Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation[a)]

Ville Pekka Sivonen[b)]
Sound Quality Research Unit (SQRU), Department of Acoustics, Aalborg University, Fredrik Bajers Vej 7 B5, 9220 Aalborg East, Denmark and Brüel & Kjær Sound & Vibration Measurement A/S, Skodsborgvej 307, 2850 Nærum, Denmark

Wolfgang Ellermeier
Sound Quality Research Unit (SQRU), Department of Acoustics, Aalborg University, Fredrik Bajers Vej 7 B5, 9220 Aalborg East, Denmark

(Received 25 July 2005; revised 7 February 2006; accepted 13 February 2006)

The effect of sound incidence angle on loudness was investigated using real sound sources positioned in an anechoic chamber. Eight normal-hearing listeners produced loudness matches between a frontal reference location and seven sources placed at other directions, both in the horizontal and median planes. Matches were obtained via a two-interval, adaptive forced-choice (2AFC) procedure for three center frequencies (0.4, 1, and 5 kHz) and two overall levels (45 and 65 dB SPL). The results showed that loudness is not constant over sound incidence angles, with directional sensitivity varying over a range of up to 10 dB, exhibiting considerable frequency dependence, but only minor effects of overall level. The pattern of results varied substantially between subjects, but was largely accounted for by variations in individual head-related transfer functions. Modeling of binaural loudness based on the at-ear signals favored a sound-power summation model, according to which the maximum binaural gain is only 3 dB, over competing models based on larger gains, or on the summation of monaural loudness indices. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2184268]

PACS number(s): 43.66.Pn, 43.66.Cb [AK] Pages: 2965–2980

I. INTRODUCTION

There is growing awareness in psychoacoustics that, for a thorough understanding of loudness perception, its binaural nature has to be taken into account. That is true for basic research, namely the construction of general loudness models (e.g., Moore *et al.*, 1997), as well as for applications to audio reproduction systems (Zacharov *et al.*, 2001) or to perceived sound quality (Bodden, 1997). Especially for instrumental loudness predictions based on Zwicker's modeling, as standardized in ISO 532 (1975), the fact that it is essentially monophonic has been regarded as a major drawback. Nevertheless, the adjustments recently made to loudness modeling rest on a fairly narrow empirical data base, which the present study hopes to extend.

To clarify the issues, it may be helpful to distinguish two stages of processing involved when the loudness of a real sound source in space is perceived: (1) the physical transformation of the "distal" stimulus emitted by the source to "proximal" stimuli arriving at the listener's ears, and (2) the neural, psychological, and cognitive process of integrating the two at-ear stimuli into a single percept.

A. Physical (HRTF) filtering

The first stage can be described in purely acoustical terms, namely by applying head-related transfer functions (HRTFs, Shaw, 1974; Wightman and Kistler, 1989a; Møller *et al.*, 1995; Blauert, 1997, Chap. 5). These account for the filtering of the source due to the physical effects of the human torso, head, and pinnae, depending on the incidence angle of the sound. Further along, through the ear canal, the physical sound transmission has been shown to be independent of the direction of the sound source (see, e.g., Hammershøi and Møller, 1996). Thus, the direction-dependent part of an HRTF can be measured at the entrance to the blocked ear canal, and described by (adopted from Møller *et al.*, 1995)

$$\mathrm{HRTF}_{\text{dir-dep}}(\phi,\theta)=\frac{P_2}{P_1}(\phi,\theta), \tag{1}$$

where ϕ is azimuth, θ is elevation, P_1 is sound pressure at the center position of head, and P_2 is sound pressure at the entrance to the blocked ear canal.

In the median plane, the HRTFs of the two ears are fairly similar due to the physical symmetry of the human body in this plane. However, level differences between HRTFs for different directions can approach 10 dB or more over a fairly wide frequency range. By contrast, large interaural time and level differences (ITDs and ILDs, respectively) between the two ears emerge in the horizontal plane, where the ILDs can reach up to 30 dB at high frequencies. HRTFs have been a major research topic during the past

[a)]Portions of the data were presented at the 147th Meeting of the Acoustical Society of America, New York, May, 2004, and at Internoise, Prague, August, 2004.
[b)]Electronic mail: vps@acoustics.aau.dk

15 years, but the focus of this research has been on adequate "auralization" or sound localization (Wightman and Kistler, 1989a,b; Bronkhorst, 1995; Møller *et al.*, 1996), not on loudness.

B. Binaural loudness summation

The second stage of processing has been termed binaural loudness summation. It describes how the acoustic inputs to the left and right ear are integrated to yield a single, binaural loudness. Starting from the observation that a sound appears louder when listened to with both ears (i.e., binaurally) than with only one (i.e., monaurally), a number of investigators conducted experiments using headphones, through which different combinations of left- and right-ear sound-pressure levels were presented in order to quantify this effect. The results are often summarized as providing evidence for a binaural-to-monaural loudness ratio of 2:1, or perfect loudness summation, corresponding to a binaural gain of approximately 10 decibels (e.g., Levelt *et al.*, 1972; Marks, 1978; Schneider and Cohen, 1997), in accordance with the sone scale of loudness. The evidence is far from unequivocal, however, with many studies finding less-than-perfect summation (e.g., loudness ratios of approximately 1.5:1; Zwicker and Zwicker, 1991), and a level dependence of the binaural gain, which appears to increase from approximately 3 dB near threshold to 6–10 dB at high sound-pressure levels (Shaw *et al.*, 1947; Reynolds and Stevens, 1960; Hellman and Zwislocki, 1963).

Interestingly, binaural loudness summation, as conceptualized in this paradigm, has not been investigated with sounds that are likely to reach the eardrums when emitted from a real source in space, i.e., with products of the first (HRTF) filtering stage. Rather, artificial sounds such as tones, or broadband noise, lacking all spatial or directional information have been used, often at interaural level differences (e.g., in monotic-to-diotic comparisons) far exceeding what would naturally occur. Such conditions of stimulation do not yield an externalized sound image, but rather more or less lateralized inside-the-head percepts. Generally, it appears that the considerable literature on binaural loudness summation has contributed more to the development of scaling methodologies than to the auditory issues involved.

C. Loudness of free and diffuse sound fields

For practical purposes, in an attempt to relate the monophonic measurement of a sound field to perceived loudness, two specific types of sound fields have been considered: The free field, where the sound incidence angle is frontal to the listener, and the diffuse field, where the sound is reaching the listener's ears with equal intensity from all directions.

In order to account for the fundamental difference in sound incidence, the standardized loudness model (ISO 532, 1975) has different computation procedures for the two sound fields. The two procedures are based on both objective and subjective data (Kuhl and Westphal, 1959; Robinson *et al.*, 1961; ISO 389-7, 1996): The objective data represent the differences in the at-ear sound pressures between the two sound fields, i.e., investigating only the effect of the first (HRTF) filtering stage; the subjective data represent the differences in perceived loudness, including effects of both the first and the second stage. Even though the agreement between the objective and subjective data is fair, these investigations do not specify how the two signals at the ears of a listener are summed into a single loudness percept, due to the fact that the stimulation of the auditory system in both sound fields is essentially diotic.

The increasing use of dummy heads for acoustical recordings and measurements, often resulting in dichotic at-ear signals, has led to growing interest in how dichotic at-ear signals should be summed to correspond to the diotic stimulation of the conventional free- and diffuse-field loudness paradigms.

D. Directional loudness

Thus, while studies of HRTF filtering have not explicitly been concerned with the loudness of dichotic sounds, the work on binaural loudness summation appears to lack ecological validity to predict the perception of real sources positioned in space. What remains, then, is less than a handful of studies that have actually investigated directional loudness of real sources in space, taking into account both stages delineated: the physical filtering due to HRTFs, and the ensuing "psychological" summation.

Sivian and White (1933) investigated the effect of direction on hearing thresholds, reporting that at absolute threshold, the binaural minimum audible field is not significantly different from the monaural one. This implies no or a very small binaural advantage, the ear receiving the higher sound pressure determining the binaural hearing threshold. While the directional HRTF effects are the same at higher sound-pressure levels, extrapolating from a detection task to suprathreshold binaural loudness and to its summation across the two ears may be unjustified.

By far, the most pertinent and complete study investigating directional loudness was published by Robinson and Whittle (1960) more than 45 years ago. The authors used a circular array of 12 equally spaced loudspeakers positioned around the listener seated in an anechoic room to obtain loudness matches between a reference and each test position. Using narrow-band sounds having six center frequencies between 1.6 and 10 kHz, and rotating their apparatus when required, they investigated the horizontal, median, and frontal planes in a sample of 16 to 20 listeners. Using probe-tube microphones they also measured sound-pressure levels at the ears of their subjects, as produced by the same stimuli, thus obtaining crude magnitudes of "HRTFs" for the six test frequencies.

As expected, the average loudness matches showed a strong frequency dependence, with the greatest directional effects (of up to 15 dB; see their Fig. 2) observed at higher frequencies (4–10 kHz). Relating the mean loudness matches to the average at-ear sound-pressure measurements, Robinson and Whittle (1960; see their Fig. 5) found the former to be reasonably well predicted by a "6-dB summation rule,"

$$L_{\mathrm{mon}} = 6 \times \log_2(2^{L_{\mathrm{left}}/6} + 2^{L_{\mathrm{right}}/6}), \tag{2}$$

where L_{mon} is the equivalent sound pressure needed for monotic stimulation to match any binaural (diotic: $L_{\mathrm{left}} = L_{\mathrm{right}}$, or dichotic: $L_{\mathrm{left}} \neq L_{\mathrm{right}}$) combination of left-ear (L_{left}) and right-ear (L_{right}) input levels. If, for example, both ears are exposed to 70 dB SPL, the equivalent monotic SPL turns out to be 76 dB SPL (i.e., a 6-dB binaural gain). Note, however, that there is not sufficient information on the fit of this heuristic (other than what can be judged from visual inspection of their Fig. 5) in Robinson and Whittle's report, or on its feasibility to predict individual subjects' data.

Two more recent studies picked up the issue, though using considerably fewer experimental conditions and subjects, and not including HRTF measurements. Both studies (Remmers and Prante, 1991; Jørgensen, 2002) used wideband noise to obtain loudness matches, thus potentially obscuring a frequency dependence, and obtained much smaller effects than Robinson and Whittle (1960), with directional loudness matches varying by less than 3 dB across incidence angles.

