Sound localization in individualized and non-individualized crosstalk cancellation systems

Piotr Majdak[a)]
Acoustics Research Institute, Austrian Academy of Sciences, A-1040 Vienna, Austria

Bruno Masiero and Janina Fels
Institute of Technical Acoustics, RWTH Aachen University, D-52056 Aachen, Germany

(Received 5 September 2012; revised 6 January 2013; accepted 29 January 2013)

The sound-source localization provided by a crosstalk cancellation (CTC) system depends on the head-related transfer functions (HRTFs) used for the CTC filter calculation. In this study, the horizontal- and sagittal-plane localization performance was investigated in humans listening to individualized matched, individualized but mismatched, and non-individualized CTC systems. The systems were simulated via headphones in a binaural virtual environment with two virtual loudspeakers spatialized in front of the listener. The individualized mismatched system was based on two different sets of listener-individual HRTFs. Both sets provided similar binaural localization performance in terms of quadrant, polar, and lateral errors. The individualized matched systems provided performance similar to that from the binaural listening. For the individualized mismatched systems, the performance deteriorated, and for the non-individualized mismatched systems (based on HRTFs from other listeners), the performance deteriorated even more. The direction-dependent analysis showed that mismatch and lack of individualization yielded a substantially degraded performance for targets placed outside of the loudspeaker span and behind the listeners, showing relevance of individualized CTC systems for those targets. Further, channel separation was calculated for different frequency ranges and is discussed in the light of its use as a predictor for the localization performance provided by a CTC system. *© 2013 Acoustical Society of America*.
[http://dx.doi.org/10.1121/1.4792355]

PACS number(s): 43.38.Md, 43.66.Qp, 43.66.Pn, 43.60.Pt [MRB] Pages: 2055–2068

I. INTRODUCTION

Sound localization is an important task for determining the direction of a sound or segregating individual sound sources in complex sound environments. While binaural disparities like interaural time and level differences (ITDs and ILDs) play an important role for sound localization in the horizontal plane (Macpherson and Middlebrooks, 2002), monaural spectral cues are known to determine the perceived sound-source position in the sagittal planes (top/down, front/back; Blauert, 1969/70). In particular, spectral cues up to 16 kHz are required for accurate sound localization in sagittal planes (e.g., Carlile and Pralong, 1994; Wightman and Kistler, 1997a; Langendijk and Bronkhorst, 2002; Best *et al.*, 2005). The spectral encoding of the spatial cues—a result of the direction-dependent filtering of the pinna, head, and torso—is described by the head-related transfer functions (HRTFs; Wightman and Kistler, 1989a) or, in particular, by their directional components, the directional transfer functions (DTFs; Middlebrooks, 1999). A monophonic signal filtered by a HRTF or DTF for a particular spatial position results in a binaural signal, which when presented to a listener via headphones, creates the impression of a virtual sound source.

When loudspeakers are used for the reproduction of a binaural signal (Bauer, 1961; Atal *et al.*, 1966), the propagation paths from the loudspeakers and the listener's ears suffer from crosstalk between the two ears (see Fig. 1). A set of *crosstalk cancellation* (CTC) filters can be used to compensate for the crosstalk. By processing the binaural signal, CTC filters generate "transaural" signals which drive the loudspeakers. The CTC filters are calculated based on the acoustic transfer functions between loudspeakers and listener's ears, i.e., the HRTFs. In a *matched* CTC system, exactly the same HRTFs are used for the filter calculation and the listening situation. The matched CTC system provides optimal cancellation and thus a good system performance is assumed (Akeroyd *et al.*, 2007). In a *mismatched* CTC system, the HRTFs do not exactly match the CTC filters and the performance is assumed to degrade. The actual localization performance of a CTC system has already been investigated in the horizontal plane (Gardner, 1997; Takeuchi *et al.*, 2001; Bai and Lee, 2006; Lentz, 2006). Furthermore, Takeuchi and Nelson (2002) collected data on localization in both horizontal and sagittal planes for mismatched non-individualized CTC systems. However, little is known about the localization performance in horizontal and sagittal planes provided by both matched and mismatched CTC systems related to the actual binaural listener-individual localization performance.

HRTFs are listener dependent (Wightman and Kistler, 1989b; Wenzel *et al.*, 1993; Møller *et al.*, 1995); Akeroyd *et al.* (2007) used, thus, HRTFs from other listeners to create mismatched CTC systems and compared their numeric performance to the matched CTC systems. In a simulation of

[a)]Author to whom correspondence should be addressed. Electronic mail: piotr@majdak.com

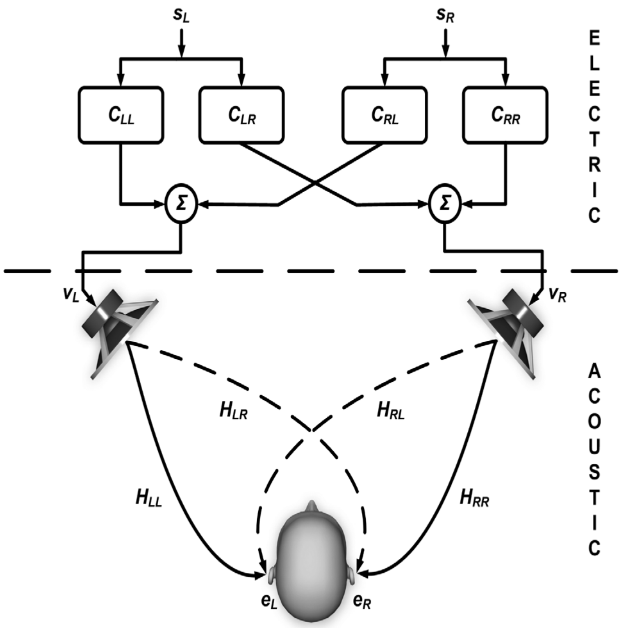

FIG. 1. Schematic representation of a binaural reproduction system via loudspeakers, i.e., CTC system. The CTC filters are shown in the upper part of the figure and acoustic paths are shown in the lower part. The solid and dashed lines show the direct and the crosstalk paths, respectively.

binaural processing, they showed disrupted ITDs and ILDs for the mismatched CTC systems. Based on their simulation results, they concluded that the mismatched system will probably yield a degraded localization performance, particularly for lateral directions.

