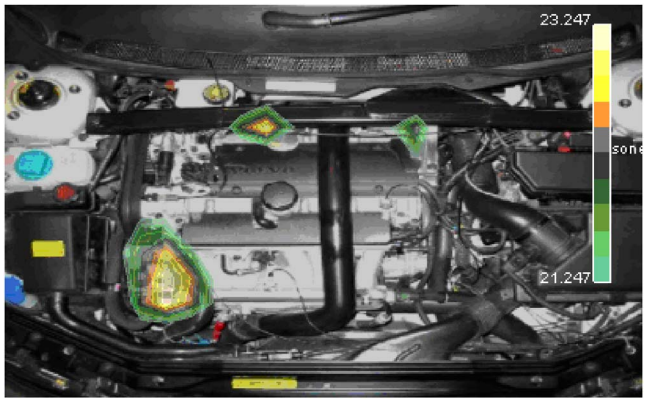


FIG. 12. (Color online) Loudness mapping of an engine compartment between 15 and 18 bark at 4000 rpm. See the text for details.

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.

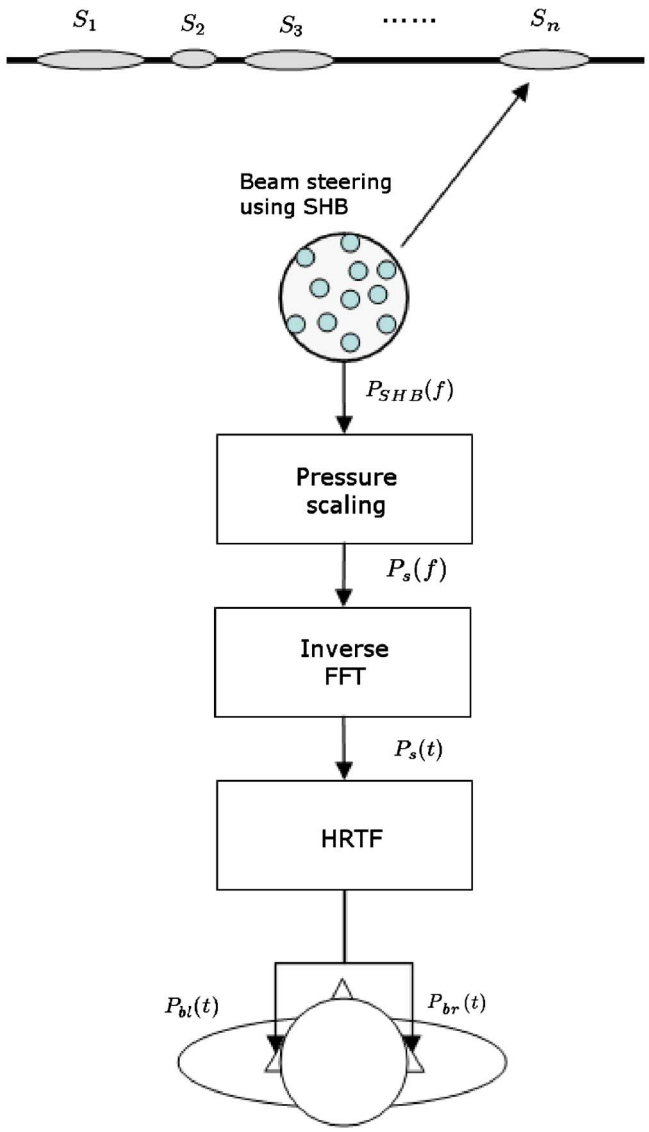


FIG. 13. (Color online) Binaural auralization of a desired sound source. Sound pressure signals are measured at each microphone position, and converted to the frequency domain. Spherical-harmonics beamforming (SHB) is applied to steer the beam toward a desired sound source and the output, $P_{\mathrm{SHB}}(f)$, 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.

Psychoacoustic evaluation of multi-channel reproduced sounds using binaural synthesis and beamforming

Wookeun Song*

Brüel & Kjær Sound & Vibration Measurement A/S, Skodsborgvej 307, 2850 Nærum, Denmark

Wolfgang Ellermeier

Institut für Psychologie, Technische Universität Darmstadt, Alexanderstr. 10, D - 64283 Darmstadt, Germany

Jørgen Hald

Brüel & Kjær Sound & Vibration Measurement A/S, Skodsborgvej 307, 2850 Nærum, Denmark

(Dated: March 2, 2008)

The binaural auralization of a 3D sound field using spherical-harmonics beamforming (SHB) techniques was investigated and compared with the traditional method using a dummy head. The new procedure was verified by comparing simulated room impulse responses with directly measured ones both monaurally and binaurally. The objective comparisons show that there is good agreement in the frequency range between 0.1 to 6.4 kHz. A listening experiment was performed to validate the auralization method subjectively. Two musical excerpts, i.e. one pop and one classical, were processed for headphone presentation in two different ways in that binaural synthesis was accomplished either (1) based on dummy head measurements or (2) SHB. Subjective responses were collected in two head motility conditions, i.e. fixed and rotating, and six spatial reproduction modes, including phantom mono, stereo, and surround sound, were applied to obtain a wide range of spatial sensations. The results show that subjective scales of width, spaciousness and preference based on the SHB auralization were not significantly different from those obtained for dummy head measurements. Thus binaural synthesis using SHB may be a useful tool to reproduce a 3D sound field binaurally while saving considerably on measurement time because head rotation can be simulated based on a single recording.

I. INTRODUCTION

Multi-channel audio has been increasingly used in automotive audio, home entertainment, and mobile phone applications, and there is a growing need for evaluating the subjective effects of such setups in listening experiments or for predicting them using objective measures. Rumsey (2002) provided a framework for conceptualizing spatial attributes based on a scene-based paradigm, which separates descriptions of sources, groups of sources, environments, and global scene parameters. Recent empirical studies (Choisel and Wickelmaier, 2006, 2007; Guastavino and Katz, 2004) investigated the identification and quantification of auditory attributes of reproduced sounds in multi-channel setups, and revealed the relationship between individual auditory attributes and overall preference.

The investigation of spatial attributes in a multi-channel setting very often requires "blind" listening experiments in order not to introduce any bias induced by prior visual exposure to the setup. Furthermore, different sets of loudspeakers and rooms may have to be compared during a listening test, e.g. to evaluate the overall audio quality of multi-channel systems in a range of different cars. For this reason, recent studies (Horbach *et al.*, 1999; Mackensen *et al.*, 2000; Spikofski and Fruhmann, 2001) have investigated methods of measuring binaural room impulse responses (BRIRs), and convolving the input signals with them according to the listener's head movement measured by a head-tracking system. This enables creating a virtual representation of the measured sound field. The method also has been used in automotive applications to estimate the subjective effects of interior car noise and to evaluate multi-channel car audio systems (Bech *et al.*, 2005; Christensen *et al.*, 2005; Farina and Ugolotti, 1997; Granier, 1996; Olive *et al.*, 2007).

It has been shown that head rotation improves the ability of sound source localization, especially for sources located in the median plane (Minnaar *et al.*, 2001; Perrett and Noble, 1997; Thurlow and Runge, 1967). Since localization may influence the judgment of other spatial auditory attributes, it appears reasonable to allow subjects to turn their head during listening tests, which involve assessing spatial sound attributes. This requires measuring BRIRs at different head rotation angles, and therefore is a very time-consuming process. By contrast, beamforming (Johnson and Dudgeon, 1993) measures a sound field with an array of microphones in a "single shot", and can by means of computation steer its beam toward a par-

*Electronic address: wksong@bksv.com

ticular direction. Furthermore, beamforming typically results in the sound pressure contribution toward the focused direction at the center of the array in the absence of the array, and this can be easily transformed to a pair of binaural signals (Song *et al.*, 2007) by incorporating binaural technology (Møller, 1992). Due to these features, beamforming may be utilized to greatly improve the efficiency of BRIR measurements when compared to traditional dummy head measurements.

The recording and analysis of a sound field using spherical microphone arrays have been thoroughly studied in recent years (Meyer, 2001; Meyer and Agnello, 2003; Petersen, 2004; Rafaely, 2004, 2005a). Since the microphones in a spherical microphone array are evenly distributed along the surface of a rigid sphere, it is possible to steer a beam in 3D space with an almost direction independent beam pattern. Park and Rafaely (2005) performed spherical microphone measurements in an anechoic chamber, and measured and classified the directional characteristics of reverberant sound fields. Rafaely (2005b) compared spherical-harmonics beamforming (SHB) with traditional delay-and-sum beamforming, and found that SHB provides similar performance when the highest spherical-harmonics order applied equals the product of the wave number and the sphere radius. However, SHB allows the use of even higher orders of spherical harmonics to provide better resolution at lower frequencies at the cost of robustness, i.e. the loss of signal-to-noise ratio.