E. Rationale for the present study

It thus appears worthwhile to take up the issue of directional loudness once more. This will be done paying special attention to five methodological issues, which are elaborated in turn:

(1) Well-defined narrow-band stimuli are needed to investigate the effects of HRTFs and binaural loudness summation. Note that, in Robinson and Whittle's (1960) report, the sounds used were not sufficiently specified beyond stating that they were "below a critical band" (p. 75), and the later studies used wideband noise which might wash out some of the directional effects.

(2) Given the evidence from earlier headphone experiments showing the binaural gain to increase from approximately 3 dB near threshold to up to 10 dB at high sound-pressure levels, level effects will be taken into account by making measurements at two overall sound-pressure levels.

(3) With the exception of Jørgensen's (2002) study, classical "method(s) of adjustment" have been used to collect the subjective data. By their transparency, and the explicit control they give listeners over the outcome, these methods are prone to subject-induced biases, such as "correcting" an adjustment due to some expectation. Forced-choice procedures (Levitt, 1971; Jesteadt, 1980), especially when interleaving adaptive tracks for different experimental conditions, are much less susceptible to such biases.

(4) Advances in the methodology to HRTFs will be brought to the study of directional loudness. Note that Robinson and Whittle's (1960) pioneering study was done before the term HRTF was coined, and that their at-ear measurements of the stimuli actually used merely provide six points along the frequency scale, and thus do not constitute HRTFs as we conceive of them today.

(5) Since HRTF filtering is known to be highly idiosyncratic, it is likely that with averaged data frequency-dependent directional effects might partially cancel each other, thus underestimating the true effect size. Therefore, a greater emphasis than in earlier studies will be on individual results and analyses.

To sum up, the present investigation will be conducted by having subjects assess loudness in a directional sound field in an anechoic room, and by relating the listening test data both to the distal stimulus given by the sound-pressure level emitted by the active loudspeaker, and to the proximal stimuli given by the participants' at-ear exposure levels as obtained via state-of-the-art HRTF measurements.

II. METHOD

A. Subjects

Eight normal-hearing listeners (between the age of 22 and 46 years; five male, three female), including the second author, participated in the experiment. The subjects' hearing thresholds were determined using standard pure-tone audiometry in the frequency range between 0.25 and 8 kHz with the requirement that none of the thresholds exceed 15 dB hearing level *re*: ISO 389-1 (1998). The five subjects who were not staff members were paid an hourly wage for their participation.

B. Apparatus

1. Loudspeaker setup in the anechoic chamber

The experiment was carried out in an anechoic chamber, which is anechoic above approximately 200 Hz, and has background noise at an inaudible level.

The loudspeaker setup for the experiment consisted of eight identical speakers (Vifa M10MD-39) mounted in hard plastic balls with a diameter of 15.5 cm. A typical frequency response of the loudspeaker can be found in Møller *et al.* (1995).

The loudspeakers were positioned both in the horizontal and median planes. In the horizontal plane, the incidence angles were 30°, 60°, 90°, and 135° of azimuth, and in the median plane the angles were 45° and 90° of elevation. Loudspeakers were also placed ahead and behind the listening position (at 0° and 180° of azimuth with 0° of elevation), where the horizontal and the median planes coincide. Due to assumed symmetry, the loudspeakers were placed only on the left-hand side in the horizontal plane. The distances from the diaphragms of the loudspeakers to the listening position at the center of the setup were 206±4 cm.

The subjects were seated in a chair, the height of which could be adjusted. The chair had a small headrest to restrict head movements of the subjects during the experiment. The subjects' heads and ears were carefully aligned with the center position of the setup by making adjustments to chair height and headrest position using a laser and two video cameras. A photograph of the setup in the anechoic chamber is shown in Fig. 1. The loudspeakers ahead, at 30° and 60° in the horizontal plane, and at 45° and 90° in the median plane are visible in the photograph. The structure suspending the

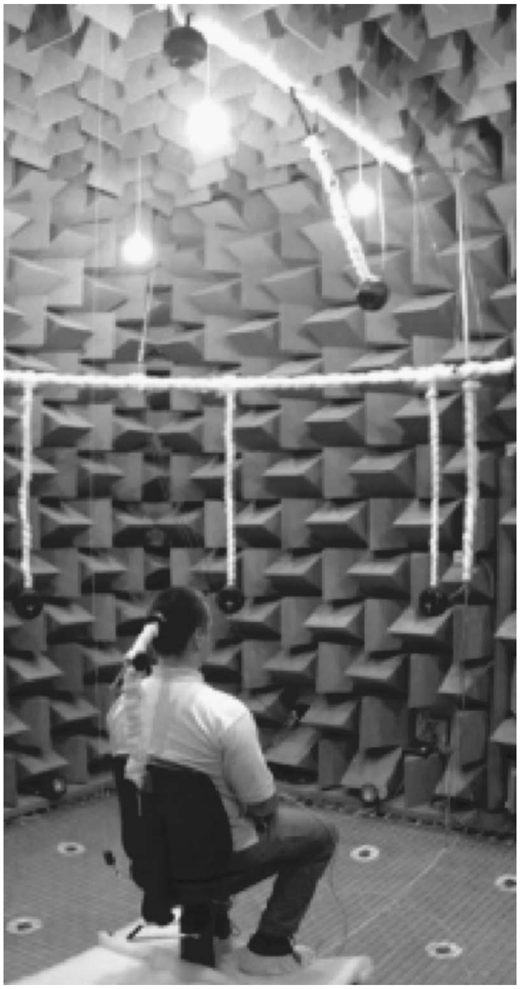

FIG. 1. The experimental setup in the anechoic chamber.

loudspeakers and the platform (an open metal grid) under the chair were covered with sound-absorbing material.

The subjective responses were collected with a two-button response box. The response box had small lights above the buttons to indicate observation intervals. An enlarged copy of the indicator lights was placed behind and slightly above the frontal loudspeaker to avoid subjects tilting their heads downwards to the response box in their hands.

2. Signal generation and control

All other equipment was placed in a control room next to the anechoic chamber. A personal computer (PC) was used for controlling the experiment and carrying out objective measurements. The PC was equipped with a digital sound card (RME DIGI96/8 PST) with eight audio channels, connected to an external AD/DA-converter (RME ADI-DS8). A custom-made eight-channel attenuator with a 128-dB dynamic range and 0.5-dB step size was used to individually control the level of the eight loudspeakers. The signals from the attenuator were amplified by power amplifiers (Rotel RB-976 Mark II), and then fed to the loudspeakers in the anechoic chamber.

The experiment was run using a program developed in LABVIEW. The program took care of reading session files, playing back appropriate stimuli, collecting subjects' responses, adapting the attenuator gains according to the responses, and writing the data into result files.

C. Measurements

Acoustical measurements were carried out using the maximum-length-sequence (MLS) system as specified by Olesen *et al.* (2000), with an MLS order of 12, preaveraging of 16, and a sampling rate of 48 kHz. The length of the acquired impulse responses was 4095 samples, which, due to the scarcity of reflections inside the anechoic chamber, was long enough to avoid time aliasing. The measurements were carried out at a level of approximately 70 dB SPL (at 1 kHz), measured in the absence of a listener at the center position of the setup.

First, responses of each loudspeaker [P_1 pressures, see Eq. (1)] were measured at the center position using a 1/4-in. pressure field microphone (Brüel & Kjær type 4136) with 90° incidence to the loudspeaker under measurement. Then, responses of each loudspeaker at each listener's ears [individual P_2 pressures, see Eq. (1)] were measured at the blocked entrance to the ear canal using two miniature microphones (Sennheiser KE 4-211-2), one microphone specifically for each ear. The miniature microphones were fitted inside foam earplugs (E·A·R Classic, halved in length), and mounted flush with the ear-canal entrance. All microphone signals were bandpass filtered between 22.5 Hz and 22.5 kHz by the measurement amplifier used (Brüel & Kjær type 2607 or type 2690 Nexus).

The above measurements were carried out three times: in the beginning, halfway through, and at the end of the experiment. The loudspeaker responses were used to equalize the stimuli for the listening experiment and to obtain reference pressures (P_1) for the HRTF calculations. The responses at each listener's ears were used to obtain individual HRTFs. The HRTF measurement procedure was as described by Møller (1995) with the following exceptions: The subjects were sitting in a chair instead of standing, the anechoic chamber was smaller, and the MLS measurement system was different.

Computation of the HRTFs involved 1024 samples from P_1 and P_2 pressures. First, individual head-related impulse responses (HRIRs) were calculated from P_1 and P_2 including a correction for the differences in the frequency responses of the two types of microphones used in the measurements. These HRIRs included reflections from the loudspeaker setup; therefore, only 140 samples from the HRIRs were used for calculating the final HRTFs. The resulting samples included all reflections from the subjects themselves (and from the chair), but excluded reflections from the other loudspeakers, the loudspeaker suspension, and any other objects inside the anechoic chamber. Note, however, that the excluded reflections were very small compared to the magnitude of the pure HRTFs.

D. Stimuli

The stimuli used for the listening experiment were third-octave noise bands centered at 0.4, 1, and 5 kHz. The length of each stimulus was 1 s.

For generating the stimuli, a 1-s white-noise signal was created, and subsequently filtered using third-octave-band filters at each center frequency. The relative differences in the frequency responses of the loudspeakers were equalized by applying minimum-phase inverse filters based on the direct sound coming from the loudspeakers. Each narrow-band signal was convolved with each of the inverse filters characterizing the individual loudspeakers, resulting in 24 stimuli for each (center frequency $\times$ loudspeaker) combination. Finally, raised-cosine rise and decay ramps of 20-ms duration were applied. The sound files thus corrected were played back at a sampling rate of 44.1 kHz, and with 16-bit resolution in the experiment proper.

The third-octave-band levels of the stimuli were aligned to 64.7 ± 0.1 dB SPL at 0.4 kHz, 64.7 ± 0.2 dB SPL at 1 kHz, and 63.9 ± 0.1 dB SPL at 5 kHz. At the highest possible playback level (75 dB SPL) the levels of the second- and third-order harmonics were more than 37 and 43 dB below the level of the center frequencies of the narrow-band noises, respectively. The distortion was measured to be highest at the lowest center frequency, but it was inaudible for all stimuli. Furthermore, the spectral envelope of the equalized stimuli was verified to be very similar between different loudspeakers.