Even though Akeroyd *et al.* (2007) used listener-individual HRTFs to create a matched CTC system, the listener-individual HRTFs do not, however, always yield a matched CTC system. For example, if the HRTF measurement is repeated for the same listener, the HRTF set will still be considered as listener-individual, but, acoustic properties of the HRTFs might slightly change, causing a mismatch to the CTC filters. This is actually a common situation even in individualized CTC systems, where the propagation paths change between the HRTF measurements and the actual use of the CTC system.

Thus, the aim of our study was to investigate two-dimensional human localization performance in CTC systems with a special focus on individualized matched, individualized but mismatched, and non-individualized CTC systems. The individualized but mismatched CTC systems used a second HRTF measurement of the same listeners. The non-individualized CTC systems used HRTFs from a mannequin and other listeners. Also, the baseline performance was acquired for binaural sound presentation without any CTC filtering.

Channel separation [(CS), see Sec. II H] is commonly used to describe the quality of a CTC system (Gardner, 1997; Bai and Lee, 2006). Akeroyd *et al.* (2007) showed much smaller channel separations in mismatched CTC systems compared to matched CTC systems. Recently, Parodi and Rubak (2011b) investigated the minimum audible channel separation in an artificial CTC system. However, it is still not clear how the channel separation is related to the localization performance. Thus, in this study, we also compared the channel separation to the sound-localization performance in CTC systems, investigating its use as a predictor for the localization performance. The channel separation, being a frequency-dependent measure, is usually averaged over a frequency range in order to describe a CTC system by a single value. Keeping in mind that different frequency regions contribute differently to the sound localization in the horizontal and sagittal planes, we investigated how channel separation calculated in specific frequency regions describes the different aspects of the sound localization.

Current CTC systems usually suffer from various technical limitations. For example, the cancellation crucially depends on the listener's alignment within the loudspeaker setup (Takeuchi *et al.*, 2001) and artifacts due to an asymmetric listener alignment might be even more prominent than those resulting from mismatched CTC filters (Rose *et al.*, 2002). While in real-time reproduction systems, this issue can be tackled by tracking the listener position, for static systems, special loudspeakers or loudspeaker combinations are required to increase the area of the sweet spot (Takeuchi and Nelson, 2002). Also the loudspeakers, usually simulated as point sources, have a non-ideal transfer function and directionality, which has a strong effect on the quality of the CTC systems (Qiu *et al.*, 2009). Further potential factors affecting the cancellation are spontaneous head movements and room reflections.

Our study aimed to control all those issues in order to focus on the effect of the mismatch. Issues like listener's misalignment, loudspeaker transfer function and directivity, spontaneous movements, and room effects were controlled by using a binaural simulation for the tests of the localization performance. In the binaural simulation of the CTC system, the stimulus was presented via headphones and listener-individual HRTFs were used to simulate the propagation paths between loudspeakers and the listener. Thus, our system consisted of three different filter stages: (i) listener-individual DTFs, used to create an acoustic target; (ii) CTC filters, used to create the transaural signals for the virtual loudspeakers, and; (iii) listener-individual HRTFs, used to simulate the virtual loudspeakers. In such a setup the loudspeaker effects reduce to that of the HRTF measurement. Also, the listeners' head is virtually fixed at exactly the same position within the sweet spot.

II. METHODS

A. Subjects

Eight listeners participated in this study, all of them having absolute hearing thresholds within the 20-dB range of the average normal-hearing population in the frequency range between 0.125 and 12.5 kHz. All listeners showed front-back confusion rates below 20% in pre-experiments with their own broadband DTFs. The study was performed as a blind experiment, i.e., none of the listeners were the authors and the listeners were not enlightened as to the nature of the experiment.

B. CTC reproduction system

Figure 1 shows the setup of our CTC reproduction system, which is based on that from Bauck and Cooper (1996). Considering the system in the frequency domain, the incoming signals at the eardrums e_L and e_R are given by

$$e_L = H_{LL}(C_{LL}s_L + C_{RL}s_R) + H_{RL}(C_{LR}s_L + C_{RR}s_R),$$
$$e_R = H_{LR}(C_{LL}s_L + C_{RL}s_R) + H_{RR}(C_{LR}s_L + C_{RR}s_R), \tag{1}$$

which, written in a matrix form, corresponds to

$$\begin{bmatrix} e_L \\ e_R \end{bmatrix} = \begin{bmatrix} H_{LL} & H_{RL} \\ H_{LR} & H_{RR} \end{bmatrix} \cdot \begin{bmatrix} C_{LL} & C_{RL} \\ C_{LR} & C_{RR} \end{bmatrix} \cdot \begin{bmatrix} s_L \\ s_R \end{bmatrix} \tag{2}$$

or

$$\boldsymbol{e} = \boldsymbol{H} \cdot \boldsymbol{C} \cdot \boldsymbol{s}, \tag{3}$$

where elements of $\boldsymbol{e}$ are the signals at the listener's ears, elements of $\boldsymbol{H}$ describe the acoustic propagation paths from a loudspeaker to an ear given by the corresponding HRTFs, elements of $\boldsymbol{C}$ are the CTC filters, and $\boldsymbol{s}$ is the binaural signal to be presented.