The possibility of recording the higher-order spherical harmonics in a sound field and reproducing them by Wavefield Synthesis or Ambisonics has been investigated by Daniel *et al.* (2003) and Moreau *et al.* (2006). But these techniques require a large number of loudspeakers in a well-controlled environment such as an anechoic chamber. The same goal may be achieved by generating binaural signals obtained through either synthesis or recording. An initial attempt at a theoretical description of binaural synthesis using SHB was made by Duraiswami and co-workers (Duraiswami *et al.*, 2005; Li and Duraiswami, 2005). The advantages of spherical-harmonics beamforming, however, have not been demonstrated by means of (a) validating the mathematical procedure through a comparison of measured and synthesized binaural room responses, and (b) conducting listening tests in which subjective audio attributes are evaluated to show that the desired 3D auditory scenes may successfully be recreated.

Therefore, the current study reports on an experiment to investigate the validity of using SHB when auralizing a 3D sound field. The goals of this study are twofold:

1. To develop a binaural auralization method of a 3D sound field dependent on the listener's head rotation using SHB. To that effect, a procedure for estimating the BRIRs of individual loudspeakers in a room will have to be suggested, and a novel scaling procedure will have to be proposed to obtain the correct binaural signals at both ears. To verify the procedure objectively, the BRIRs of individual loudspeakers acquired by SHB will have to be compared with those derived from dummy head measurements.

2. To validate the proposed auralization method by obtaining subjective responses on auditory attributes, such as width, spaciousness, and preference, in a listening experiment. Synthesis based on dummy head measurements and SHB will be compared on subjective scales, and with both auralization methods the subject's head movement shall be controlled in such a way that they either rotate (with a head tracking system) or fix their head during listening tests.

To achieve these goals, BRIRs were calculated based on SHB and also measured using a dummy head in a multi-channel loudspeaker setup. By simulating BRIRs using SHB, it is possible to investigate whether measuring the sound field with a spherical microphone array and processing it via SHB will create the same subjective impression as does the dummy head technology. The proposed measurement technique will reduce measurement time dramatically, and thereby be useful in situations in which the operating conditions cannot be kept constant for repetitive measurements with different head rotations, e.g. when making on-road vehicle measurements.

II. THEORETICAL BACKGROUND

A. Binaural synthesis

Binaural signals can be recorded using a dummy head placed in a real sound field, but they can also be synthesized on a computer. This requires measuring two transmission paths. One is from a "dry" source signal to free-field pressure at the center of the head and the other is from the free-field pressure to each of the two ears, i.e. the head-related transfer functions (HRTFs). Binaural signals can then be produced by convolving a "dry" source signal with the total transmission path. The method of binaural synthesis has been verified in source localization experiments (Hammershøi, 1995; Wightman and Kistler, 1989). These investigations show that subjects judge the spatial location of properly synthesized stimuli to be the same as that of stimuli presented in free field.

For the free field situation, the transmission path from a source to the center of head position is a single function and may be measured with a microphone. However, in a listening room a direct sound as well as reflections have to be considered as separate "dry" source signals. In order to convolve each incoming acoustic wave with the HRTFs at the corresponding direction, the individual waves have to be separated from each other, and this may be achieved by obtaining the approximation of each

incoming wave at the listener's position using beamforming.

B. Spherical-harmonics beamforming

A theoretical description of SHB has been developed in Song *et al.* (2007), and it also includes the scaling procedure to estimate the free-field pressure at the center of the array contributed from a focused direction. The fundamental concepts are briefly reviewed, and a procedure taking room reflections into account is introduced in this section.

1. Fundamental formulation

Consider focusing the beamformer at a distance r_0 of a mono point source. The sound pressure produced by this monopole at the array center under free-field conditions is:

$$p_{center} = \frac{e^{ikr_0}}{kr_0} \tag{1}$$

where k is the wave number, and $i = \sqrt{-1}$. The output of SHB can be derived as a function of direction Ω (Song *et al.*, 2007).

$$w(\Omega) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \frac{P_{nm}}{R_n(ka)} Y_n^m(\Omega) \tag{2}$$

where a is the radius of the hard measurement sphere, Y_n^m are the spherical harmonics, and N is the maximum order of spherical harmonics (Williams, 1999). Park and Rafaely (2005) suggested that the maximum spherical harmonics order in SHB should be limited to $N \leq ka$, i.e. frequency dependent, in order to avoid noise originating from the high-order spherical harmonics. The pressure coefficients P_{nm} in the spherical Fourier transform are defined as (Song *et al.*, 2007):

$$P_{nm} \approx \sum_{i=1}^{M} c_i p(\Omega_i) Y_n^{m*}(\Omega_i) \tag{3}$$

where "*" represents complex conjugate, M is the number of microphones in the array, and c_i is the weights applied to the individual microphone signals $p(\Omega_i)$ at the direction Ω_i. The radial function R_n is defined as (Bowman *et al.*, 1987):

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

Here, j_n is the spherical Bessel function, $h_n^{(1)}$ the Hankel function of the first kind, and $j_n^{'}$ and $h_n^{(1)'}$ are their derivatives with respect to the argument. This shows that the directional distribution of spherical waves can be obtained by dividing the pressure coefficients P_{nm} by the radial function R_n in the spherical Fourier domain. If the goal is to make the beamformer output to be the free-field pressure at the center of the array (Eq. 1) when focused on the monopole source, then we have to scale the beamformer output by the following factor (Song *et al.*, 2007).

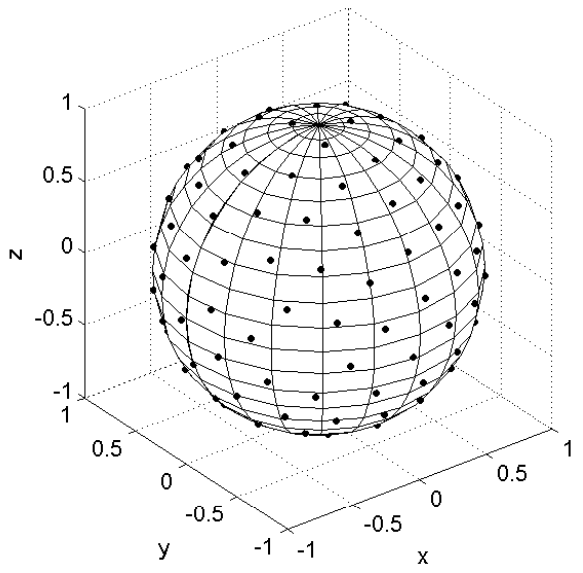


FIG. 1. The distribution of beam-focused directions along the integration sphere.

$$\frac{4\pi e^{ikr_0}}{(N+1)^2 kr_0} \tag{5}$$

This enables estimation of the correct sound pressure contributed from a focused direction, and thereby it provides a way of calculating directional contributions induced by both direct sound from individual loudspeakers and reflections from walls and physical objects located in the room.

2. Integration of beams on a sphere

In order to recreate a 3D sound field information using SHB, the contribution from individual focus (beam) directions have to be integrated. This requires determining a spatial resolution, with which the beamformer can separate major sound sources in space. For example, if the sound field is generated by a complex physical structure, the spatial resolution of SHB should be fine enough to catch spatial detail of the sound field produced. Denoting by θ the angular radius of the beam, a single beam (focus direction) covers a circular area of diameter 2θ

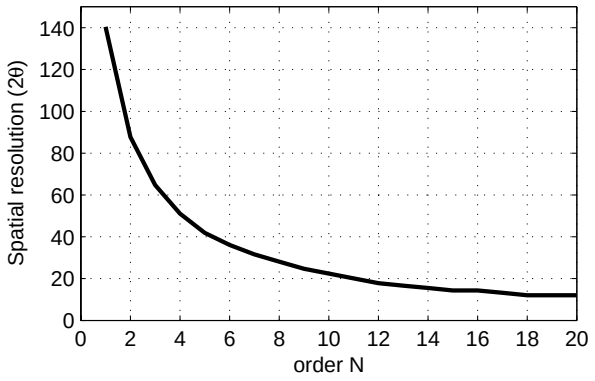


FIG. 2. Spatial resolution as a function of spherical harmonics order.

equal to the resolution. The number of directions (N_p) covering the full sphere can then be calculated as:

$$N_p = \frac{2}{1 - cos\theta} \tag{6}$$

In the present study, six loudspeakers were placed in a listening room and they will have to be simulated by SHB. A spatial resolution of $2\theta = 20°$ should be fine enough to render the spatial images of the sound field. In order to cover the full sphere, 132 directions are required according to Eq. 6.

To distribute these 132 directions evenly on a sphere of integration, the directions were initially placed randomly on a unit sphere. Subsequent iterations involved finding the two closest points and then moving them apart with a predefined step size. The optimal number of iterations was found by inspecting the resulting directional distribution visually, and the same coordinates were used for integrating beams for all loudspeaker excitations. Fig. 1 displays the distribution of points on a unit sphere, which represents the directions resulting from these iterations.