In the experiment proper, the stimuli were played back at two overall levels as measured at the listener's position; a "low" overall level of around 45 dB SPL and a "high" overall level of around 65 dB SPL. Even though the actual measured sound-pressure levels deviated slightly from these values, note that the misalignment between the loudspeakers was less than ± 0.2 dB at each center frequency.

E. Procedure

The aim of the experiment was to determine how loudness is affected by the sound incidence angle at three center frequencies and two overall levels. This was accomplished by matching the loudness of test sounds emanating from each of the loudspeakers in the setup to the loudness of the same sound coming from the reference loudspeaker positioned in front of the subject at 0° of azimuth and elevation.

1. Adaptive matching procedure

Matches were obtained using a two-interval, adaptive forced-choice (2AFC) procedure (Levitt, 1971) converging on the point of subjective equality (PSE) by following a simple 1-up, 1-down rule. On each trial, the (variable) test sound, and the (fixed) frontal reference were presented in random order, with a 500-ms pause in between. Synchronized with the sounds, two light-emitting diodes were successively lit both on the hand-held response box, and on its larger model in order to mark the observation intervals to be compared. The subject's task was to judge which of the two noises sounded louder by pressing one of the two buttons aligned with the observation-interval lights. The participants were instructed to judge the loudness of the sounds only, and to disregard any other differences (due to direction, or timbre, for example) they might perceive.

For each adaptive track, the overall level of the frontal reference was fixed to either 45 or 65 dB SPL, as was the center frequency of the sounds to be played, and the test loudspeaker to be matched. The level of the test loudspeaker, however, was controlled by the adaptive procedure: Whenever the subject judged the test sound to be louder than the (frontal) reference, its sound-pressure level was lowered by a given amount; whenever the subject judged the reference to be louder, the level of the test loudspeaker was increased by that same amount. The initial step size was 4 dB; after two reversals (i.e., changes in the direction of the adaptive track) it was decreased to 1 dB. A total of eight reversals was collected in each adaptive track; the arithmetic mean of the last six of them was used to estimate the PSE. Two different starting levels were employed for the adaptive tracks, one 10 dB above, one 10 dB below the level of the reference loudspeaker, thus providing clearly discriminable loudness differences at the outset of each track.

2. Experimental design

For a given overall level, the experimental design required loudness matches to be determined in 44 different experimental conditions. These resulted from the factorial combination of 7(test loudspeakers) $\times$ 3(center frequencies) $\times$ 2(adaptive starting levels), and additional two conditions of the reference loudspeaker being matched to itself for the 1-kHz center frequency only (using both starting levels) to obtain a measure of the baseline variability of the matches.

Collection of these data was organized as follows: In order to allow subjects to adapt to a given loudness range, "high-SPL" (65 dB; "*A*"), and "low-SPL" (45 dB; "*B*") measurements were strictly separated in different sessions, which were counterbalanced following a succession of *ABBA* (respectively, *BAAB*) schemes. The order of the 44 experimental conditions to be investigated at each level was randomized, and subsequently divided into blocks of eight (the remaining four being assigned to the next block, i.e., the following replication of the measurements). Thus, within a given block of trials, eight adaptive tracks were randomly interleaved on a trial-by-trial basis, providing some random sampling of loudspeaker locations, center frequencies, and starting levels. Consequently, it was impossible for the subjects to track the immediate "adaptive" consequences of their judgments, and from their perspective the task was just a succession of unrelated paired comparisons with respect to loudness.

Each listening session consisted of four such blocks (containing eight adaptive tracks each). Completing a block of trials took approximately 10 min. While it lasted, the subjects were instructed to sit as still as possible in order to maintain the alignment with the loudspeaker setup. A short break was taken after each first and third block in a session, and participants were allowed to move their heads and upper body during those breaks, but not to leave the chair. After each second block they had a longer break during which they

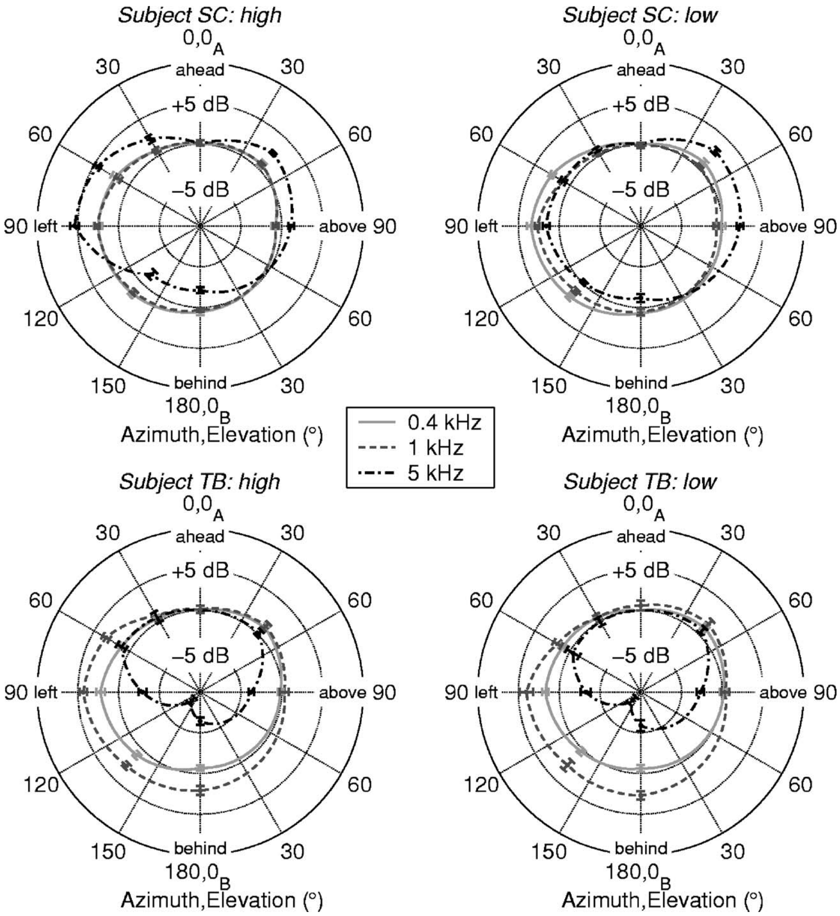

FIG. 2. Directional loudness sensitivities at the two overall levels for subjects SC and TB with 95% confidence intervals of 16 replications. The graphs on the left show the results for the high overall level (65 dB SPL), the graphs on the right for the low overall level (45 dB SPL). Elevations 0°_{A} and 0°_{B} are ahead and behind the listener, respectively.

left the anechoic chamber, thus requiring them to realign the seating position upon return. An entire session lasted approximately 1 h.

Since 16 replications of the matches (eight with each of the two adaptive starting levels) were collected per experimental condition, all subjects had to participate in 22 listening sessions. The participants completed a maximum of two sessions per day with a minimum of 1 h between sessions. With three additional sessions reserved for audiometry, HRTF measurements, and practice (one block in each of the high-SPL and low-SPL conditions), the total number of hours amounted to 25 per subject.

III. RESULTS

A. Directional loudness sensitivities

The adaptive procedure matched the loudness of a sound of a given center frequency coming from one of the loudspeakers in the horizontal or the median plane to the loudness of the same sound with frontal incidence. Thus, the raw data from the experiment were the sound-pressure levels (in dB SPL) the loudspeakers would have to be set to, in order to be perceived equally loud as the frontal reference. These raw data were averaged across the 16 repetitions that each participant accumulated in each condition, and normalized by subtracting the result from the fixed level of the respective frontal reference (65 or 45 dB SPL). That way, (relative) directional loudness sensitivities[1] were obtained, positive values of which indicate loudness enhancement, i.e., a lower sound-pressure level required for that direction to achieve a match with the frontal reference.

1. Individual data

Individual directional loudness-sensitivity curves are depicted for two subjects, SC (upper panels) and TB (lower panels), representing extremes of performance, in Fig. 2. The data are rendered in polar coordinates, though in a particular, asymmetrical way: The left-hand side of each polar graph shows the data for the horizontal plane as the loudspeakers were physically positioned in the setup. On the right-hand side of each polar graph the data are shown for the median plane where the loudspeakers were actually above the subjects. Note that these two planes coincide ahead of and behind the subjects.

For subject SC, loudness matches at 0.4 and 1 kHz vary as a function of sound incidence angle over a range of approximately 3 dB, the subject being most sensitive to loudness for sounds coming from the side, i.e., from 90° to the left in Fig. 2. That holds for both overall levels used. At 5 kHz, by contrast, this pattern is observed at the high overall level only, whereas at the low level the loudness pattern is fairly omnidirectional in the horizontal plane. In the median plane the directional patterns are similar across overall levels.

For subject TB, loudness matches vary over a range of less than 3 dB at 0.4 kHz. At 1 kHz the direction has a larger

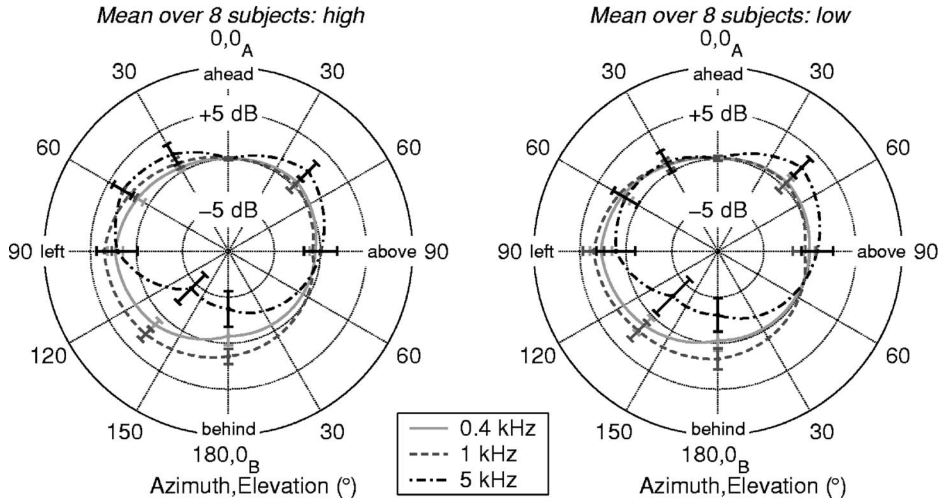

FIG. 3. Directional loudness sensitivities at the two overall levels for means across all eight subjects with 95% confidence intervals of the means. The graph on the left shows the results for the high overall level (65 dB SPL), the graph on the right for the low overall level (45 dB SPL). Elevations 0°_{A} and 0°_{B} are ahead and behind the listener, respectively.

effect on loudness, the sensitivity being up to 4 dB higher on the left-hand side than straight ahead. At 5 kHz the directional effect is even more pronounced, the minimum sensitivity at 135° in the horizontal plane being approximately 8 dB below and the maximum being close to the frontal sensitivity. The directional sensitivity patterns for this subject do not appear to be level dependent.