In a theoretical CTC system, the CTC matrix is the inverse of the HRTFs matrix. Thus, Eq. (3) reduces to $\boldsymbol{e} = \boldsymbol{s}$, and the binaural signal $\boldsymbol{s}$ can be presented to the listener. Unfortunately, $\boldsymbol{H}$ is not always well-conditioned; its inversion might yield an ill-posed problem, in which case CTC filters may produce very high gains. The usual method to approximate the inverse of $\boldsymbol{H}$ is the Moore-Penrose pseudoinverse combined with the Tikhonov regularization (Kirkeby *et al.*, 1998)

$$\boldsymbol{C} = (\boldsymbol{H}^*\boldsymbol{H} + \beta \boldsymbol{I})^{-1}\boldsymbol{H}^*, \tag{4}$$

where $\boldsymbol{H}^*$ is the conjugate transpose of $\boldsymbol{H}$, β is the regularization parameter, and $\boldsymbol{I}$ is the identity matrix. Note that for $\beta = 0$, Eq. (4) is the ideal linear least-squares approximation of $\boldsymbol{H}^{-1}$, and for $\beta \neq 0$, Eq. (4) imposes a gain limitation to the resulting CTC filter.

C. HRTF measurements

HRTFs were measured individually for each listener. Twenty-two loudspeakers (custom-made boxes with 10 BGS, VIFA as drivers) were mounted on a vertical circular arc at fixed elevations from $-30°$ to $80°$, with a $10°$ spacing between $70°$ and $80°$, and $5°$ spacing elsewhere. The listener was seated in the center point of the circular arc on a computer-controlled rotating chair. The distance between the center point and each speaker was 1.2 m. Microphones (KE-4-211-2, Sennheiser) were inserted into the listener's ear canals and their output signals were directly recorded via amplifiers (FP-MP1, RDL) by the digital audio interface.

For the spatial setup and system identification, we used the same procedure as in Majdak *et al.* (2010). A 1729-ms exponential frequency sweep from 0.05 to 20 kHz was used to measure each HRTF. To speed up the measurements, for each azimuth, the multiple exponential sweep method was used (Majdak *et al.*, 2007). At an elevation of $0°$, the HRTFs were measured with a horizontal spacing of $2.5°$ within the range of $\pm 45°$ and with the horizontal spacing of $5°$ otherwise. With this rule, the measurement positions for other elevations were distributed with a constant spatial angle, i.e., the azimuthal spacing increased toward the poles. In total, HRTFs for 1550 positions within the full $360°$ horizontal span were measured for each listener. The measurement procedure lasted for $\sim$20 min. The acoustic influence of the equipment was removed by equalizing the HRTFs with the transfer functions of the equipment. The equipment transfer functions were derived from the reference measurements in which the microphones were placed at the center point of the circular arc and the measurements were performed for all loudspeakers.

The DTFs were calculated (Middlebrooks, 1999). The magnitude of the common transfer function (CTF) was calculated by averaging the log-amplitude spectra of all HRTFs for each individual listener. The phase spectrum of the CTF was set to the minimum phase corresponding to the amplitude spectrum. The DTFs were the result of filtering HRTFs with the inverse complex CTF. Finally, the impulse responses of all HRTFs and DTFs were windowed with an asymmetric Tukey window (fade in of 0.5 ms and fade out of 1 ms) to a 5.33-ms duration.

Two sets of HRTFs were measured for each listener.[1] The first measurements were performed for a previous study and the second measurements were performed for the present study. The interval between the two measurements was approximately 5 years.

D. Acoustic targets

Lateral and polar angles from the horizontal-polar coordinate system (see Fig. 2) were used to describe the acoustic target's position (Morimoto and Aokata, 1984). The tested lateral angle ranged from $-90°$ (right) to $90°$ (left). The polar angle of the targets ranged from $-30°$ (front, below eye-level) to $210°$ (rear, below eye-level). The targets were pseudo-uniformly distributed on the surface of the sphere by

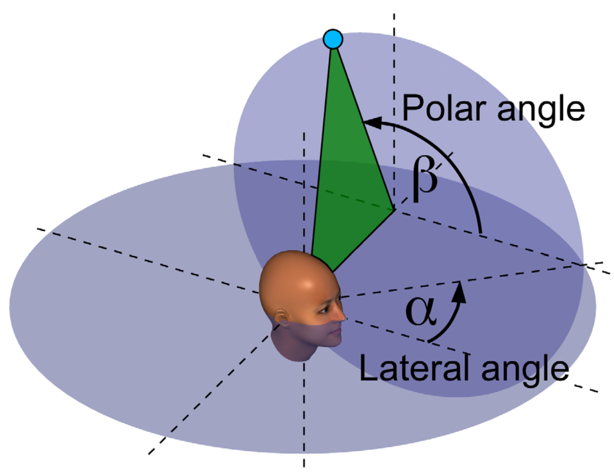

FIG. 2. (Color online) The coordinate system for the acoustic targets used in the localization experiments.

using a uniform distribution for the polar angle and an arcsine-scaled uniform distribution for the lateral angle.

The acoustic targets were Gaussian white noises with a duration of 500 ms and 10-ms fade-in and fade-out, filtered with the listener-specific DTFs. Prior to filtering, the position of the acoustic target was discretized to the grid of the available DTFs.

The level of the presented stimuli was 50 dB above the individual absolute hearing threshold in each condition. The threshold was estimated in a manual up-down procedure individually for each condition using an acoustic target positioned at azimuth and elevation of 0°. In the experiments, the stimulus level for each presentation was randomly roved within the range of ±5 dB to reduce the possibility of localizing spatial positions based on overall level.

E. Binaural CTC simulation

In the tested CTC conditions (see Sec. II G), the acoustic targets were processed with a binaural CTC simulation. The simulation was used to guarantee that subjects were always in the sweet-spot and to fully control the correspondence between the acoustic paths and CTC filters.

The CTC filters were calculated for a pair of virtual loudspeakers with one loudspeaker placed at 45° left and a second loudspeaker placed at 45° right to the listener. Thus, the loudspeaker span angle was 90°. The CTC filters were calculated for frequencies up to 16 kHz in order to provide all relevant spectral cues for accurate sound localization. The propagation paths from the loudspeakers to the listener's ears are described by so-called "setup HRTFs." The corresponding impulse responses were zero padded to 85.33 ms.[2] The CTC filters were calculated in the frequency domain according to Eq. (4) with $\beta = 0.005$ for each frequency.[3] Note that the effect of the regularization parameter depends on the overall level of the HRTFs. In order to use the same regularization parameter for all HRTF sets, the level of each HRTF was divided by the root-mean-square (RMS) level calculated for all directions in a HRTF set, yielding the same average level across all HRTF sets. Then, the CTC filters were calculated according to Eq. (4), converted back to the time domain, and circularly shifted by 3.125 ms to avoid non-causality.