The normalized directivity pattern of SHB, $W_N(\Theta)$, can be formulated as (Rafaely, 2004):

$$W_N(\Theta) = \frac{N+1}{4\pi(cos\Theta - 1)} \left[P_{N+1}(cos\Theta) - P_N(cos\Theta)\right] \tag{7}$$

where Θ is the angle between the focused and the considered direction, and P is the Legendre polynomial. The resolution is calculated from the point -6 dB below the maximum amplitude of the calculated directivity. Fig. 2 shows the spatial resolution of SHB as a function of spherical harmonics order and indicates that higher orders of spherical harmonics should be used to achieve better spatial resolution. To obtain a spatial resolution of 20°, the spherical harmonics up to the 11th order should be taken into account according to Fig. 2.

In Song *et al.* (2007), an order limiting criterion $N \leq N_{max}$ modified after Park and Rafaely (2005) is employed to avoid noise from higher spherical harmonics orders. Since the number of spherical harmonics, $(N+1)^2$, should not exceed the number M of microphones in order to achieve good sidelobe suppression, the maximum order of spherical harmonics was decided to be $N_{max} = 7$ in the previous study where 64 microphones were used. The same way of limiting orders is used in the current investigation except that the maximum order of spherical harmonics is set to $N_{max} = 11$ due to the desired spatial resolution with the known cost of increasing sidelobes at high frequencies.

3. Frequency dependent beam width correction

The spatial resolution of SHB processing is dependent on the order of spherical harmonics employed, and an overlap of adjacent beams cannot be avoided due to the integration procedure outlined in II.B.2. One possible way of avoiding overlaps between beams is to change the number of directions for the integration process according to the spatial resolution calculated from the order of spherical harmonics. But, this may be not be practical since it requires more computation than a simple frequency weighting function. Therefore, it is proposed that the response error caused by different overlapping beams is calculated, and the inverse function of that is compensated for during the integration of beams.

The response error due to overlapping beams was calculated by simulating a monopole sound source in the frontal direction, subsequently integrating beams over a full sphere as described in the previous section, and comparing that with the true response. The design frequency is defined by $ka = N_{max}$, leading to

$$f_d = \frac{N_{max} C}{2\pi a} \tag{8}$$

where C is the speed of sound. For the microphone array used in this study, the designed frequency is 2.7 kHz. Response errors up to the design frequency were compensated for during the spatial integration of beams, but the errors at the higher frequencies were not because they are caused by spatial aliasing and not predictable.

The calculated response error as a function of frequency is shown in Fig. 3. The error increases at low frequencies as a result of the greater beam overlap, and the curve has a staircase shape due to different orders of spherical harmonics being applied dependent on frequency. A correction filter was obtained by taking the inverse of the function displayed in Fig. 3, and was applied to the calculated BRIRs.

4. Binaural auralization of a multi-channel setup using SHB

The procedure to derive binaural signals from SHB measurements was described in Song *et al.* (2007) for the free field situation. In the current investigation, however, the reflections in a listening room have to be taken into

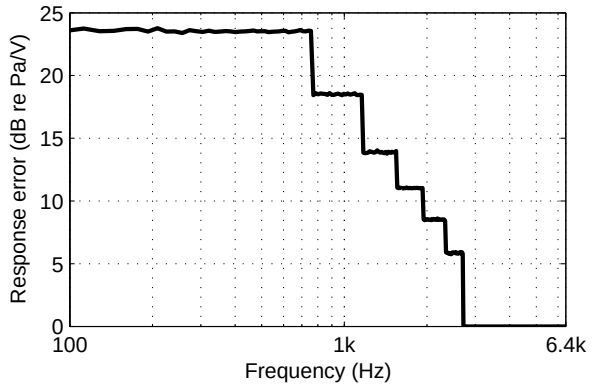


FIG. 3. Response error caused by different beam widths.

account, and this necessitates modifications of the procedure. An overview will be given in this section for the general procedure to simulate binaural signals from the measurements performed in a 3D sound field.

If $t(\Omega_i)$ is the Frequency Response Function (FRF) from a loudspeaker in a listening room to each microphone position on the sphere, the coefficients of the loudspeaker FRF's spherical Fourier transform T_{mn} can then be obtained by the following equation (Song *et al.*, 2007).

$$T_{nm} \equiv \sum_{i=1}^{M} c_i t(\Omega_i) Y_n^{m*}(\Omega_i) \tag{9}$$

Substituting Eq. 9 in Eq. 2 yields the directional response of the beamformer to a unit excitation of the speaker.

$$s(\Omega) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \frac{T_{nm}}{R_n(ka)} Y_n^m(\Omega) \tag{10}$$

The directional impulse response can then be obtained by taking the inverse temporal FFT of $s(\Omega)$. The binaural room response of the ith loudspeaker in a multi-channel setup can now be calculated as:

$$b_{li} = \sum_{p=0}^{N_p} s_i(\Omega_p) f_{inv} h_l(\Omega_p)$$

$$b_{ri} = \sum_{p=0}^{N_p} s_i(\Omega_p) f_{inv} h_r(\Omega_p) \tag{11}$$

where b_{li} and b_{ri} are the binaural room responses, s_i is the directional impulse response function of the ith loudspeaker in the direction of Ω_p, f_{inv} is the beam-width correction filter (see II.B.3), and h_l and h_r are the HRTF of the left and right ear. The BRIRs can be obtained by taking the inverse temporal FFT of b_{li} and b_{ri}.

The BRIRs of individual loudspeakers should be convolved with the input signal of each loudspeaker, and subsequently summed in order to calculate the binaural signal at both ears. The input signal of each loudspeaker will be dependent on the reproduction mode, which will be detailed in the following sections. Typically, a BRIR database is required to utilize a real-time convolution program with a head-tracking system, and the database should contain the contributions from all loudspeakers in a multi-channel setup. This was achieved by treating the BRIRs of individual loudspeakers as if they were the input of loudspeakers in the reproduction modes (e.g. phantom mono, stereo, surround) employed in this study.

III. METHOD

A. Subjects

Sixteen normal-hearing listeners between the age of 27 and 55 (15 male, 1 female) participated in the experiment. All listeners were employees of Brüel & Kjær Sound & Vibration Measurement A/S. The subjects' hearing thresholds were checked using standard pure-tone audiometry in the frequency range between 0.25 and 6 kHz and it was required that their pure-tone thresholds should not fall more than 20 dB below the normal curve (ISO 389-1, 1998) at more than one frequency. None of the thresholds exceeded 20 dB hearing level except a subject who had 30 dB hearing level at one frequency. The subjects were also screened for known hearing problems and they were not paid for their participation. The subjects were not exposed to the sounds employed prior to the experiment.

B. Apparatus and stimuli

1. Experimental setup

The experiment was carried out in a small listening room with sound-isolating walls and ceiling. The listeners were seated in a height-adjustable chair. They were instructed to look straight ahead, and were not allowed to move their head in the fixed-head condition. They were instructed to rotate their head continuously within $\pm 30°$ while listening to stimuli in the rotating-head condition. Their head movement was monitored through a window placed between the control room and the listening room. The subjects were told in the beginning of each listening test whether they had to rotate their head or not. The computer installed in the control room generated sound signals indicating the progress of the experiment as well as breaks.

A computer with a sound card (RME DIGI96) was used to transfer digital sound signals to an external D/A converter (RME ADI-8 DI). 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) was installed between the headphone amplifier and the D/A converter, and used for calibrating the playback level. A 500-Hz sine tone with a full scale was generated, and its level was checked to be 94 dB SPL at the left ear of the dummy head when the calibration tone was played with the phantom mono reproduction mode in the frontal direction. This ensured that the setup could be restored in case of some hardware changes.

Subject's head rotation was measured by a head tracker (Polhemus Fastrak) connected to a computer using an RS-232 connection. The receiver was attached to the headphones, and the transmitter was positioned on the table in front of the listeners. The update rate of the head tracker was 120 Hz. A real-time convolution software (customized for this kind of experiment by AM3D A/S) was employed to convolve the program materials with the selected BRIRs according to the subject's head rotation and to switch between different BRIR databases corresponding to different reproduction modes. The processed BRIRs had a length of 500 ms, and contained impulse responses from $-30°$ to $+30°$ of head rotation with an angular step size of $2°$. In total there were 6 reproduction modes and 2 processing modes, which led to 12 BRIR databases, and they were loaded to the real-time convolution software before the listening experiment started. Two types of databases corresponding to the two different head motility conditions were generated, and the type of database was selected by the listening test program. The maximum response time of the real-time convolution software to movements of the listener's head is guaranteed to be 15 ms at a 44.1 kHz sampling rate, which is sufficient for the current investigation.

The experiment was controlled by a customized software developed in C#. The software read the session files to assign a subject to the defined session, controlled the real-time convolution software to select the program material and to switch between BRIR databases, collected subjects' responses, and wrote the responses into text files.

2. Program materials

Two musical program materials, i.e. one pop and one classical, were selected from commercially available CDs, and they are listed in Table I. The classical music has a duration of 5:46 and the pop song of 4:41 min. The musical excerpts were repeated until the subjects completed their judgment of all reproduction modes presented on a given trial. The two program materials were selected to investigate whether their different musical content, spatial information, and recording techniques influenced the perception of spatial attributes as well as of overall quality as a function of the various reproduction modes.