The confidence intervals for the matches of subjects SC and TB in Fig. 2 are small. Average individual standard deviations of the loudness matches across all subjects were 1.0 and 1.2 dB at the high and low overall levels, respectively. All participants adjusted the (identical) 1-kHz frontal test sound to a sound-pressure level close to the (fixed) reference, indicating that there was no systematic bias in the matches. The standard deviation of the identical-direction matches (0.9 dB) was only marginally lower than that of the across-direction matches, suggesting that these were of no greater difficulty.

2. Group data

Figure 3 shows mean loudness sensitivities when data are aggregated across all of the eight subjects. When the listener-specific idiosyncrasies are thus removed, directional loudness sensitivity still varies over some 3 dB at the two lowest center frequencies, whereas at 5 kHz the directional effect is approximately twice as large. Also, the error bars are larger at the highest center frequency due to a wider spread in the individual data. The overall level does not seem to have a marked effect on the patterns when considering the average data: the left and the right panels of Fig. 3 are hardly distinguishable.

The data and the subsequent analyses show that loudness is not constant over sound incidence angles, and the directional loudness-sensitivity patterns change considerably with center frequency, and to a lesser extent, with overall sound-pressure level.

3. Statistical analysis

The significance of the effects observed in the averaged data was confirmed by a $7\times3\times2$ (directions $\times$ center frequencies $\times$ levels) repeated-measures analysis of variance (ANOVA) on the means obtained from each subject in each of the experimental conditions.

In addition to a significant main effect of direction, $F(6,42)=28.35$, $p<0.001$, indicating that directional loudness-sensitivity differences persist, even when collapsing across levels and frequencies, all its interactions were highly significant:

(1) As expected, the direction $\times$ frequency interaction produced the highest F value, $F(12,84)=31.29$, $p<0.001$, confirming that the way in which directional loudness varies is strongly frequency dependent (see Fig. 3). It should be noted that this interaction is also highly significant for each of the eight subjects when statistical analyses are done individually.
(2) Furthermore, there is a significant direction $\times$ level interaction in the pooled data, $F(6,42)=7.29$, $p<0.001$. Inspecting the average directional loudness sensitivities in Fig. 3, it appears that—ignoring center frequency—the directional effects on loudness are slightly more pronounced at the higher overall level (65 dB SPL).
(3) More importantly, there is a three-way (direction $\times$ frequency $\times$ level) interaction, indicating that the frequency-dependent directional effects show a different pattern for the two overall levels, $F(12,84)=7.42$, $p<0.001$. This appears to be largely due to the 5-kHz data showing a slightly larger gain in sensitivity in front of the listener, and a slightly larger loss behind when comparing the high with the low overall level (see Fig. 3). Again, this interaction is significant for all of the eight subjects, even though the patterns show strong individual differences (see Fig. 2).

B. Head-related transfer functions

Individual head-related transfer functions were measured to investigate how sound is being filtered from a free field to the subjects' ears, depending on the angle of incidence. As an example, the HRTF magnitude spectra for subject IA from all eight directions are plotted in Fig. 4. Each panel depicts curves for the three separate sets of measurements made at different stages of the experiment. These mea-

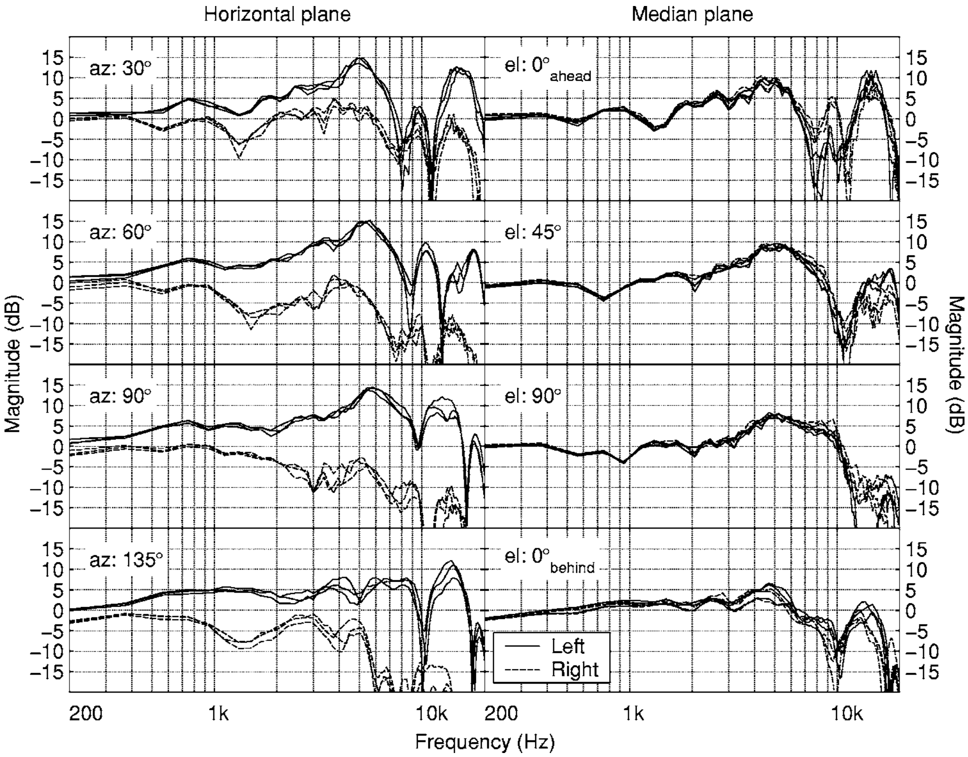

FIG. 4. Three HRTF measurements, performed at different stages of the experiment. The figure shows data for the left and right ears of a single subject (IA), for stimulation from all eight directions. The left panel depicts measurements obtained in the horizontal plane (azimuths of 30°, 60°, 90°, and 135°), the right panel those obtained in the median plane (elevations of 0°_{ahead}, 45°, 90°, and $0^{\circ}_{\text{behind}}$). 0°_{ahead} incidence is the frontal (reference) direction.

surements include individual fitting and positioning of the microphones, aligning of the subjects to the listening position, calibration, and acoustic measurements. As seen in Fig. 4, the measurements are highly repeatable, the variation below 1 kHz on average being within ±0.4 to ±0.6 dB (comparable to e.g., Møller *et al.*, 1995).

Figure 4 also shows that the interaural level differences in the median-plane HRTFs are very small up to around 7 kHz. In the horizontal plane, however, HRTFs of the left and right ears differ considerably due to a pressure buildup at the ipsilateral ear and head shadowing at the contralateral ear, especially at high frequencies. For the fairly representative subject whose HRTFs are depicted in Fig. 4, the maximum magnitudes of the ipsilateral (left) ear in the horizontal plane are around 15 dB for azimuths from 30° to 90° (front-left side), while the magnitudes at the contralateral (right) ear are typically below 0 dB.

C. HRTFs and directional loudness

1. Calculating normalized at-ear exposure

In order to investigate the effects of the physical HRTF filtering on the directional loudness matches on an individual basis, the objective HRTF measurements and the subjective loudness data were combined. This was done in order to obtain the actual frequency-specific at-ear exposure, and to evaluate whether the peculiarities of individual HRTFs might account for some of the interindividual variation seen in the directional loudness matches. Note that this analysis was based on the magnitude spectra of the HRTFs, and that the effect of the interaural time difference was disregarded.

The individual HRTFs were averaged across the three repetitions by calculating the mean of the linear magnitude spectra. These means were then converted to the corresponding third-octave-band levels in decibels. Finally, the left- and right-ear SPLs were normalized, for each incidence angle and at each center frequency, by subtracting the respective frontal left- and right-ear levels from them, since the loudness matches were always made to the frontal reference.

2. Relating loudness matches to HTRFs

For each of the eight participants, the normalized at-ear levels and directional loudness matches are combined in Fig. 5 and Fig. 6.

a. Horizontal plane. The combined data for the horizontal plane are plotted in Fig. 5. As seen in Fig. 5, in this plane the individual ILDs reach a maximum of 5 dB at 0.4 kHz, of 12 dB at 1 kHz, and of up to 30 dB at 5 kHz for the calculated third-octave-band at-ear SPLs.

For all subjects, except for subject IA at 0.4 kHz, the subjective directional loudness sensitivities at the high and low overall levels largely fall between the objective at-ear sound-pressure levels. It thus seems that the agreement between the two types of data is fair: For example, by considering the 5-kHz data for subjects TB and WE in Fig. 5, the idiosyncrasies in their at-ear SPLs are reflected in equally individual directional loudness sensitivities. However, the picture is not as clear when considering the two overall levels ("high" at 65 dB SPL and "low" at 45 dB SPL): Generally, the subjective data at the two overall levels are fairly congruous. In some cases, however, the most extreme case being subject SC at 5 kHz in Fig. 5, a clear overall level dependence can be observed.