Finally, the impulse responses where windowed with a one-sided Tukey window with a fade out of 18.6 ms at their end. These CTC filters were used in the experiments, i.e., the transaural signals were calculated by processing the acoustic target with the CTC network according to Eq. (1). Then, the transmission of the transaural signals from the loudspeakers to the listener's ears was simulated by filtering the transaural signals with the listener-individual HRTFs, so-called "playback HRTFs." Note that listener-individual HRTFs were used for the playback HRTFs in all conditions—only the setup HRTFs were varied in this study. All the signal processing was done in MATLAB (Mathworks) with the ITA-Toolbox[4] at the sampling rate of 48 kHz.

F. Apparatus and procedure

The virtual acoustic stimuli were presented via headphones (HD 580, Sennheiser) in a double-wall sound-proof room. The headphones were diffuse-field-compensated circumaural headphones and no additional headphone correction was applied. The listener stood on a platform enclosed by a circular railing. Stimuli were generated using a computer and output via a digital audio interface (ADI-8, RME) with a 48-kHz sampling rate. A virtual visual environment was presented via a head-mounted display (3-Scope, Trivisio). It provided two screens with a field of view of 32° × 24° (horizontal × vertical dimensions). The virtual visual environment was presented binocularly with the same picture for both eyes. A tracking sensor (Flock of Birds, Ascension) was mounted on the top of the listeners' head, which captured the position and orientation of the head in real time. A second tracking sensor was mounted on a manual pointer. The tracking data were used for the three-dimensional graphic rendering and response acquisition.

The listeners were immersed in a spherical virtual visual environment (Majdak *et al.*, 2010). They held a pointer in their right hand. The projection of the pointer direction on the sphere's surface, calculated based on the position and orientation of the tracker sensors, was visualized and recorded as the perceived target position. The pointer was visualized whenever it was in the listeners' field of view.

Prior to the tests, listeners performed a visual and an acoustic training. The goal of the visual training was to train subjects to perform accurately within the virtual environment. The visual training was a simplified game in the first-person perspective where listeners had to find a visual target, point at it, and click a button within a limited time period. This training was continued until 95% of the targets were found with a RMS angular error in the range of 2°. This performance was reached within a few hundred trials. Then the acoustic training was performed with listener-individual DTF (Majdak *et al.*, 2010). The goal of the acoustic training was to settle a stable localization performance of the subjects. The acoustic training consisted of 6 blocks, 50 acoustic targets each, lasting for ∼2 h.

In the actual acoustic tests, at the beginning of each trial, the listeners were asked to align themselves with the reference position and click a button. Then, the stimulus was presented. During the presentation, the listeners were instructed not to move. The listeners were asked to point to the perceived stimulus location and click the button again. This response was recorded for the data analysis. The tests were performed in blocks; each block consisted of 100 acoustic targets and lasted for ∼15 min. Within a block, the targets were sampled randomly with replacement from the 1550 possible spatial positions. After each block, subjects had a break of ∼15 min. The procedure was controlled by LocaCTC from the ExpSuite.[5]

G. Conditions

Eight conditions were tested in three blocks each. The order of the blocks was randomized in such a way that within eight blocks all conditions were in a randomized order.

The first two conditions consisted of pure acoustic targets, i.e., binaural signals without the CTC simulation. The former, *binOwn*, used the same DTFs as those used for the acoustic training while the latter, *binOwnB*, used DTFs from the latter HRTF measurement.

In the individual matched CTC condition, *ctcOwn*, the acoustic targets were presented via the simulated CTC system using the same setup and playback HRTFs, namely the listener-individual HRTFs from the condition binOwn.[6] The condition ctcOwn corresponds to the matched case from Akeroyd *et al.* (2007) and represents an *ideal individualized* CTC system, where the CTC filters match exactly the acoustic paths between the loudspeakers and the listener.

In the individual but mismatched CTC condition, *ctcOwnB,* the playback HRTFs were the same as in the matched CTC condition, but the setup HRTFs were those from the latter measurement and corresponded to those used for the condition binOwnB. The condition ctcOwnB represents a *realistic individualized* CTC system, where for the calculation of the CTC filters the listener-individual HRTFs have been measured, but during the signal presentation the acoustic propagation paths do not exactly match these measured HRTFs. In particular, distortions of the spectral cues were expected at high frequencies where (because of the small wavelength) small changes in the measurement setup have a potentially large impact on the measured HRTFs. Thus, from an acoustic point of view, this condition is clearly a mismatched condition.

The last CTC conditions were *non-individual* mismatched conditions, i.e., the setup HRTFs were those from other sources, while the playback HRTFs did not change. In the condition *ctcKemar,* the setup HRTFs were those from measurements on a mannequin (Gardner and Martin, 1995). Note that in contrast to all other HRTFs in our study, these HRTFs were measured using microphones included in an ear simulator, yielding a HRTF set containing the direction independent ear-canal transfer function. In the remaining non-individual conditions, the setup HRTFs were those from other listeners, namely, NH57, NH64, and NH68. These particular listeners were also tested with setup HRTFs from NH12 in order to obtain the same number of tested conditions for each listener. We refer to those conditions as *ctcNH57, ctcNH64, ctcNH68,* and *ctcNH12*. For the sake of simplicity, all non-individual conditions are referred to as *ctcOther*.