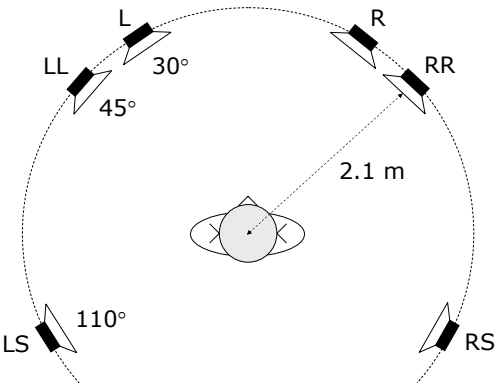


FIG. 4. The loudspeaker configuration in the multi-channel setup: left (L), right (R), left-of-left (LL), right-of-right (RR), left surround (LS), and right surround (RS).

3. Reproduction modes

The following equations were used to calculate the input of the four loudspeakers from the stereo program materials:

$$Y_L = X_L + (1-w)X_R$$

$$Y_R = X_R + (1-w)X_L$$

$$Y_{LS} = (X_L - X_R)s$$

$$Y_{RS} = (X_R - X_L)s \tag{12}$$

where X_L and X_R are the stereo signals, w is a coefficient determining the width of the stereo image, and s is a coefficient adjusting the level of surround channels. Notice that 'phantom mono' (identical signals being played through the stereo speakers) can be computed by using $w = 0$ and $s = 0$, and 'wide' stereo by using $w = 1$ and $s = 0$ while feeding the signals to the outer loudspeaker pairs, LL and RR (see Fig. 4). Six different reproduction modes (phantom mono, weak stereo, stereo, wide stereo, weak surround, and surround) were generated by selecting proper values of w and s, and the loudspeakers to play (see Table II). This selection of reproduction modes was made in order to create a wide range of spatial perception changes, and thereby the comparison between the two auralization methods based on dummy head measurements and SHB can be conducted more generally.

C. Measurements

The three different types of measurements using a microphone, a dummy head and a spherical microphone

TABLE I. List of musical program materials used

Type	Title	Album	Track	Artist
Pop	Rapunzel	Before These Crowded Streets	2	Dave Matthews Band
Classical	Concerto for bassoon 2 violins, viola, and continuo in B flat major, Allegro	Instrumental and vocal music vol. 1	8	Christoph Graupner

Name	w	s	Speakers
phantom mono (PM)	0	0	L,R
weak stereo (s)	0.5	0	L,R
stereo (S)	1	0	L,R
wide stereo (WS)	1	0	LL,RR
weak surround (snd)	1	0.5	L,R,LS,RS
surround (SND)	1	1	LL,RR,LS,RS

TABLE II. List of reproduction modes

array were performed in a listening room. The room complies with the IEC 268-13 (1985) standard, which describes an "average living room" acoustically, and has dimensions of $2.8\times4.2\times7.8m$ ($H\times W\times L$). Six loudspeakers (Genelec 1031A) were positioned at 2.1 m from the center of the setup, and their positions are shown in Fig. 4. The microphone, the two ears of the dummy head, and the center of the spherical microphone array were all 1.25 m above the floor, aligned with the tweeters of the loudspeakers. Four of the six loudspeakers were arranged in accordance with the ITU-R BS.775-1 (1994) standard: two additional speakers (LL and RR) were placed at $\pm45°$ to generate a wider stereo image than the standard one based on $\pm30°$ angular separation.

Since the microphone array and the required hardware was available for a very limited time only, time data were recorded to permit changing some of the analysis parameters without repeating the measurements. The input and output time data were recorded by means of the Data Recorder in the Brüel & Kjær PULSE software (type 3560) with a frequency range of 6.4 kHz for the microphone array, and 25.6 kHz for the microphone and the dummy head. The microphone and the dummy head signals were low-pass filtered during the calculation of impulse response functions to have the same frequency range as the array measurements. The loudspeaker input was random pink noise, and the measurement was done one loudspeaker at a time. The impulse response functions (IRF) were calculated using the auto-spectrum and cross-spectrum of input and output and taking the inverse FFT of the calculated FRF using Matlab.

The monaural room impulse response functions 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 loudspeaker during the measurement with the help of two laser beams mounted in the room. The measured IRFs were compared with the simulated ones to validate the IRFs obtained using SHB. The BRIRs of each loudspeaker were measured by placing an artificial head (VALDEMAR; Christensen and Møller, 2000) at the center of the loudspeaker setup. The dummy head was rotated from $-30°$ to $30°$ with a $2°$ angular step size to allow the rotation of the subject's head during the experiment. The dummy head measurements were compared with the ones calculated from SHB. The HRTFs employed in this study to perform binaural synthesis using SHB were taken from a database containing artificial-head HRTFs measured with a resolution of $2°$ (Bovbjerg *et al.*, 2000; Minnaar, 2001), using the same dummy head as the current study.

The 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 sound pressure mapping generated by the beamforming process could localize the correct angular position of each loudspeaker. The array with a radius of 14 cm consisted of 64 microphones (1/4 in. microphone, Brüel & Kjær type 4951) that were evenly distributed around a hard sphere in order to achieve the constant directivity pattern in all directions (see Fig. 5). In total, six loudspeaker positions and 64 microphones produced 384 impulse response functions. The headphone transfer functions (PTF) have been measured in connection with an earlier study (Song *et al.*, 2007), and theses PTFs were applied inversely to the synthesized binaural signals in this study.

D. Procedure

1. Loudness equalization of the reproduction modes

If one attribute, particularly loudness, is dominant over others perceived in the stimuli, this may affect subjective judgments during the experiment. It is often desired to minimize such attribute dominance, especially when the

FIG. 5. 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.

dominant attribute is not the subject of investigation. To this effect, the loudness of the six reproduction modes was equalized by performing a listening experiment using the method of adjustment (MA) with five normal hearing subjects.

The same user interface as in the main experiment (see Fig. 6) was presented to the subjects. The top scale was always 'phantom mono' (PM in Table II) based on the dummy head measurement, and served as the reference. The slider was positioned in the middle of this scale and disabled. The attribute to be judged was identified as loudness on the screen, and the two end points of each ruler were labelled as "softer" and "louder". The volume of the selected reproduction mode was changed according to the slider position (see Fig. 6). The subject was instructed to adjust the slider for each scale until the selected reproduction mode had the same loudness as the reference. The equalization was done for each program material. The subjects were instructed to fix their head position while they were listening to the stimuli. The order of the reproduction modes and the program materials was randomized across subjects. Fig. 7 shows the average loudness ratings as a function of reproduction mode. The upper graph shows the results from the dummy head measurements marked as HATS, and the lower from the spherical microphone array measurements marked as SHB. Similar results were obtained for the classical and pop musical excerpts, though the loudness ratings for the classical music were slightly higher on average than those of the pop music. In general higher ratings were acquired when surround channels were active (see the reproduction modes snd and SND). It can also be seen that the phantom mono reproduction mode (PM) based on SHB were perceived louder than that based on the dummy head synthesis, and it may be due to the effect of sidelobes. The channel gains were calculated as a function of reproduction mode and processing mode (HATS/SHB) based on the average loudness ratings, and were applied to the playback channels during the main experiment.

2. Training

There was a training session prior to the main experiment. The goal of the training was that the subjects become familiar with the experimental procedure, musical excerpts, and reproduction modes. Three attributes ("width", "spaciousness", "preference") were selected for the training, and different musical excerpts were used when judging each of these attributes. To achieve this, one additional pop music sample was selected, and the presentation sequence of the musical excerpts was randomized. The subjects were asked to fix their head while listening to the stimuli. For training, three reproduction modes were presented in a random order, and they were PM (phantom mono), S (stereo), and SND (surround).

It was required that the subjects should be able to distinguish between phantom mono and stereo to participate in the main experiment. If subjects were not able to distinguish these in the first training, they repeated the training session once more. Three out of original nineteen subjects were screened out based on this criterion. After the training, the subjects had a chance to discuss their listening experience with the experimenter.

3. Main experiment

The experiment consisted of two head motility conditions, i.e. fixed and rotating, to investigate the influence of head rotation on the audio quality of the auralization using SHB. Half of the subjects started judging the music samples in the fixed-head condition, and the other half in the rotating-head condition to minimize any order effects.

Quantification of two specific auditory attributes, width and spaciousness, as well as of overall preference was achieved by asking subjects to rate their subjective impression on the rating scales like those shown in Fig. 6 by positioning a slider, which assigned a value between 0 to 100. The attribute to be judged was pointed on the top of the page, and a set of scales were displayed below. Each scale had two end points, which were "narrow" and "wide" for width, "like a cigarette box" and "like a church" for spaciousness, and "not preferred" and "preferred" for preference. Definitions of the two attributes as given by Choisel and Wickelmaier (2007) were presented to the subjects prior to the experiment. The subjects were allowed to choose their own criteria to judge overall preference.