If loudness were perceived as being constant over sound incidence angles, the subjective directional sensitivity data would follow the 0-dB horizontal in Fig. 5 or, equivalently, the 0-dB circles in Figs. 2 and 3. That would imply loudness

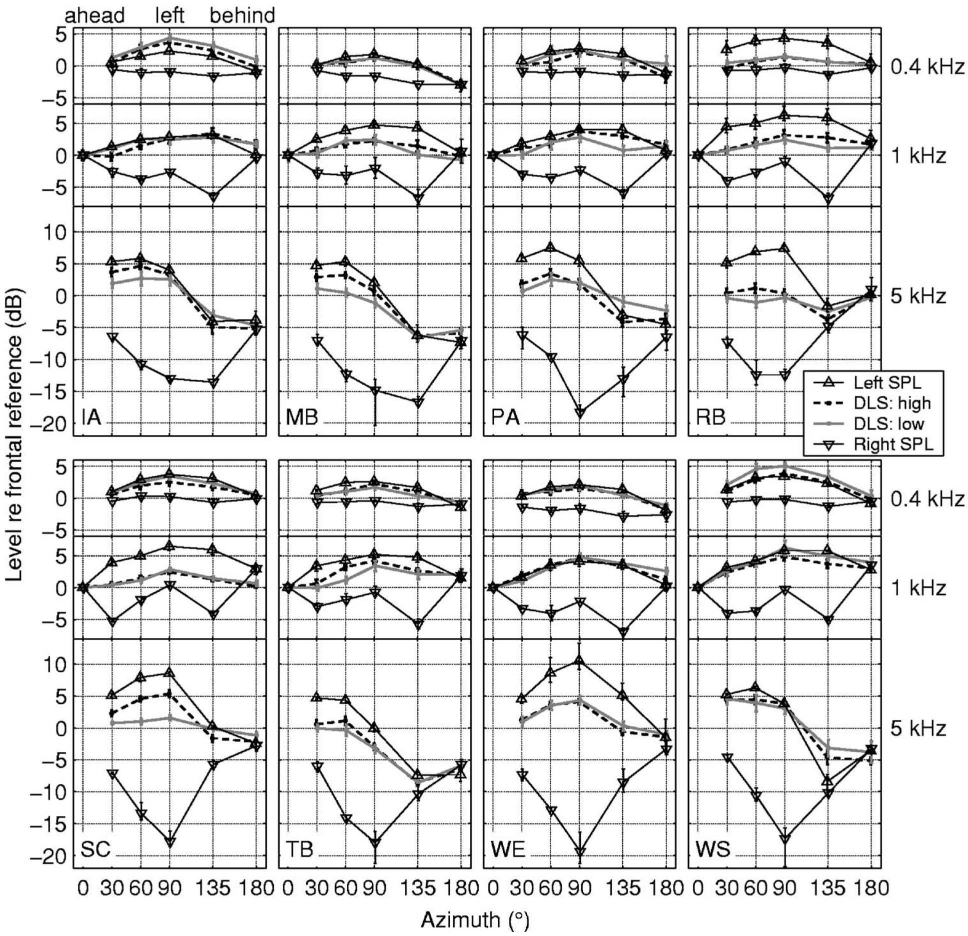

FIG. 5. Horizontal plane: Directional loudness sensitivities (DLS) at the high and low overalls level (DLS high: 65 dB SPL and DLS low: 45 dB SPL) with 95% confidence intervals, along with left- and right-ear sound-pressure levels, plotted relative to the frontal reference (see the text).

to be governed solely by the sound-pressure level of the source measured in the absence of a listener, irrespective of the changes in the at-ear sound-pressure levels as a function of sound incidence angle. This does not seem to be the case for any of the data sets.

If, on the other hand, the subjective loudness data always followed the ear with the higher SPL, this would imply *no* binaural loudness summation, i.e., loudness would be determined by the ear getting the higher input alone. Evidence for this kind of behavior may be seen in the data of IA, WS, and to some extent in those of WE and PA, though not at 5 kHz.

b. Median plane. In the median plane the ILDs are small, and the two ears are getting approximately the same

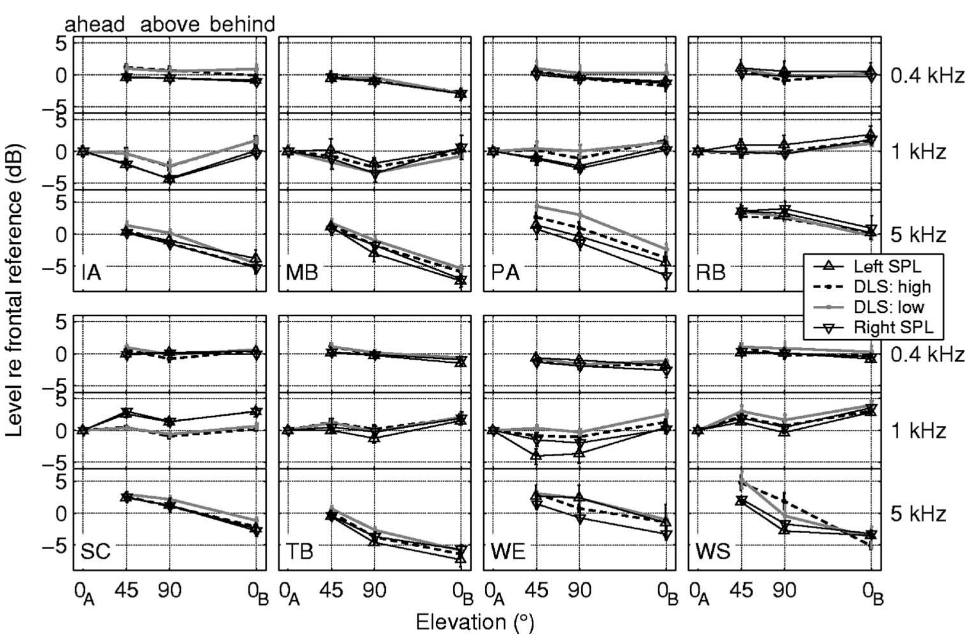

FIG. 6. Median plane: Directional loudness sensitivities (DLS) at the high and low overall levels (DLS high: 65 dB SPL and DLS low: 45 dB SPL) with 95% confidence intervals, along with left- and right-ear sound-pressure levels, plotted relative to the frontal reference (see the text). Elevations 0°_{A} and 0°_{B} are ahead and behind the listener, respectively.

input at all sound incidence angles; see Fig. 6. The differences between the ears are largest for subject WE, producing ILDs as large as 3 dB.

The normalized at-ear levels as a function of direction vary over less than 3 dB at 0.4 kHz, by up to 5 dB at 1 kHz, and over a range of almost 10 dB at 5 kHz. In this plane a change in the at-ear SPLs with incidence angle should presumably be reflected in a similar change in directional loudness sensitivity, which is true for most of the subjects. Occasional exceptions from this rule can be seen, however, for example for subjects SC and IA at 1 kHz, and for subject WS at 5 kHz.

c. Summary. Both in the horizontal and median planes, the patterns of the individual directional loudness sensitivities can largely be explained by directional effects the individual HRTFs have on at-ear sound-pressure levels. The way the subjects combine their left- and right-ear SPLs to a single loudness percept is further explored in the next section concerned with modeling binaural loudness.

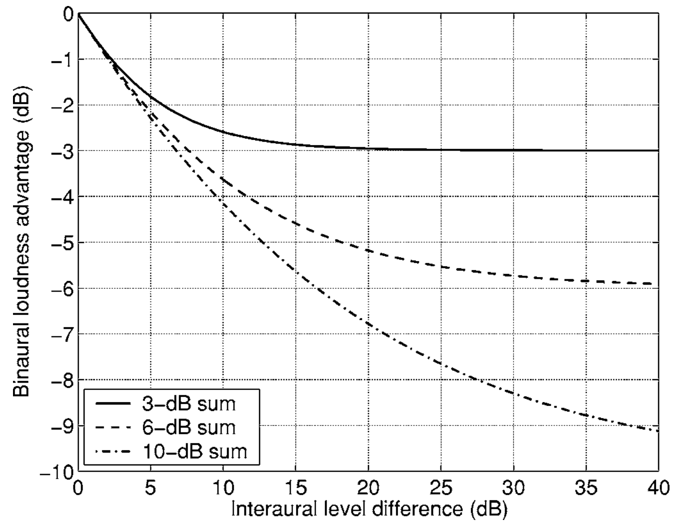

FIG. 7. Binaural loudness advantage as a function of interaural level difference; solid: 3-dB, dashed: 6-dB, and dash-dotted line: 10-dB summation rule.

IV. MODELING OF BINAURAL LOUDNESS

Large interindividual variation was found in subjects' directional loudness sensitivities. As seen in the previous section, these sensitivities exhibit systematic dependencies on the directional variations in individual HRTFs. Thus, a straightforward strategy in modeling binaural loudness is to take the HRTF effects into account, and to relate the physical changes in the at-ear signals—independent of direction—to the corresponding changes in loudness as perceived in a real sound field.

In the median plane, where the loudspeakers were positioned symmetrically with respect to the subjects' left and right ears, the listening situation was close to diotic. In this plane, the sound-pressure levels at the two ears were similar at the elevations under investigation (0°_{A}, 45°, 90°, and 0°_{B}); see Fig. 6. In such a situation, the actual amount of summation across the two ears has no effect on binaural modeling. This is due to the fact that the same binaural listening advantage takes effect both for the reference and the comparison to be matched. Note that the same applies for the traditional free- and diffuse-field loudness paradigms.

Dichotic stimulation, with different at-ear levels, thus constitutes the most interesting case for the modeling of binaural loudness. Dichotic at-ear SPLs were observed for the azimuths of 30°, 60°, 90°, and 135° in the horizontal plane (see Fig. 5). At these azimuths subjects typically had to match a dichotic sound to the diotic frontal reference.

Narrow-band stimuli were used in the listening experiment in order to simplify the modeling of binaural loudness, by being able to ignore spectral summation of loudness across critical bands. Also, assuming that perceived loudness is doubled when the listening is binaural (diotic) instead of monaural, a relationship between the psychophysical dimension of loudness (as measured in sones) and its physical correlate, the sound-pressure level (in dB SPL) can be established. By definition, a loudness of 1 sone is produced by a 40-dB SPL, 1-kHz tone, and doubling or halving loudness (in sones) corresponds to a 10-dB increment or decrement in sound-pressure level, respectively. Due to the shape of the equal-loudness contours (ISO 226, 2003), the increments within the range of sound-pressure levels used in the present experiment are approximately 10.5 and 9.5 dB SPL at 0.4 and 5 kHz, respectively, for a doubling of loudness. Thus, at all three center frequencies (0.4, 1, and 5 kHz) used in the present study, doubling in sones corresponds fairly closely to a 10-dB gain in sound-pressure level.