H. Channel separation

In the ideal CTC system, one would be able to present the binaural signal $\boldsymbol{s} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ without any artifacts, namely, $\boldsymbol{e} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ for all frequencies. For realistic CTC systems, this assumption is not achieved, i.e., $\boldsymbol{e} \neq \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. CS has been proposed to describe the quality of such systems (Gardner, 1997) and has been defined as the logarithmic difference between the signals at the two ears (Bai and Lee, 2006). With $\boldsymbol{s} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, Eq. (1) reduces to

$$e_{\mathrm{L}} = H_{\mathrm{LL}}C_{\mathrm{LL}} + H_{\mathrm{RL}}C_{\mathrm{LR}},$$
$$e_{\mathrm{R}} = H_{\mathrm{LR}}C_{\mathrm{LL}} + H_{\mathrm{RR}}C_{\mathrm{LR}}. \tag{5}$$

Thus, for the left ear, the CS is

$$\mathrm{CS_L} = 20\log_{10}\left(\frac{|H_{\mathrm{LL}}C_{\mathrm{LL}} + H_{\mathrm{RL}}C_{\mathrm{LR}}|}{|H_{\mathrm{LR}}C_{\mathrm{LL}} + H_{\mathrm{RR}}C_{\mathrm{LR}}|}\right). \tag{6}$$

The same applies for the right ear. Note that with our definition, a larger CS suggests a better CTC system, which follows Akeroyd *et al.* (2007) and Qiu *et al.* (2009), but is contrary to Parodi and Rubak (2010), and Bai and Lee (2006). Note also that the CS is given for each frequency separately and, in order to obtain a single-valued metric for a CTC system, the CS is usually averaged over a frequency range (Bai and Lee, 2006; Akeroyd *et al.*, 2007).

Without the CTC filters, i.e., $\boldsymbol{C} = \boldsymbol{I}$, the CS for the left ear would be

$$\hat{CS}_L = 20\log_{10}\left(\frac{|H_{\mathrm{LL}}|}{|H_{\mathrm{LR}}|}\right). \tag{7}$$

The same definition applies for the right ear. $\hat{CS}$ represents the natural CS caused by the head shadow, and it is equivalent to that obtained with a simple stereophonic reproduction system. Hence, not only CS, but also $\hat{CS}$ directly depends on the particular loudspeaker position, it is frequency dependent, and it has its maximum (approximately 30 dB) at higher frequencies (Blauert, 1997). In a well-designed realistic CTC, based on Eq. (4) with $\beta \neq 0$, the CS should be substantially larger than $\hat{CS}$.

III. RESULTS AND DISCUSSION

A. Spectral features

In the top panels, Fig. 3 shows the left-ear DTFs of an exemplary listener (NH64) for the conditions binOwn and binOwnB as amplitude spectra for the median plane. The spectral peaks and notches, i.e., the spectral features assumed to be relevant for sagittal-plane localization (Wightman and Kistler, 1997b; Macpherson and Middlebrooks, 2002) are well-represented in both conditions. The differences between the two conditions are, however, clearly evident. For example, the notch beginning at 9 kHz and $-30°$ and extending to 10 kHz and $+30°$ (type 1 in Takemoto *et al.*, 2012) seems to be more pronounced in binOwn than in binOwnB. On the other hand, the notch extending from 11 to 13 kHz at the upper rear directions (type 3 in Takemoto *et al.*, 2012) seems to be more pronounced in binOwnB than in binOwn. The differences between the both DTF sets can be attributed to small differences in details of the HRTF measurement like the insertion depth of the microphones or the head position. Both DTF sets, however, seem to provide similar spectral localization cues, as indicated by the similar localization performance (as discussed in Sec. III B).

In the bottom panels, Fig. 3 shows the "CTC DTFs" for the conditions ctcOwn and ctcOwnB. The CTC DTFs describe the total filtering performed by our setup in order to

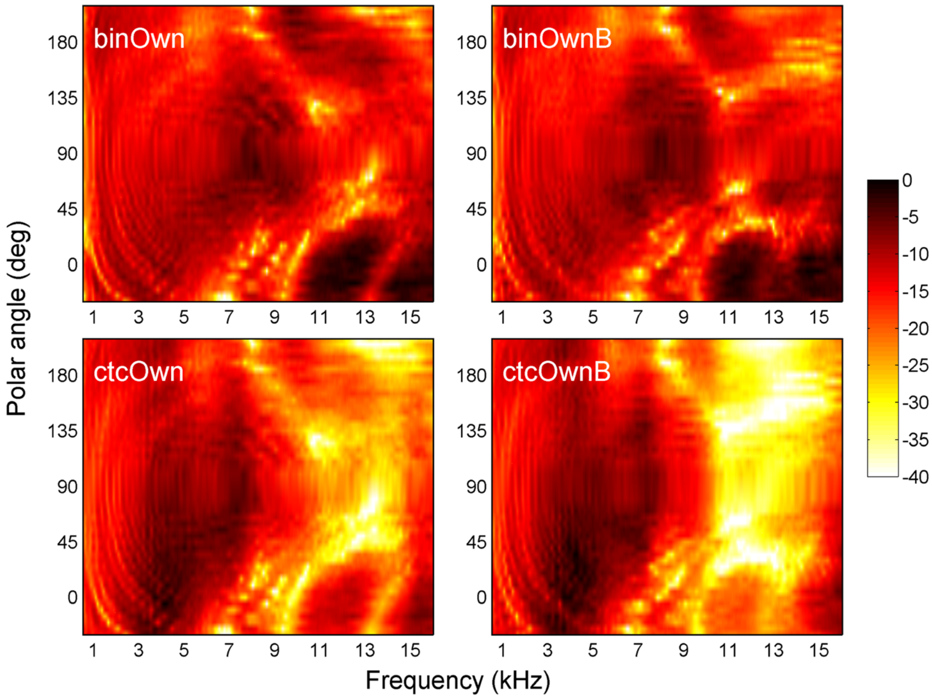

FIG. 3. (Color online) Top panels: Left-ear DTFs of an exemplary listener (NH64) shown as amplitude spectra for the median plane. Polar angles of 0° and 180° encode the position at the eye level in front and back, respectively. The color encodes the relative magnitude in dB. The panels binOwn (top left panel) and binOwnB (top right panel) correspond to the binaural conditions. Bottom panels: Left-ear CTC DTFs (see Sec. III A) with other details as in the top panels. The conditions ctcOwn (bottom left panel) and ctcOwnB (bottom right panel) correspond to the individualized matched and mismatched CTC conditions, respectively.

position a virtual sound source at a given angle in a CTC condition. Each CTC DTF contains the filtering by the listener-individual DTF for a given angle, the CTC filtering according to Eq. (3), and the simulation of the propagation paths.