The two processing modes (HATS, SHB) and six reproduction modes resulted in twelve scales being presented to the subjects on a given trial. Next to each scale, there was a corresponding button, which served to

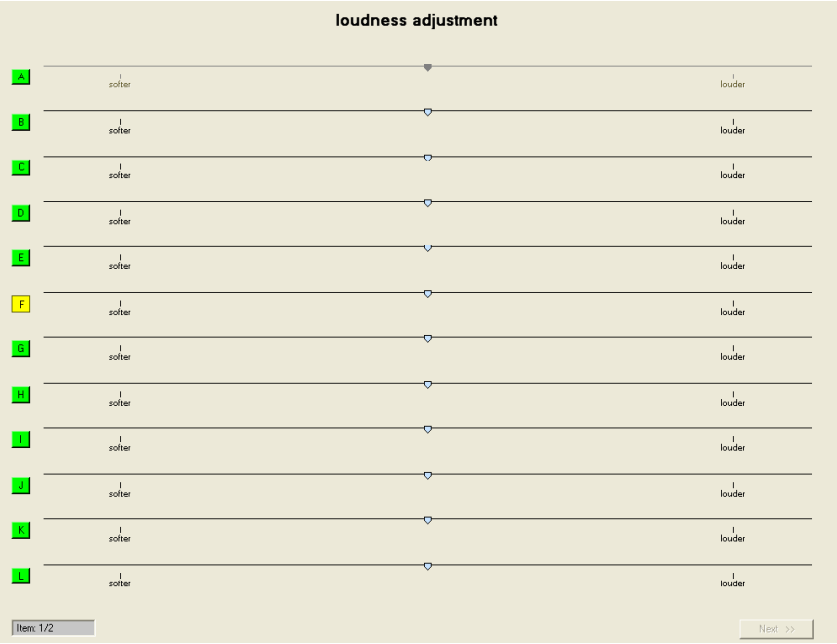


FIG. 6. The user interface employed in the experiment.

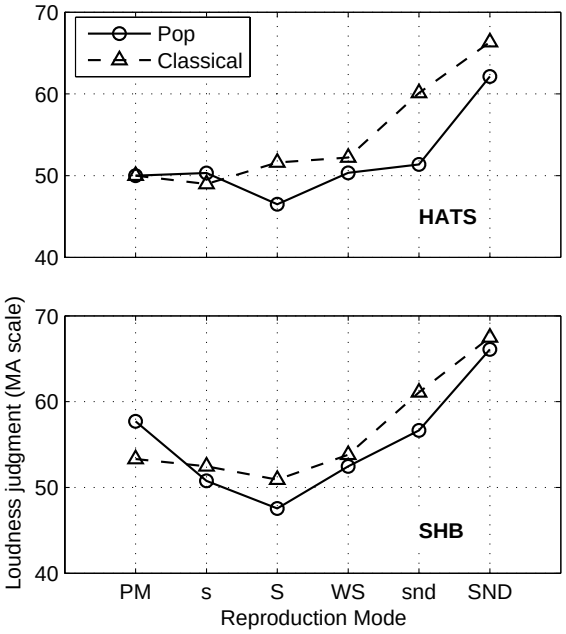


FIG. 7. Loudness judgments on average for the six reproduction modes. Top: results for binaural synthesis based on dummy-head measurements; bottom: based on SHB auralization.

activate the selected reproduction mode. The activation of the selected reproduction mode resulted in a cross-fading from the previous BRIR database to the selected one. The three attributes and the two program materials required six trials per session, run either in the fixed or the rotating-head condition. The six trials were divided into three groups within which each group the same attribute was presented in two trials with the two musical excerpts. These three groups of trials as well as the two program materials within a group were presented in a random order to the subjects. The subjects were allowed to take a short break of 1 minute after each trial, during which they stayed in the listening room. A longer break of 10 minutes was taken outside of the listening room after every other trial. The subjects spent approximately 1.5 hour per day working on each head motility condition, resulting in 3 hours total.

IV. RESULTS

In this section, the simulated room responses are compared with the measured ones both monaurally and binaurally to show any potential physical level differences caused by the beamforming process. Furthermore, the differences in perceptual quality between the two processing modes (HATS, SHB) are analyzed based on the ratings of auditory attributes in the listening tests. Notice that any discrepancies in perceptual quality emerging in the experiment may also be influenced by the procedures used in the real-time convolution software employed.

A. Comparison of measured and simulated responses

The validation of the SHB processing was conducted by comparing the measured and simulated room responses

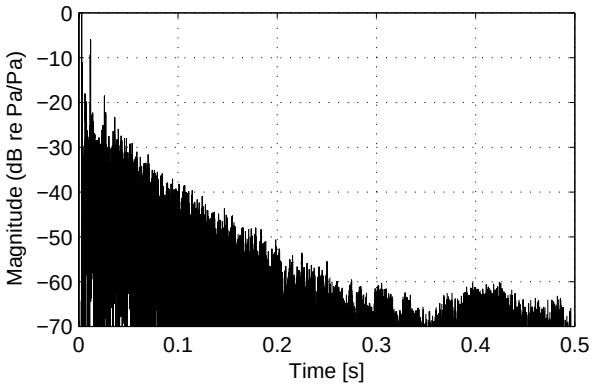


FIG. 8. Measured monaural room impulse response of the loudspeaker placed at 30° to the left (L).

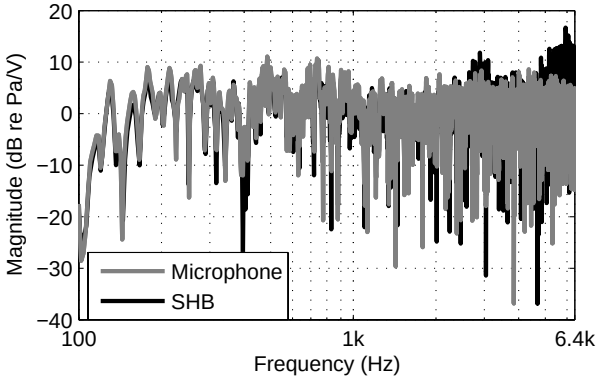


FIG. 9. Monaural room response of the loudspeaker placed in 30° to the left (L): measured with a microphone (Microphone) and synthesized using SHB (SHB).

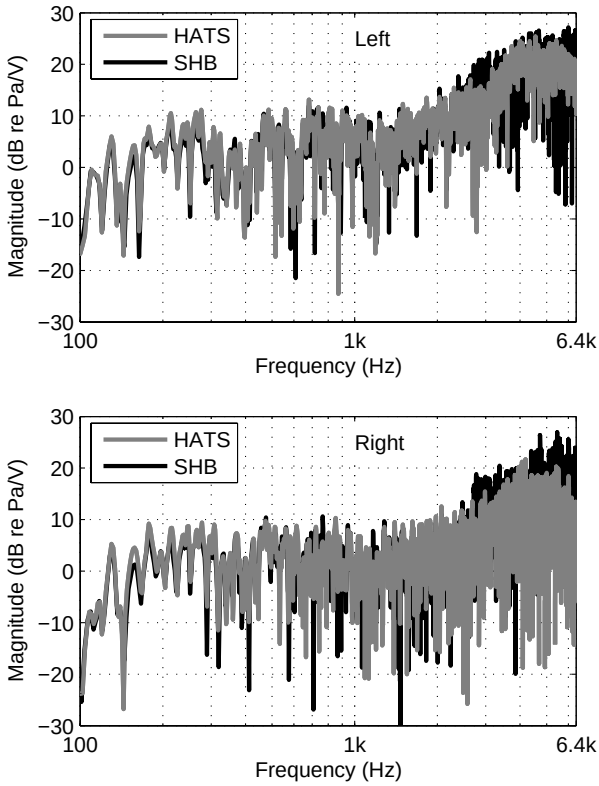


FIG. 10. Binaural room responses of the loudspeaker placed in 30° to the left (L): measured with a dummy head (HATS) and synthesized using SHB (SHB).

monaurally and binaurally. First, the IRFs from the loudspeaker input to the microphones used in the dummy head and the spherical microphone array were calculated. An example of room IRFs measured by a microphone and normalized by the peak value is shown in Fig. 8. A signal-to-noise ratio (SNR) of 60 dB is achieved in the measurement, and it can be seen from the noise part of the response after 0.3 second.

For the SHB calculation, the IRFs measured by the 64 microphones placed on the hard sphere were used as an input to the procedure. The monophonic room impulse responses were calculated by integrating the beams covering a full integration sphere, scaling the output by the inverse function of the beam-width effect, and subsequently taking the inverse FFT of the output signal (see II.B.2 and II.B.3). The BRIRs were also calculated by taking into account HRTFs while integrating the individual beams (see II.B.4).

The simulated and measured mono room responses were compared in the frequency range from 0.1 to 6.4 kHz, and an example for the loudspeaker placed 30° to the left hand side (L) is displayed in Fig. 9. There is good agreement between the measured and simulated responses up to the designed array frequency, i.e. 2.7 kHz in this case, and the error grows towards higher frequencies. This may be due to high sidelobe levels caused by spatial under sampling at high frequencies. This tendency was the same for all loudspeakers.