In order to illustrate how binaural loudness is affected by various interaural level differences, theoretical curves can be obtained utilizing Eq. (2), taken from Robinson and Whittle (1960). It is reasonable to assume that the summation of sound-pressure levels across the two ears is nonlinear, as suggested by Eq. (2): At large ILDs, the ear receiving the lower sound-pressure level presumably has little effect on overall binaural loudness, and the stimulation is effectively monaural. When approaching a diotic situation, however, the signals at the two ears tend to be weighted equally in contributing to overall loudness.

Theoretical curves for three hypothetical binaural loudness-summation rules are plotted as a function of the interaural level difference in Fig. 7. In addition to the 6-dB summation rule adopted from Robinson and Whittle (1960), two other curves were derived by changing the binaural gain factor in Eq. (2): A 3-dB summation rule corresponding to the "power summation" of the linear at-ear magnitude spectra, and a 10-dB summation rule, which for the stimuli used in the present study roughly corresponds to perfect binaural summation in sones.

The different curves in Fig. 7 are normalized so that they all coincide in the origin of the graph: it represents the diotic case with an ILD of zero. As the ILD increases, loudness decreases by different amounts, depending on the summation rule with the "loss" to be read from the ordinate corresponding to the "binaural loudness advantage" achievable by switching from dichotic to diotic stimulation. The 3-dB summation rule fairly quickly converges to the −3-dB level in the graph: when the ILD increases beyond 15 dB, binaural loud-

TABLE I. Least-squares estimates for the amount of binaural loudness summation [x in Eq. (3), in dB], at the three center frequencies (f_c: 0.4, 1, and 5 kHz) at high (65 dB SPL) and low (45 dB SPL) overall levels. The two right-most columns show the best fits when pooling center frequencies, and the best fits across center frequencies when the data are averaged across subjects (bottom row). Extreme values are marked with stars (see the text for details).

	f_c							
	0.4 kHz		1 kHz		5 kHz		Best fit across f_c	
Subject	High	Low	High	Low	High	Low	High	Low
IA	0.1	0.1	0.1	0.1	0.7	1.5	0.1	0.1
MB	0.3	0.4	3.3	3.5	2.2	4.8	2.4	4.6
PA	1.0	0.4	0.7	1.3	3.1	3.0	2.1	2.4
RB	99.9*	99.9*	4.0	9.3	10.0	18.4*	9.1	17.6*
SC	1.6	0.3	13.1	29.7*	3.1	8.2	3.8	8.5
TB	1.1	2.8	2.1	3.0	3.7	5.0	3.3	4.6
WE	0.1	0.1	0.1	0.3	4.9	4.5	2.8	2.6
WS	0.1	0.1	1.7	0.5	0.1	0.1	0.7	0.1
Median	0.7	0.4	1.9	2.2	3.1	4.7	2.6	3.6
Averaged data							2.6	3.9

ness is no longer affected. At these ILDs loudness is determined by the ear with the higher sound-pressure level alone, and dichotic loudness is 3 dB lower than the corresponding diotic one. With the 6- or 10-dB summation rules, much larger ILDs are required until the curve asymptotes at −6 and −10 dB, respectively. For the 10-dB summation rule, at an ILD of 40 dB (far exceeding the ILDs observed in the present study) binaural loudness still continues to decrease.

A. Individual data

The third-octave-band at-ear sound-pressure levels computed from the HRTFs were used in the modeling, in order to find the best-fitting binaural summation rule to predict the directional loudness-sensitivity data. Robinson and Whittle (1960) reported their average data to support a 6-dB loudness summation across their listeners' ears [see Eq. (2)]. This type of modeling was explored for the present data, but on an individual basis. The modeling was carried out by relaxing the factor 6 in Eq. (2).

To that effect, the optimal amount of binaural loudness summation (x)—assumed to be fixed at 6 dB in Eq. (2)—was *estimated* by minimizing the sum-of-squares of the errors (SSE) between the actual directional loudness sensitivity (DLS) and the sensitivity predicted (L_{mon}) from the changes in at-ear sound-pressure levels using Eq. (3). All 16 (j) repetitions of each condition, and the mean at-ear sound-pressure levels for each of the four horizontal-plane angles of incidence (i; 30°, 60°, 90°, and 135°) were included in the modeling, which was performed individually for each subject, and separately for the three center frequencies and the two overall levels.

$$\mathrm{SSE} = \sum_{i=1}^{4} \sum_{j=1}^{16} \{\mathrm{DLS}_{\mathrm{high/low}_{i,j}} - [L_{\mathrm{mon,comp}_i}(x) - L_{\mathrm{mon,ref}}(x)]\}^2, \quad (3)$$

where

$$L_{\mathrm{mon,comp}_i}(x) = x \times \log_2(2^{L_{\mathrm{left,comp}_i}/x} + 2^{L_{\mathrm{right,comp}_i}/x}), \quad (4)$$

and

$$L_{\mathrm{mon,ref}}(x) = x \times \log_2(2^{L_{\mathrm{left,ref}}/x} + 2^{L_{\mathrm{right,ref}}/x}). \quad (5)$$

In these equations, $L_{\mathrm{left,comp}}$ and $L_{\mathrm{right,comp}}$ refer to the third-octave-band levels for the comparison incidence calculated from the individual left- and right-ear HRTFs, respectively. Likewise, $L_{\mathrm{left,ref}}$ and $L_{\mathrm{right,ref}}$ refer to the corresponding levels for the frontal reference at the left and right ears, respectively.

The subjective directional loudness sensitivities had been normalized to the frontal reference (see Figs. 2 and 5). Therefore, the predictions were normalized as well by subtracting Eq. (5) from Eq. (4) in the minimization of the sum of squares of the errors. Due to this normalization, the overall level (65 vs 45 dB SPL) does not have an influence on the predictions. The possible dependence of binaural loudness summation may nevertheless show up in the subjective directional loudness sensitivities at the high and low overall levels, and may thus influence the estimate of the variable x, the binaural gain estimated from the data. Forty-eight such estimates (for eight subjects, three center frequencies, and two overall levels) for the amount of binaural loudness summation are listed in Table I. The minimization algorithm was limited to a summation value between 0.1 and 99.9 dB.

As was already seen in Fig. 5, the amount of binaural loudness summation varies greatly across subjects, and also within subjects across the three center frequencies. The best-fitting binaural gain estimates roughly fall into three categories: The summation is minor (less than 1 dB) for 19, moderate (from 1 to 10 dB) for 24, and extreme (greater than 10 dB) for 5 out of the 48 cases analyzed. There is a tendency for the summation values to increase with center frequency, but due to the fact that the center frequencies are confounded with variations in ILDs, the comparison may not be fair.

The smaller the amount of binaural loudness summation, the more binaural loudness is determined by the ear getting a higher input. By contrast, the higher the summation value, the more influence the ear receiving the lower sound-pressure level has on binaural loudness. Some extreme values marked with stars in Table I, e.g., subject RB at 0.4 kHz, seem to imply the latter behavior. Closer inspection of Fig. 5, however, reveals that for this subject the directional loudness sensitivity remains close to the 0-dB line, even if the at-ear sound-pressure levels vary over a fairly wide range. As Robinson and Whittle (1960) pointed out, the actual value of the summation parameter (at the natural ILDs in question) does not have a great effect on the directional loudness sensitivities predicted from the at-ear SPLs. For these reasons the minimization algorithm can reach very high summation values (up to the limit of 99.9 dB) when searching for the best fit. However, it is unrealistic that the binaural gain (i.e., the loudness match between monotic and diotic stimulation) for a normal-hearing subject is much larger than 10 dB.

To get a more stable estimate, the amount of binaural summation was also determined by pooling across the three center frequencies; see the two right-most columns in Table I. This was achieved by aggregating the data across center frequencies, and finding the best-fitting summation rule to the aggregated data set. The individual differences are still retained, and the summation values again fall into the three categories defined above.

In order to deal with the variance inherent in the subjective data, a partial F-test (Bates and Watts, 1988, Chap. 3) was performed to investigate whether the subjects summed their at-ear levels in significantly different ways. In a "restricted" model *one* least-squares fit of binaural loudness summation [x in Eq. (3)] common to all subjects was estimated, whereas in a "full" model the summation value was relaxed to estimate *different* parameters for the eight subjects. The data were aggregated across incidence angles, overall levels and center frequencies. The partial F-test showed that the error sum of squares between the subjective data and the estimate was significantly larger for the restricted model having a common parameter for all subjects [$F(7,3064)=211.58$; $p<0.001$]. Therefore, the full model allowing for individually different binaural-gain parameters predicted the data better, and hence, the differences in the way the subjects summed the at-ear levels appear to be significant.

B. Group data

The individual third-octave-band HRTFs and directional loudness sensitivities were averaged across subjects, to make an estimate for the mean data thus obtained. Aggregating over center frequencies, as before, the best fits for the averaged data came fairly close to suggesting a 3-dB summation rule both at the high and the low overall level (see the bottom row of Table I).

Thus far the prediction was entirely based on the at-ear sound-pressure levels at the center frequency of the narrow-band noises used. However, by using a loudness model, the possible spread of excitation to neighboring critical bands can be taken into account in the modeling. Furthermore, given a relatively large dynamic range, the shape of the loudness function may be better accounted for when using a loudness model.

Therefore, the most widely accepted loudness model by Moore *et al.* (1997) was tested in predicting the present data. This model facilitates the use of eardrum pressures for loudness computations, i.e., using at-ear signals as a product of the HRTF-filtering stage. The model also predicts monaural loudness, by assuming perfect loudness summation in sones between the two ears, and calculating monaural loudness simply as being one half of the binaural, diotic loudness. Dichotic loudness can then be computed as a sum of the two monaural loudness values in sones.