The CTC DTFs for the condition ctcOwn are shown in the bottom left panel of Fig. 3. Recall that ctcOwn uses binOwn for both the CTC filtering and simulation of the propagation paths and represents the matched CTC system. Thus, ideally, the bottom left panel should be identical to the top left panel. Indeed, the spectral features appear to be similar between ctcOwn and binOwn. The small differences can be attributed to the regularization in the calculation of the CTC filters.

The condition ctcOwnB is shown in the bottom right panel of Fig. 3. Recall that ctcOwnB represents a mismatched CTC system and thus, artifacts are expected. Indeed, for the rear directions, drastic artifacts can be observed, especially in the frequency range above 9 kHz. Even though fewer artifacts can be observed for the front directions, the spectral features appear to be generally smeared across frequencies and directions. Such artifacts are reasonable considering the differences between the two DTF sets.

B. Localization performance

1. General

Figures 4 and 5 show results of the localization experiment for an exemplary listener (NH64). The target and response angles are shown on the horizontal and vertical axes, respectively, of each panel. For the polar dimension, the results are shown for targets with lateral angles within ±30° only. Responses that resulted in absolute polar errors larger than 90° are plotted as filled circles. All other responses are plotted as open squares. The performance seems to be similar for both binaural conditions and the differences to the ctcOwn condition seem to be negligible. A generally degraded performance can be observed for ctcOwnB and also for all other mismatched conditions.

Localization errors were calculated by subtracting the target angles from the response angles. The lateral error (LE) was the RMS of the localization error in the lateral dimension. In the polar dimension, we separated our data analysis in confusions between the hemifields and the local performance within the correct hemifield. Only responses within the lateral range of ±30° were considered (Middlebrooks, 1999). The rate of confusions was represented by the quadrant error (QE), which is the percentage of responses where

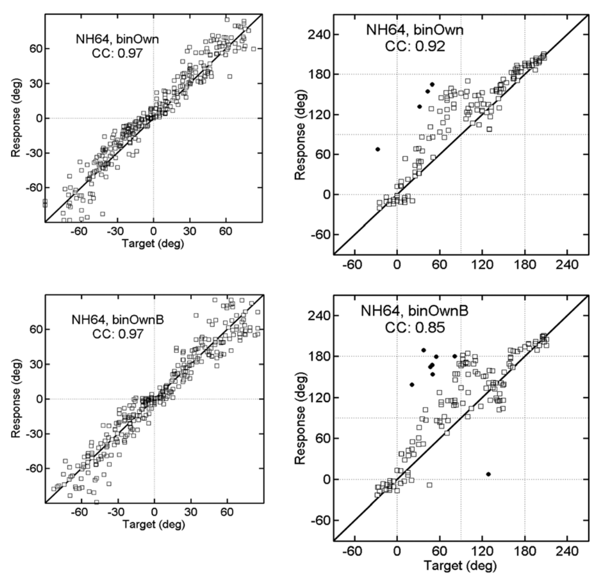

FIG. 4. Localization results of an exemplary listener (NH64) for the binaural conditions. Lateral results are plotted in the smaller panels, the polar results in the larger panels. Polar results outside the lateral range of ±30° are not shown. Filled circles: Responses with errors outside the ±90° range. CC: Correlation coefficient between responses and targets.

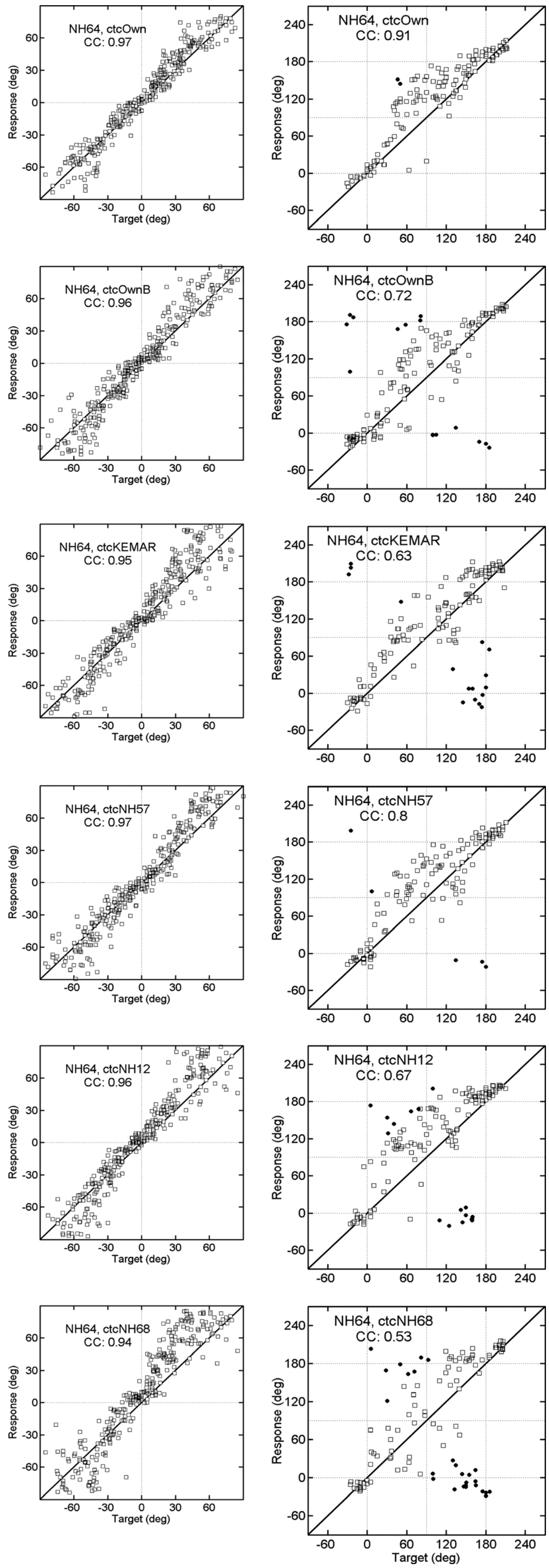

FIG. 5. Localization results of an exemplary listener (NH64) for the CTC conditions. Other details as per Fig. 4.