The binaural room responses simulated by SHB for the same loudspeaker are plotted together with the measured ones in Fig. 10. The head was placed in the frontal direction for this comparison. The same tendency is observed as for the monophonic measurements in that the curves are quite similar, and the response errors increase at high frequencies. These investigations confirm that binaural auralization using SHB produces binaural signals physically close to the measured ones while greatly saving time when measuring a 3D sound field.

B. Scaling of auditory attributes

The ratings of the three auditory attributes were averaged across the 16 subjects for each reproduction mode in the two processing modes (HATS, SHB) and 95%-confidence intervals were determined. The outcome is shown in Figs. 11 to 14. The results of the dummy head

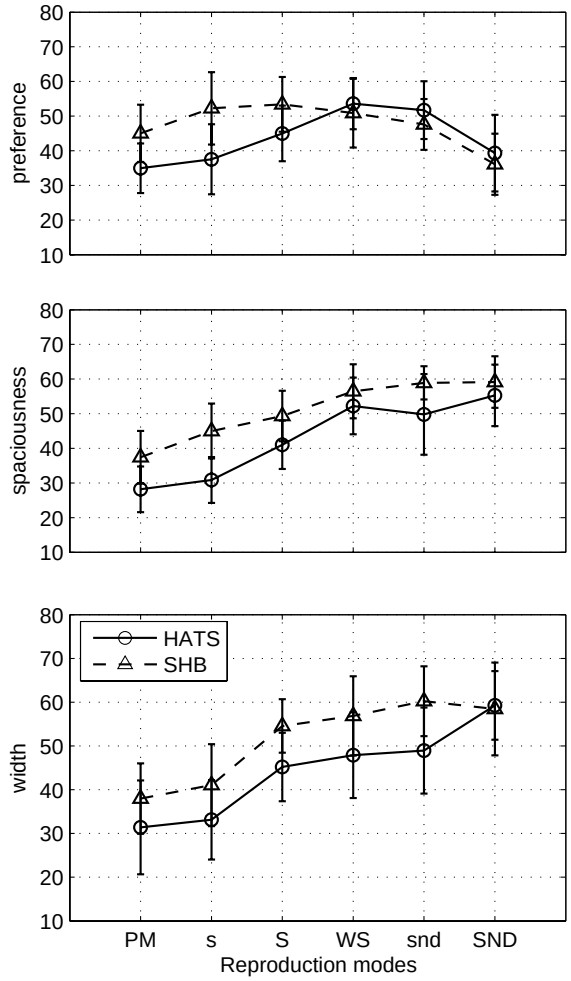


FIG. 11. Sound quality ratings of the pop music excerpt in the fixed-head condition. Top: overall preference; center: spaciousness; bottom: width. Dashed line: SHB processing; sold line: HATS synthesis.

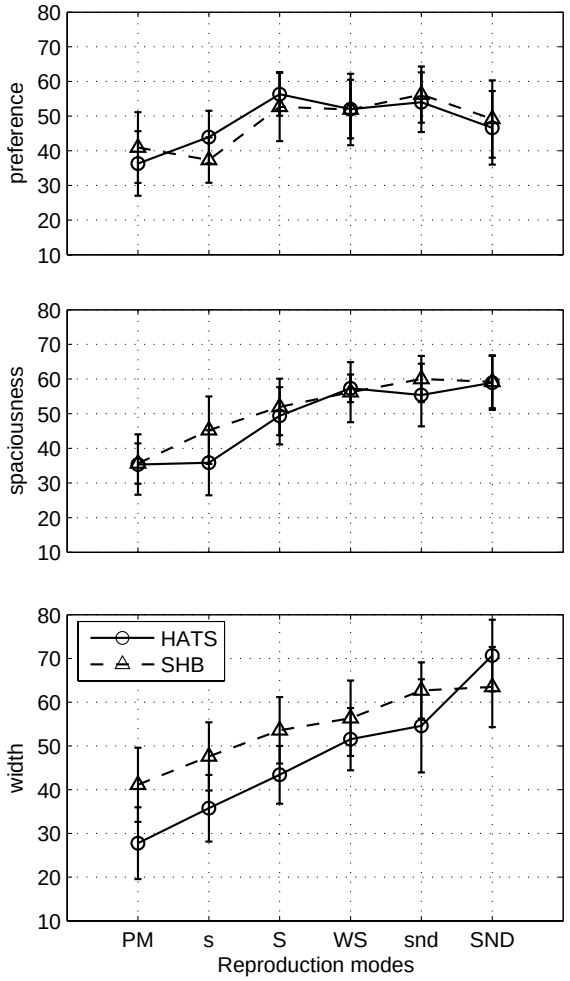


FIG. 12. Sound quality ratings of the classical music excerpt in the fixed-head condition. Data arranged as in Fig. 11.

measurements (HATS) are drawn with solid lines, and those of SHB with dashed lines. Notice that the graphical scales presented to the subjects were coded with values from 0 to 100, while the figures display a range between 10 to 80 to emphasize the effects.

When the pop music was presented in the fixed-head condition (see Fig. 11), as in all other conditions (see Figs. 12 - 14), the six reproduction modes differed markedly in preference, and in the ratings of the two spatial auditory attributes. The significance of this effect of the experimental manipulation was confirmed by performing a three-factor analysis of variance (ANOVA; Montgomery, 2004) with the 6 reproduction modes, the 2 processing modes (SHB, HATS), and the 3 attributes all constituting within-subjects factor. This analysis indicated a highly significant effect of the reproduction mode [$F(5, 75) = 13.38$, $p < 0.001$], which incidentally was of similar magnitude in all other conditions studied (see Figs. 12 - 14). Furthermore, largely similar curves were obtained for the two processing modes, but the SHB processing produced higher responses than the dummy head synthesis, particularly for width and spaciousness. The statistical significance of this discrepancy shows up to a main effect of processing mode [$F(1, 15) = 6.51$; $p = 0.022$] in the ANOVA. It may be the effect of ghost images generated by sidelobes, which create the percept of additional diffuseness in the reproduced sounds.

As regards overall preference, the wide stereo (WS) and the two multi-channel reproduction modes (snd, SND) were judged quite similarly when comparing the two processing modes, but the subjects preferred the SHB processing over the dummy head synthesis in the three two-channel reproduction modes (PM, s, S). This may be due to the fact that the additional diffuseness created spatial impressions resembling those produced by the surround channels. It can also be seen that the subjects made quite similar responses when asked about width or spaciousness, and thus for this particular mate-

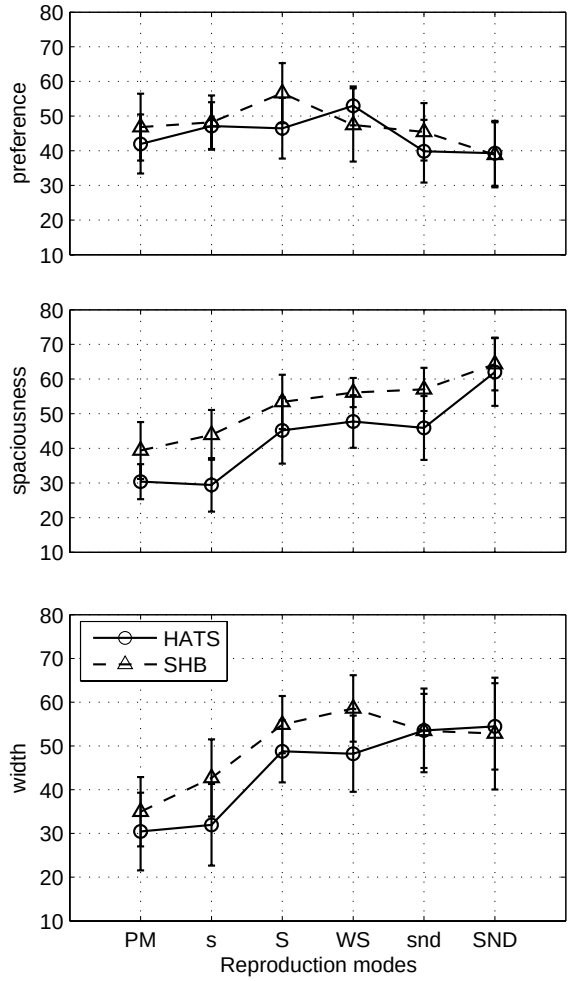


FIG. 13. Sound quality ratings of the pop music excerpt in the rotating-head condition. Data arranged as in Fig. 11.