Since the HRTFs of the present study had been measured at the entrance to the blocked ear canal, a direction-independent transfer from the measurement point to the eardrum (mean P_4/P_2) was adopted from Fig. 13 in Hammershøi and Møller (1996). In contrast to the summation rule explored in the previous section, here absolute binaural loudness values were computed. The effects of the HRTFs were taken into account, as before, but now the entire at-ear spectra were included (instead of only using the level at the center frequency). The input data to the loudness model thus were third-octave-band spectra based on the measured stimulus spectrum in the absence of a listener (P_1), combined with the left- and right-ear HRTFs (P_2/P_1), and corrected by the eardrum-to-the-measurement-point transfer function (P_4/P_2).

Monaural loudness values were computed for (dichotic) left- and right-ear signals, subsequently summed, and compared to the loudness produced by the (close to) diotic frontal reference. First, binaural loudness values for each of the frontal reference stimuli were computed, as described above. Then, sone values for the comparison directions were computed by varying the level of the P_1 pressures, within the range of ±10 dB from the frontal reference level, in steps of 0.5 dB. The P_1 sound-pressure levels yielding the binaural loudness values closest to that of the frontal reference were selected. In this way the loudness model was used to find equal-loudness sound-pressure levels for each incidence angle, including the effects of the HRTFs. The inverses of these sound-pressure levels relative to the frontal reference were taken as the directional loudness sensitivities predicted by the model.[2]

Figure 8 contrasts the predictions made by the loudness model (Moore *et al.*, 1997) with the 3-dB power summation, which fared best in the earlier analysis. Since the effect of overall sound-pressure level on directional loudness was minor for the averaged data (see Fig. 3), only the high-level (65 dB SPL) directional sensitivities are plotted.

For all dichotic situations (horizontal plane, left column in Fig. 8), the 3-dB summation rule predicts the obtained mean loudness-sensitivity data quite well. At each center frequency, the patterns of the 3-dB prediction and the actual matches made are congruous, and only in two instances (at 0.4 kHz, azimuths of 90° and 135° in Fig. 8) do the 95% confidence intervals of the subjective data not include the 3-dB prediction. By contrast, the prediction of the loudness

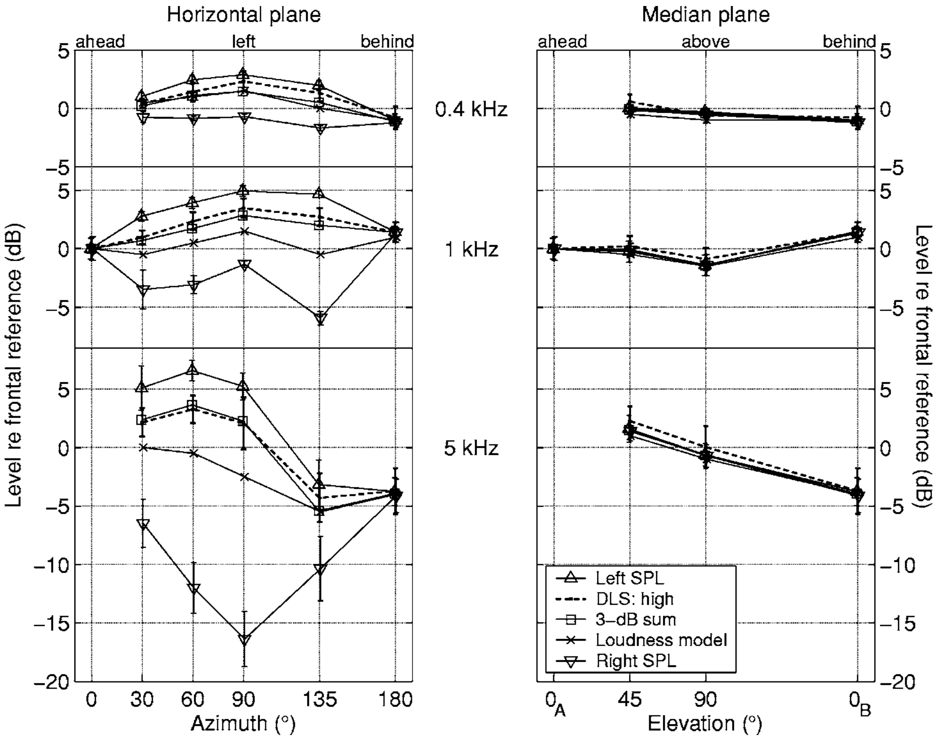

FIG. 8. Average left- and right-ear sound-pressure levels, a 3-dB summation rule, loudness summation in sones, and obtained average DLS at the high overall level (65 dB SPL). The error bars denote the 95% confidence intervals of the means across the eight subjects. The left panel depicts data and predictions for the horizontal plane, the right panel for the median plane. Elevations 0°_A and 0°_B are ahead and behind the listener, respectively.

model markedly deviates from the obtained directional loudness sensitivities, particularly at the two higher center frequencies (1 and 5 kHz). These are the situations in which the interaural level differences range from 6 to over 20 dB. For these ILDs, the prediction is not bracketed by the confidence intervals of the data for seven out of the eight dichotic conditions, the difference between data and predictions reaching up to 5 dB (5 kHz, azimuth 90° in Fig. 8). It thus seems that the 3-dB summation rule of at-ear sound-pressure levels predicts the directional loudness of dichotic sounds considerably better than the assumption of perfect binaural loudness summation in sones.

In the median plane, all five curves (at-ear levels, directional loudness sensitivities, and model predictions) are nearly indistinguishable; see the right panels in Fig. 8. The 95% confidence intervals of the subjective data include both the physical changes in left- and right-ear sound-pressure levels, and the predicions of 3-dB sum and loudness summation in sones. Obviously, the diotic stimulation condition does not provide a critical test for these models.

V. DISCUSSION

A. Comparison with previous work

When comparing the present results to the work of Robinson and Whittle (1960), it may be observed that the average directional effect sizes they obtained are comparable to those measured in the present study: For the incidence angles presented here, the average directional loudness sensitivities Robinson and Whittle (1960) obtained at center frequencies below 6.4 kHz varied from −6.5 to +5.0 dB (see their Fig. 2) relative to the frontal reference level. The corresponding range for the average data in the present study is −4.3 to +3.5 dB (see Fig. 3), although the actual stimulus center frequencies used differed somewhat between the two investigations.

In both investigations, direction had a smaller effect on loudness at lower center frequencies, and the effect increased with stimulus center frequency. Qualitatively, this can be explained by the fact that with increasing frequency the physical dimensions of a listener start to obstruct the sound field. The obstruction also becomes more direction dependent at higher frequencies (as can be seen in the sample HRTFs plotted in Fig. 4), and this is reflected in its increasing effect on the directional loudness sensitivities.

The present empirical data collection, however, goes beyond previous work by reporting individual analyses. Consequently, and as expected from research on HRTFs, idiosyncratic directional loudness-sensitivity patterns were found. The individual data also showed that all participants were highly consistent in their judgments, even though the loudness of two sounds coming from different directions, and typically having different timbres, had to be compared.

The consistency in the participants' directional loudness matches provided considerable statistical power. On the one hand, that means that the significance of the major frequency-dependent effects of the direction of incidence on perceived loudness may be ascertained with great confidence. On the other hand, that entails that even small effects on the range of 1–2 dB level will emerge as statistically significant, and thus require further interpretation. That is the case for the effects of overall presentation level, [3] and its interaction with the directional and frequency-specific effects.

Comparison of both individual data (e.g., Fig. 2, top row) and of the group averages (Fig. 3) shows a tendency for the frequency-dependent directional effects to become more pronounced with increasing level. Likewise, small but systematic level effects are found when trying to estimate the amount of binaural gain from the data (Table I). Contrary to what is reported in the literature (Shaw *et al.*, 1947; Reynolds and Stevens, 1960; Hellman and Zwislocki, 1963; Scharf and Fishken, 1970), this gain appears to be smaller at the higher overall level. That may be due to the low-level directional sensitivities being less affected by the ear receiving the greater input than the high-level directional sensitivities (see Fig. 5). Due to the small magnitudes of the overall-level effects, the present authors consider them to be negligible for most practical purposes, at least in the midlevel range investigated here (45–65 dB SPL).

Furthermore, the relatively low binaural-gain parameter derived from the present data is in conflict with the outcome of most of the classical studies (such as Reynolds and Stevens, 1960; Hellman and Zwislocki, 1963; Marks, 1978; Zwicker and Zwicker, 1991, among others) employing headphones, and largely focusing on monotic-to-diotic comparisons. But, note that—apart from other methodological distinctions—a key feature of these earlier studies is that signals may have been generated that would never reach the two ears when being emitted by a real source positioned in space, and fail to produce an externalized auditory event. It is unclear whether the results of the two paradigms (binaural loudness summation versus directional loudness) can be compared directly, since the auditory events produced are so drastically different. The directional loudness paradigm, however, is not only closer to "real-world" stimulation, but also to the application of measuring sound fields using a dummy head, where the signals at the ears of the dummy are due to the physical obstruction in the sound field.

B. Individual differences

Even though tentative general conclusions on the computation of binaural loudness may be drawn from the present data, it is striking how large the interindividual differences in loudness matches (see Figs. 2 and 5), and hence, in directional loudness sensitivity are when comparing the eight listeners participating. The original hope, that *all* of this interindividual variance might be accounted for by the equally large differences in individual HRTFs (e.g., Fig. 4) does not seem to be warranted, as is evident from our analysis of individual "summation rules" displayed in Table I. Obviously, using the actual at-ear sound-pressure levels rather than the levels emitted by the loudspeakers in the analysis still leaves us with considerable residual individual variance.

Several potential reasons for that variance might be explored: An obvious reason may be that the third-octave-band levels derived from the HRTF measurements do not reflect the actual at-ear stimulation well enough. However, the quality of the HRTFs may be examined by contrasting the present results with data obtained in individual binaural synthesis where the directional sound sources are recreated via virtual acoustics, the crucial difference being that the at-ear levels are precisely known in that situation. Performing such an experiment with six listeners from the original sample of eight (Sivonen *et al.*, 2005), we found no appreciable, or statistically significant, differences between the two sets of data (real vs virtual sound field). Rather, the individual differences remained, leading us to look for factors other than differences in the physical shape of pinnae, heads, and torsos.