the absolute polar error exceeded 90° (Middlebrooks, 1999). In order to quantify the local performance in the polar dimension, the local polar RMS error (PE) was calculated, i.e., the RMS of the polar errors calculated for the responses with the absolute polar error smaller than 90°.

The results described by the error metrics QE, PE, and LE are shown in Tables I, II, and III, respectively. In both binaural conditions, the group average performance was within the range of previously reported performance for localization of virtual broadband noises under comparable conditions. For the sagittal planes, the average QE of 8.6% (for binOwn, 7.1% for binOwnB) was similar to QE of 7.7% from Middlebrooks (1999) and QE of 9.4% from Goupell *et al.* (2010). Also, the average PE of 31.0° (31.6° for binOwnB) was similar to PE of 28.7° from Middlebrooks (1999) and PE of 33.8° from Goupell *et al.* (2010). For the horizontal planes, the average LE of 10.7° (10.4° for binOwnB) was similar to LE of 14.5° from Middlebrooks (1999) and LE of 12.4° from Majdak *et al.* (2010) (reported in Majdak *et al.*, 2011).

Repeated-measures (RM) analyses of variance (ANOVA) were used for the statistical analysis of the results. Each of the three tested blocks was treated as within-subject repetition. For the binaural conditions, RM ANOVAs were calculated on the LE, PE, and QE with the factor condition at two levels (binOwn, binOwnB). The analyses showed neither significant effect for the QE ($p = 0.44$), nor for PE ($p = 0.68$), nor for LE ($p = 0.38$). This indicates that despite five years of interruption between the two HRTFs measurements, both HRTF sets provided localization performance at a similar level.

2. Individualized CTC systems

In order to investigate the performance in individualized CTC systems, RM ANOVAs with the factor condition at four levels (binOwn, binOwnB, ctcOwn, and ctcOwnB) was performed. The results were significant for the QE ($p = 0.005$) and the LE ($p < 0.001$), but not the PE ($p = 0.20$). Tukey-Kramer *post hoc* tests were used to test the statistical significance of particular levels. The significance was considered at $p < 0.05$. *Post hoc* tests showed that only the ctcOwnB condition yielded significantly larger QE and LE compared to all other conditions. Note that even though for PE the differences were not significant, the PE was larger for ctcOwnB (34.2°) than for ctcOwn (31.8°) or the binaural conditions (31.0° and 31.6°).

The lack of significance in the differences between the ideal CTC and binaural systems indicates that our ideal CTC system provided localization performance at the level of the binaural reproduction systems. In a realistic application of the individualized CTC, this situation is, however, unachievable because the propagation paths would (slightly) change as soon as a listener leaves the HRTF measurement setup and enters the CTC system. In our study, such a situation was represented by the condition ctcOwnB where individual HRTFs from the latter measurement were used to calculate CTC filters. The performance in such a realistic CTC system was worse than that in an ideal CTC system in terms of

TABLE I. Quadrant errors, QEs, in % for all listeners and conditions tested. The condition ctcOther represents the median of the non-individual conditions. Averages and standard deviations are across the listeners.

Condition	NH12	NH14	NH15	NH57	NH62	NH64	NH68	NH72	Average ± standard deviation
binOwn	3.3	5.3	7.4	25.4	5.9	2.7	6.3	12.5	8.6 ± 7.4
binOwnB	1.7	5.2	1.5	17.5	2.8	6.1	13.3	9.1	7.1 ± 5.8
ctcOwn	4.2	7.4	6.8	15.2	6.7	1.4	13.5	8.9	8.0 ± 4.5
ctcOwnB	0.6	6.3	5.0	33.0	4.0	11.0	32.3	16.7	13.6 ± 12.7
ctcOther	6.3	9.7	16.3	25.8	10.0	13.2	33.2	20.7	16.9 ± 9.1
ctcKemar	2.0	7.8	10.1	23.1	5.4	12.2	36.6	5.1	12.8 ± 11.6
ctcNH57	10.6	11.5	14.2	—	11.8	4.4	36.6	23.9	16.1 ± 10.7
ctcNH64	0.0	4.4	18.5	28.3	10.8	—	29.8	23.2	16.4 ± 11.6
ctcNH68	21.9	12.1	27.1	23.3	9.1	23.1	—	18.3	19.3 ± 6.5
ctcNH12	—	—	—	28.3	—	14.2	25.3	—	22.6 ± 7.4

significantly larger QEs and LEs. Note that in ctcOwnB, not only the average performance degraded, but also the variance across our listeners increased. Since the degree of mismatch depends on the differences in the two DTF sets, namely binOwn and binOwnB, the increased variance in the performance arises from the variance in the two DTF sets. There are many sources for such differences in the measurements: insertion depth of the microphones, positioning of the head, or even aging of our listeners. While we are not able to rule out any of these factors, our results demonstrate that a mismatch between the playback and setup HRTFs results in a degraded localization performance in a CTC system, even when both HRTFs provide a similar performance in a binaural system.