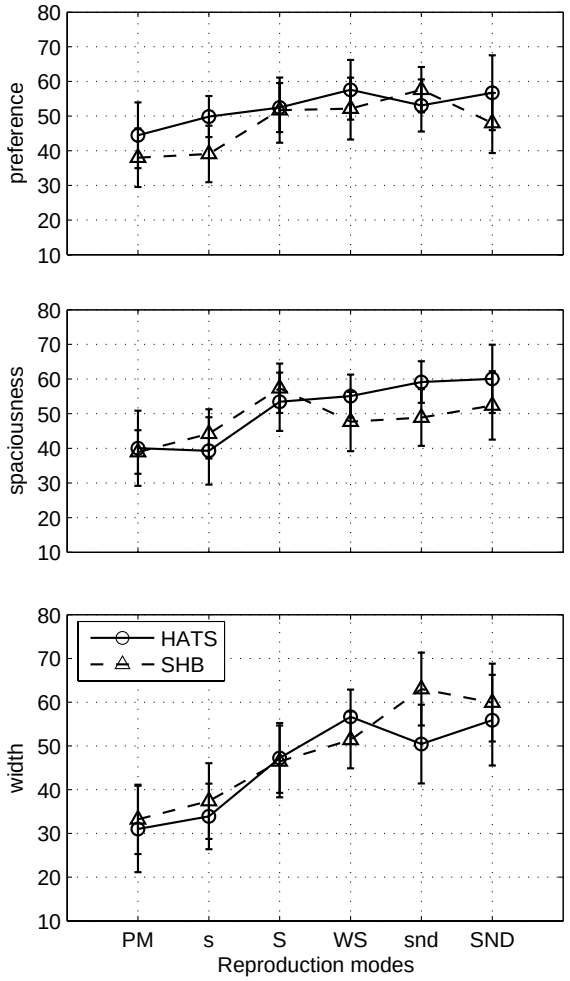


FIG. 14. Sound quality ratings of the classical music excerpt in the rotating-head condition. Data arranged as in Fig. 11.

rial hardly distinguished these two attributes. The participants generally preferred the wide stereo (WS) and the multi-channel reproduction (snd), while they disliked the reproduction mode with a higher level of surround channels (SND).

Judging the classical music excerpt reduced the differences between the two processing modes (HATS, SHB), except for judgments of width (see Fig. 12). Here, the main overall effect of processing mode did not reach statistical significance [$F(1,15) = 1.43$; $p = 0.25$], but the three-way interaction between processing, the reproduction modes, and the attributes did [$F(10, 150) = 1.91$; $p = 0.049$], indicating that the divergence seen for the width ratings for the less complex reproduction modes (PM, s, S; bottom panel in Fig. 12) appears to be significant.

This indicates that the SHB processing can approximate listening to the sound fields recorded with a dummy head in terms of spaciousness, overall audio quality, and to some extent, width. For the classical music, the interpretation may be that the effect of ghost images only influences the perception of width, but not of spaciousness. It can still be seen that the two stereo (S, WS) and the two multi-channel reproduction (snd, SND) modes are almost equally preferred while the subjects did not prefer phantom mono (PM) and the narrow reproduction (s).

The results discussed so far imply that auditory attributes of recorded 3D sound fields may be faithfully rendered by measuring the sound field with a spherical microphone array, and reproducing it in a fixed-head condition. Width is the most sensitive attribute and somewhat affected by the beamforming processing, and the perception of the multi-channel reproduction modes (snd, SND) was less affected than that of the simpler reproduction schemes. The results seem to be dependent on the musical excerpts for spaciousness and preference, but not for width. The effect of head rotation will be analyzed in the following.

When the subjects were asked to rotate their head

while listening to the pop music excerpt (see Fig. 13), almost identical responses were obtained for width and spaciousness. A four-factor analysis of variance with the two head motility conditions (fixed and rotating) constituting an additional within-subjects factor revealed no significant main effect of head motility condition [$F(1, 15) = 0.02$, $p = 0.89$], as well as no significant interactions of head motility with any of the other factors ($p > 0.22$). Nevertheless, the preference judgments appear to show less of an effect when compared to the fixed-head condition. The two multi-channel reproduction modes (snd, SND) are no longer preferred, and the two stereo reproduction modes (S, WS) are slightly preferred over the others.

For the classical music excerpt (see Fig. 14), the two head-motility conditions again yielded quite similar results, except for ratings of width (compare the bottom panels of Figs. 12 and 14). The effect of processing mode in width became smaller in the rotating-head condition. It is also interesting that in the rotating-head condition spaciousness of wide stereo (WS) and the two multi-channel reproductions yielded smaller values for SHB compared to HATS while preference are quite similar to the fixed-head condition. This was evident in the significant interaction of the attribute judged with the head condition [$F(2, 30) = 7.59$, $p = 0.002$].

These results indicate that allowing for head rotation may modify sound quality judgments to some extent like seen in the rating of width for the classical music and of preference for the pop music, but it certainly does not reveal further differences between the two processing modes (SHB, HATS) when compared to a fixed-head listening test. The results from these investigations thus show that binaural auralization using SHB can be used for reproducing recorded 3D sound fields while listeners are allowed to rotate their head freely.

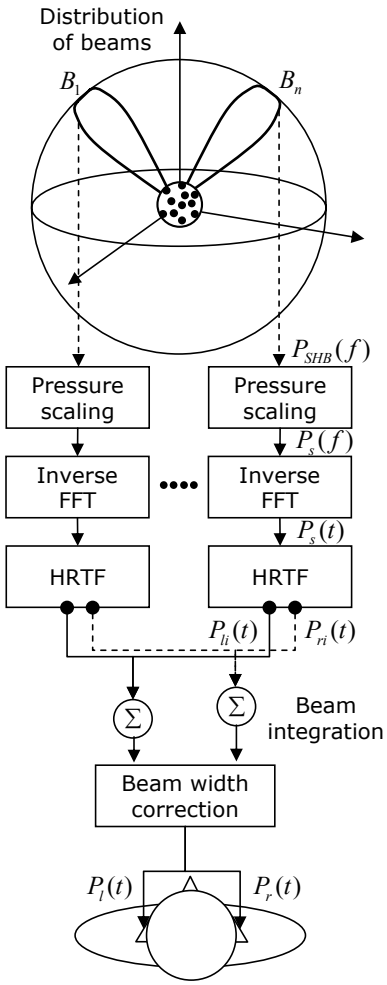


FIG. 15. Proposal for the binaural auralization of a 3D sound field using spherical harmonics beamforming (SHB). For details, see text.

V. DISCUSSION

Recently, it has become a popular strategy in automotive audio engineering to develop target sounds by modifying the contributions from sub-systems, e.g. the exhaust, by manipulating level, frequency dominance, or order balance (e.g. Brassow and Clapper, 2005). This is done by recording interior sounds in various operating conditions, decomposing the recorded sounds, and subsequently modifying the decomposed sounds. This strategy helps to define realistic target sounds through a series of subjective listening tests in which automotive experts or potential customers may participate.

A Noise, Vibration, and Harshness (NVH) simulator improves the process of defining target sounds by delivering the right context and enabling back-to-back comparisons (e.g. Jennings *et al.*, 2007). Such simulators employ source-path-contribution analysis (Schuhmacher and Tcherniak, 2006) to estimate the transfer paths from sub-systems to both ears of a dummy head, and thereby provide methods of synthesizing vehicle sounds by modifying the measured transfer paths from each component.

This process typically requires to measure sets of FRFs from the defined point sources to both ears of a dummy head placed in a car cabin, and measuring source strengths during a variety of operating conditions. It is very time consuming work to prepare such data for a NVH simulator, and it may not be feasible at all to measure FRFs at a large number of head-rotation angles to allow for head rotation during a listening experiment. By contrast, when a spherical microphone array is placed in a vehicle instead of a dummy head, much more efficient transfer function measurements can be made, and FRFs at different head-rotation angles may easily be calculated by the procedure outlined in this study. The proposed method will also be useful to record interior noise during on-road testing, and to reproduce it with a head-tracking

system in a laboratory. Note that this is not possible with traditional methods, such as dummy head recordings, since it is nearly impossible to replicate the exact operating conditions, background noise, and road noise.

Thus, a procedure for deriving binaural signals from spherical microphone array measurements has been developed and experimentally validated to enable audio engineers to reproduce 3D sound fields binaurally based on "one-shot" measurements. Since the procedure utilizes beam steering to derive binaural signals dependent on the listener's head rotation, no physical rotation of the measurement devices is required in contrast to typical dummy head measurements. The details of the proposed method will be given in the following.

Fig. 15 shows the flow of the calculation proposed here. First sound pressure signals are measured at each microphone of the array, and the SHB processing is applied to steer the beam toward each direction that is defined for the distribution of beams over a full sphere (see II.B.2). The output of the beamforming ($P_{SHB}(f)$) is scaled to produce the correct free-field pressure contribution ($P_s(f)$) from a given direction in the absence of the array. The corresponding pressure time data, $P_s(t)$, are calculated by taking the inverse FFT of $P_s(f)$. The binaural signals contributed from the single direction ($P_{li}(t)$, $P_{ri}(t)$) can be acquired by convolving the free-field pressure with the HRTF for that direction. The HRTF can be selected from the closest measured points in the database or interpolated between the two nearest points (Algazi *et al.*, 2004). Finally, the binaural signals of the whole 3D sound field ($P_l(t)$, $P_r(t)$) can be obtained by summing the individual binaural signal contributions and subsequently applying the correction for the beam width.