A more speculative explanation for the individual differences found might be that the participants exhibited different degrees of "loudness constancy" in our experimental setup. The notion of "perceptual constancy" refers to situations in which a percept remains constant despite profound changes in the physical stimulation affecting the sensory receptors (Zahorik and Wightman, 2001). Typically, loudness constancy is observed when the loudness of a source (e.g., a musical instrument, a human voice) is judged to remain constant, even though its distance to the observer is varied. Stretching this notion somewhat, we might also speak of loudness constancy when listeners judge sounds to be equally loud, despite variations in their angle of incidence (which greatly affects the at-ear stimulation). It might be speculated that observers have learned to deconvolve the signals with their HRTFs in order to infer the loudness at the source.

Do the present data show evidence for loudness constancy defined in this way? The answer is clearly negative: Note that perfect constancy would mean that all of the identical-distance, identical-power sources used in the present experiment should be judged to be equally loud, i.e., the matches should fall on the 0-dB (reference) circle in Fig. 2, or on the 0-dB horizontal in Figs. 5 and 6. That, obviously, is not the case. Nevertheless, subjects might have a tendency to preserve constancy to varying degrees, thus producing different amounts of bias towards the zero line. Potentially, they could do so by using the localization and timbre cues available, as well as the fact that the loudspeakers producing the sounds are in plain view.

The constancy problem is related to that of the "listening attitude" a participant might adopt: In a pioneering investigation of loudness constancy (Mohrmann, 1939), this was operationalized as judging hidden sources at various distances while either adopting a sender attitude ("Sendereinstellung;" p. 155), or a receiver attitude ("Empfangseinstellung") which yielded appreciably different results. In modern terminology one would refer to judging the distal stimulus vs the proximal stimulus, and in the present situation that would be equivalent to judging the sound power of the loudspeaker as opposed to judging how it affects the listener. It is unclear, however, whether subjects can make that distinction in an anechoic situation, and the present authors know of no published reports implementing the instructional variations required.

Nevertheless, it may safely be said that a "bias" towards constancy can only play a minor role in accounting for the present data. The fact that knowing the individual HRTFs goes such a long way in accounting for the idiosyncrasies seen in the matches argues against constancy being a major factor in directional loudness perception, at least for the synthetic sounds and the anechoic environment studied here.

VI. CONCLUSIONS

(1) Loudness matches obtained with narrow-band noises in an anechoic environment showed that loudness is not constant over sound incidence angles. Rather, directional loudness sensitivities varied by up to 10 dB in individual, and up to 8 dB in averaged data.
(2) The directional effects on loudness showed considerable dependency on center frequency, with greater directional effects being observed at higher center frequencies, and to some extent on the overall sound-pressure level of the stimuli.
(3) Large, but highly reliable individual differences in directional loudness perception were observed.
(4) The individual patterns of directional loudness could largely be accounted for by the corresponding changes in physical stimulation, as determined by head-related transfer functions (HRTFs).
(5) These transfer functions were utilized for modeling binaural loudness based on the at-ear sound-pressure levels encountered. A 3-dB binaural loudness-summation ("power-summation") rule predicted the obtained mean data best, but sizable interindividual differences remained, even after the effect of individual HRTFs was taken into account.
(6) The results can be used for predicting loudness in any type of sound field (be it free, diffuse, or directional, resulting in diotic or dichotic at-ear signals) using a dummy head.

ACKNOWLEDGMENTS

The authors would like to thank Claus Vestergaard Skipper for his help in setting up the apparatus used in the present experiment. Henrik Møller and Dorte Hammershøi are thanked for their advice on how to carry out HRTF measurements. Pauli Minnaar helped in analyzing the HRTFs and with including them in the modeling, Florian Wickelmaier advised us with the statistical analyses. We would further like to thank Brian C.J. Moore for giving us access to and advice in utilizing his loudness model. Finally, we would like to thank the participants for their patience and devotion donated to extensive and at times exhaustive listening tests. This research was carried out as part of the "Centercontract 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).

[1]Directional loudness sensitivities are thus defined in loose analogy to the directivity characteristics of microphones (Beranek, 1986, Chap. 6). Being the inverse of the relative sound-pressure level required to produce a loudness match, they are—despite the similarity in terms—not related to sensitivity parameters (such as d') as conceptualized in signal detection theory (Green and Swets, 1988).

[2]These predictions were made both for the individual and the mean data, essentially yielding the same conclusions. Thus, only the results for the mean data are presented here.

[3]Even though a 20-dB range may not appear sufficient to investigate the effects of overall level, note that when considering the extra headroom required for the adaptive starting values (±10 dB), and HRTF effects boosting or attenuating levels by approximately the same amount, the effective range listeners were exposed to in the experiment was quite large, covering what can be handled in a loudness-matching experiment without encountering floor ("too soft") or ceiling ("too annoying") problems.

Bates, D. M., and Watts, D. G. (**1988**). *Nonlinear Regression Analysis and Its Applications* (Wiley, New York).
Beranek, L. L. (**1986**). *Acoustics* (American Institute of Physics, New York).
Blauert, J. (**1997**). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA).
Bodden, M. (**1997**). "Instrumentation for sound quality evaluation," Acust. Acta Acust. 83, 775–783.
Bronkhorst, A. W. (**1995**). "Localization of real and virtual sound sources," J. Acoust. Soc. Am. **98**, 2542–2553.
Green, D. M., and Swets, J. A. (**1988**). *Signal Detection Theory and Psychophysics* (Peninsula, Los Altos, CA).
Hammershøi, D., and Møller, H. (**1996**). "Sound transmission to and within the human ear canal," J. Acoust. Soc. Am. **100**, 408–427.
Hellman, R. P., and Zwislocki, J. (**1963**). "Monaural loudness function at 1000 cps and interaural summation," J. Acoust. Soc. Am. **35**, 856–865.
ISO 226 (**2003**). "Acoustics—Normal equal-loudness-level contours," (International Organization for Standardization, Geneva, Switzerland).
ISO 389-1 (**1998**). "Acoustics—Reference zero for the calibration of audiometric equipment—Part 1: Reference equivalent threshold sound pressure levels for pure tones and supra-aural earphones," (International Organization for Standardization, Geneva, Switzerland).
ISO 389-7 (**1996**). "Acoustics—Reference zero for the calibration of audiometic equipment—Part 7: Reference threshold of hearing under free-field and diffuse-field listening conditions," (International Organization for Standardization, Geneva, Switzerland).
ISO 532 (**1975**). "Acoustics—Method for calculating loudness level," (International Organization for Standardization, Geneva, Switzerland).
Jesteadt, W. (**1980**). "An adaptive procedure for subjective judgments," Percept. Psychophys. **28**, 85–88.
Jørgensen, S. L. (**2002**). "MATLAB model for loudness perception—directional loudness," Unpublished masters thesis, Technical University of Denmark (DTU), Copenhagen.
Kuhl, W., and Westphal, W. (**1959**). "Unterschiede der Lautstärken in der ebenen Welle und im diffusen Schallfeld" ("Differences in loudness levels between a plane wave and a diffuse sound field,") Acustica **9**, 407–408.
Levelt, W. J. M., Riemersma, J. B., and Bunt, A. A. (**1972**). "Binaural additivity of loudness," Br. J. Math. Stat. Psychol. **25**, 51–68.
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.
Mohrmann, K. (**1939**). "Lautheitskonstanz im Eutfernungswechsel" ("Loudness constancy with changing distance,") Z. Psychol. **145**, 145–199.
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**, 224–239.
Olesen, S. K., Plogsties, J., Minnaar, P., and Møller, H. (**2000**). "An improved MLS measurement system for acquiring room impulse responses," in Proceedings of NORSIG 2000, IEEE Nordic Signal Processing Symposium, Kolmården, Sweden, 117–120.
Remmers, H., and Prante, H. (**1991**). "Untersuchung zur Richtungsabhängigkeit der Lautstärkeempfindung von breitbandigen Schallen" ("Investigation of the directional dependence of the loudness of broadband sounds"), *Fortschritte der Akustik*-Deutsche Arbeitsgemeinschaft für Akustik (DAGA), 537–540.
Reynolds, G. S., and Stevens, S. S. (**1960**). "Binaural summation of loudness," J. Acoust. Soc. Am. **32**, 1337–1344.
Robinson, D. W., and Whittle, L. S. (**1960**). "The loudness of directional sound fields," Acustica **10**, 74–80.
Robinson, D. W., Whittle, L. S., and Bowsher, J. M. (**1961**). "The loudness

of diffuse sound fields," Acustica **11**, 397–404.

Scharf, B., and Fishken, D. (**1970**). "Binaural summation of loudness: Reconsidered," J. Exp. Psychol. **86**, 374–379.

Schneider, B. A., and Cohen, A. J. (**1997**). "Binaural additivity of loudness in children and adults," Percept. Psychophys. **59**, 655–664.

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.

Shaw, W. A., Newman, E. B., and Hirsh, I. J. (**1947**). "The difference between monaural and binaural thresholds," J. Exp. Psychol. **37**, 229–242.

Sivian, L. J., and White, S. D. (**1933**). "On minimum audible sound fields," J. Acoust. Soc. Am. **4**, 288–321.

Sivonen, V. P., Minnaar, P., and Ellermeier, W. (**2005**). "Effect of direction on loudness in individual binaural synthesis," in Proceedings of the 118th Convention of the Audio Eng. Soc., Convention Paper No. 6512.

Wightman, F. L., and Kistler, D. J. (**1989a**). "Headphone simulation of free field listening. I. Stimulus synthesis," J. Acoust. Soc. Am. **85**, 858–867.

Wightman, F. L., and Kistler, D. J. (**1989b**). "Headphone simulation of free field listening. II. Psychophysical validation," J. Acoust. Soc. Am. **85**, 868–878.

Zacharov, N., Tuomi, O., and Lorho, G. (**2001**). "Auditory periphery, HRTF's and directional loudness perception," in Proceedings of the 110th Convention of the Audio Eng. Soc., Convention Paper No. 5315.

Zahorik, P., and Wightman, F. L. (**2001**). "Loudness constancy with varying sound source distance," Nat. Neurosci. **4**, 78–83.

Zwicker, E., and Zwicker, U. T. (**1991**). "Dependence of binaural loudness summation on interaural level differences, spectral distribution, and temporal distribution," J. Acoust. Soc. Am. **89**, 756–764.