The method proposed in this study is not limited to a specific type of sound field, and therefore can be used to record sound fields like a listening room or a vehicle interior. On the other hand, the presence of the array may influence the characteristics of the original sound fields in some cases, and it is recommended that the size of a spherical microphone array be approximated to the average size of human heads to minimize such effects. It is planned to apply the suggested algorithm directly to automotive sounds fields, and to perform a series of listening experiments to prove that the technique can reproduce automotive interior sounds with the aid of a head-tracking system while preserving the original auditory scenes as well as the overall quality of the sound field in a car cabin.

VI. CONCLUSION

1. A theoretical method for integrating beams formed by spherical harmonics beamforming (SHB) over a sphere was proposed to include direct transmissions from sound sources to the listener's position as well as reflections from walls and physical objects in a room. The contributions from individual beams were convolved with the HRTFs of a dummy head in the same direction, and subsequently summed to generate binaural signals. As a final step, the effect of different beam widths dependent on frequency was simulated, and an inverse filter of the simulated response was applied to scale the binaural signals.

2. This procedure of binaural auralization using SHB was validated by comparing measured and simulated room frequency responses (both monaural and binaural) for individual loudspeakers. It was found that there is good agreement in the frequency range between 0.1 to 6.4 kHz, and response errors grow above the designed array frequency due to the effect of sidelobes.

3. A listening experiment was performed to validate the procedure, i.e. to show that the SHB auralization yields similar results as does binaural synthesis based on measurements made with a head-and-torso simulator. When subjects were asked to rate auditory attributes elicited by a multichannel loudspeaker array, by and large, the two auralization methods produced quite similar results. For a pop music sample investigated, however, subjects judged the width and spaciousness of the stimuli processed by SHB to be slightly higher than with the synthesis based on the dummy head measurements, and thus preferred the SHB processing over the dummy head synthesis in the three two-channel reproduction modes (PM, s, S). This may be due to the fact that the additional diffuseness caused by sidelobes created spatial percepts in the two-channel reproduction modes similar to what the surround channels did. On the other hand, both processing modes (HATS, SHB) produced similar responses for preference judgments. When the subjects judged a classical music excerpt, the difference between the two processing modes became even smaller for spaciousness and preference.

4. No significant effects of head rotation was observed except that the preference for the pop music reproduction scenarios became less discriminated and that the two processing techniques resulted in more similar subjective width ratings for the classical music in the rotating-head condition.

5. The implications of the current study in terms of practical benefits for sound-quality engineering were outlined and a procedure for deriving binaural signals using SHB for binaural synthesis was sketched. The suggested procedure can be applied to situations in which more efficient recording of 3D sound fields is required or where defined operating conditions cannot be repeated for measuring an entire set of head rotation angles, e.g. when auralizing on-road vehicle testing.

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). The authors would like to thank Sylvain Choisel for sharing his knowledge of multi-channel sound reproduction.

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

Bech, S., Gulbol, M.-A., Martin, G., Ghani, J., and Ellermeier, W. (**2005**). "A listening test system for automotive audio Part 2: Initial verification", in *Audio Engineering Society, 118th Convention*, preprint 6359 (Barcelona, Spain).

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*, preprint 5264 (Los Angeles, CA, USA).

Bowman, J. J., Senior, T. B. A., and Uslenghi, P. L. E. (**1987**). *Electromagnetic and acoustic scattering by simple shapes* (Hemisphere Publishing Corp., New York, USA).

Brassow, B. and Clapper, M. (**2005**). "Powertrain Sound Quality Development of the Ford GT", in *The Society of Automotive Engineers Noise and Vibration Conference and Exhibition*, preprint 2480 (Traverse City, MI, USA).

Choisel, S. and Wickelmaier, F. (**2006**). "Extraction of auditory features and elicitation of attributes for the assessment of multichannel reproduced sound", J. Audio Eng. Soc. **54**, 815–826.

Choisel, S. and Wickelmaier, F. (**2007**). "Evaluation of multichannel reproduced sound: Scaling auditory attributes underlying listener preference", J. Acoust. Soc. Am. **121**, 388–400.

Christensen, F., Martin, G., Minnaar, P., Song, W., Pedersen, B., and Lydolf, M. (**2005**). "A listening test system for automotive audio Part 1: System Description", in *Audio Engineering Society, 118th Convention*, preprint 6359 (Barcelona, Spain).

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

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*, preprint 5788 (Amsterdam, Netherlands).

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 Headphones with HRTF Cues", in *Audio Engineering Society, 119th Convention*, preprint 6540 (New York, NY, USA).

Farina, A. and Ugolotti, E. (**1997**). "Subjective comparison of different car audio systems by the auralization technique", in *Audio Engineering Society, 103rd Convention*, preprint 4569 (New York, USA).

Granier, E. (**1996**). "Comparing and Optimizing Audio Systems in Cars", in *Audio Engineering Society, 100th Convention*, preprint 4283 (Copenhagen, Denmark).

Guastavino, C. and Katz, B. F. G. (**2004**). "Perceptual evaluation of multi-dimensional spatial audio reproduction", J. Acoust. Soc. Am. **116**, 1105–1115.

Hammershøi, D. (**1995**). "Binaural technique – a method for true 3d sound reproduction", Ph.D. thesis, Aalborg University.

Horbach, U., Karamustafaoglu, A., Pellegrini, R., Mackensen, P., and Theile, G. (**1999**). "Design and Applications of a Data-based Auralization System for Surround Sound", in *Audio Engineering Society 106th Convention*, preprint 4976 (Munich, Germany).

IEC 268-13 (**1985**). "Sound system equipment, part 13: Listening tests on loudspeakers", International Electrotechnical Commission, Geneva, Switzerland .

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

ITU-R BS.775-1 (**1994**). "Multichannel stereophonic sound system with and without accompanying picture", International Telecommunication Union, Geneva, Switzerland .

Jennings, P., Giudice, S., Fry, J., Williams, R., Allman-Ward, M., and Dunne, G. (**2007**). "Developing best practice for sound evaluation using an interactive nvh simulator", Transactions Of Japanese Society Of Automotive Engineers **38**, 31–36.

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

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*, 106–109 (New York, NY, USA).

Mackensen, P., Fruhmann, M., Thanner, M., Theile, G., Horbach, U., and Karamustafaoglu, A. (**2000**). "Head Tracker-Based Auralization Systems: Additional Consideration of Vertical Head Movements", in *Audio Engineering Society, 108th Convention*, preprint 5135 (Paris, France).

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*, preprint 5975 (New York, NY, USA).

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

Minnaar, P., Olesen, S. K., Christensen, F., and Møller, H. (**2001**). "The importance of head movements for binaural room synthesis", in *Proceedings of the 2001 International Conference on Auditory Display*, 21–25 (Espoo, Finland).

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

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

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, France).

Olive, S., Welti, T., and Martens, W. L. (**2007**). "Listener loudspeaker preference ratings obtained in situ match those

obtained via binaural room scanning measurement and playback system", in *Audio Engineering Society, 122nd Convention*, preprint 7034 (Vienna, Austria).
Park, M. and Rafaely, B. (**2005**). "Sound-field analysis by plane-wave decomposition using spherical microphone array", J. Acoust. Soc. Am. **118**, 3094–3103.
Perrett, S. and Noble, W. (**1997**). "The effect of head rotations on vertical plane sound localization", J. Acoust. Soc. Am. **102**, 2325–2332.
Petersen, S. O. (**2004**). "Localization of sound sources using 3D microphone array", Master's thesis, University of Southern 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. (**2005**a). "Analysis and design of spherical microphone arrays", IEEE Transactions of Speech and Audio Processing **13**, 135–143.
Rafaely, B. (**2005**b). "Phase-mode versus delay-and-sum spherical microphone array processing", IEEE Signal Processing Letters **12**, 713–716.
Rumsey, F. (**2002**). "Spatial quality evaluation for reproduced sound: Terminology, meaning, and a scene-based paradigm", J. Audio Eng. Soc. **50**, 651–666.
Schuhmacher, A. and Tcherniak, D. (**2006**). "Engine contribution analysis using a noise and vibration simulator", in *International Conference on Modal Analysis Noise and Vibration Engineering*, preprint 4003.
Song, W., Ellermeier, W., and Hald, J. (**2007**). "Using beamforming and binaural synthesis for the psychoacoustical evaluation of target sources in noise", Under revision within J. Acoust. Soc. Am. .
Spikofski, G. and Fruhmann, M. (**2001**). "Optimisation of Binaural Room Scanning (BRS): Considering inter-individual HRTF-characteristics", in *19th Conference of the Audio Engineering Society*, 272–286 (Schloss Elmau, Germany).
Thurlow, W. R. and Runge, P. S. (**1967**). "Effect of induced head movements on localization of direction of sounds", J. Acoust. Soc. Am. **42**, 480–&.
Wightman, F. L. and Kistler, D. J. (**1989**). "Headphone simulation of free-field listening. II: Psychophysical validation", J. Acoust. Soc. Am. **85**, 868–878.
Williams, E. G. (**1999**). *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic Press, London, Great Britain).