Compared to ctcOwnB, the ctcOwn condition yielded a better performance in the horizontal plane. This result confirms the results for modeling interaural differences in matched and mismatched CTC systems (Akeroyd *et al.*, 2007), where for mismatched CTC systems, the model predicted large ITD and ILD errors. Looking more closely at their results, it seems like the errors are large for large interaural differences only (compare Fig. 10 of Akeroyd *et al.*, 2007). For smaller interaural differences, the errors seem to be negligible, which would suggest a correct reproduction of central targets. In order to investigate this issue, the targets were grouped to those within (central) and those outside (lateral) the loudspeaker span, and the LEs were calculated as a function of the target lateral angle (Fig. 6). While in the ctcOwnB condition the performance seems to slightly degraded for the central targets, the performance appears to be much worse for the lateral targets. An RM ANOVA was performed on the LEs for the factors target direction (central, lateral) and condition (ctcOwn, ctcOwnB). Both main effects ($p < 0.001$) and their interaction ($p = 0.048$) were significant. The significant interaction suggests a different impact of the condition for the two target directions. The *post hoc* test showed that the only significant difference was that for the lateral targets tested with ctcOwnB (17.2°) when compared to the lateral targets tested with ctcOwn (12.5°) or the central targets (11.5° for ctcOwnB and 9.7° for ctcOwn).

Thus, for targets placed outside the loudspeaker span, the mismatch significantly affects the lateral localization performance, and for targets placed inside the span, the mismatched CTC system may yield a similar performance as the matched CTC system. This seems to confirm our observations in the details of binaural modeling in Akeroyd *et al.* (2007). It further suggests a correspondence between central targets in a mismatched CTC system and phantom sources in a stereophonic reproduction system.

Targets placed at elevations near the loudspeakers may also correspond to a phantom source stereophonic reproduction. If such targets were well-localized in sagittal planes even in a mismatched CTC system, then the difference between ctcOwn and ctcOwnB would depend on the target polar angle, with a larger difference in the performance for targets placed behind the listener. Figure 7 shows the QEs as a function of the polar angle, with targets grouped to four groups with a polar angle span of 60°. The QE seems to

TABLE II. Local polar errors, PEs, in degrees for all listeners and conditions tested. Other details as per Table I.

Condition	NH12	NH14	NH15	NH57	NH62	NH64	NH68	NH72	Average ± standard deviation
binOwn	28.3	26.5	30.5	37.0	27.0	28.5	31.1	38.9	31.0 ± 4.6
binOwnB	25.0	30.1	31.2	35.6	26.6	34.8	34.2	35.5	31.6 ± 4.1
ctcOwn	26.7	25.8	36.2	33.9	35.0	32.0	26.7	38.4	31.8 ± 4.8
ctcOwnB	26.8	35.6	32.2	36.7	29.0	32.0	41.2	39.8	34.2 ± 5.1
ctcOther	35.4	31.9	40.1	39.4	34.1	33.7	37.4	41.5	36.7 ± 3.5
ctcKemar	35.2	26.0	39.9	33.7	35.2	31.3	40.6	36.8	34.8 ± 4.7
ctcNH57	35.5	31.4	40.4	—	34.4	33.5	42.9	42.7	37.2 ± 4.7
ctcNH64	26.1	32.3	36.0	42.2	31.3	—	32.3	43.6	34.8 ± 6.2
ctcNH68	36.8	32.8	41.1	38.2	33.7	33.9	—	40.4	36.7 ± 3.4
ctcNH12	—	—	—	40.6	—	38.0	34.2	—	37.6 ± 3.2

TABLE III. Lateral errors, LEs, in degrees for all listeners and conditions tested. Other details as per Table I.

Condition	NH12	NH14	NH15	NH57	NH62	NH64	NH68	NH72	Average ± standard deviation
binOwn	8.1	10.1	10.9	16.7	9.7	9.4	10.0	10.7	10.7 ± 2.5
binOwnB	8.2	9.6	12.7	13.6	9.9	9.2	9.4	10.6	10.4 ± 1.8
ctcOwn	8.0	9.1	11.7	14.0	9.6	8.7	10.4	11.1	10.3 ± 1.9
ctcOwnB	10.8	13.4	14.3	15.5	9.5	10.2	11.8	18.2	13.0 ± 2.9
ctcOther	11.0	11.3	14.4	15.5	10.9	10.8	13.0	14.0	12.6 ± 1.9
ctcKemar	9.8	9.8	14.1	14.8	9.4	11.3	13.7	9.3	11.5 ± 2.3
ctcNH57	11.9	8.9	14.8	—	12.8	10.2	13.6	14.4	12.4 ± 2.2
ctcNH64	10.0	12.8	13.7	16.2	10.4	—	12.3	13.5	12.7 ± 2.1
ctcNH68	13.0	16.2	17.2	17.6	11.4	15.8	—	14.6	15.1 ± 2.3
ctcNH12	—	—	—	14.7	—	10.0	11.4	—	12.0 ± 2.4

increase with the polar distance between the targets and the loudspeaker elevation. An RM ANOVA was performed with factors target hemifield (frontal targets: 0°, rear targets: 180°) and condition (ctcOwn, ctcOwnB). The main factors ($p < 0.018$ for both) and their interaction ($p = 0.023$) were significant. The significant interaction suggests a different impact of the conditions in the two target hemifields. The *post hoc* test showed that while for the frontal targets, the difference between the conditions was not significant (5.5% for ctcOwn, 5.9% for ctcOwnB), for the rear targets, highly significantly ($p < 0.005$) more QEs occurred in the ctcOwnB (16.8%) than in the ctcOwn (5.7%) condition. This indicates a strong impact of the mismatch on the localization performance for the targets placed behind the listener.

In summary, for the targets placed in the same hemisphere as the loudspeakers, the individual but mismatched condition ctcOwnB yielded a similar sagittal-plane performance as that for the ideal matched condition ctcOwn. For the targets placed in the opposite hemisphere as the loudspeakers, the QE were substantially larger. This indicates that mismatched but individualized CTC systems are able to provide a good performance *only* for the frontal targets. The exact match of the CTC filters to the propagation paths seems to be highly relevant for targets virtually placed at the other hemisphere than the loudspeakers. Our results further indicate that only an ideal, individualized, matched CTC system provides a correct reproduction of the spectral cues required for accurate sagittal-plane sound localization in both hemifields.

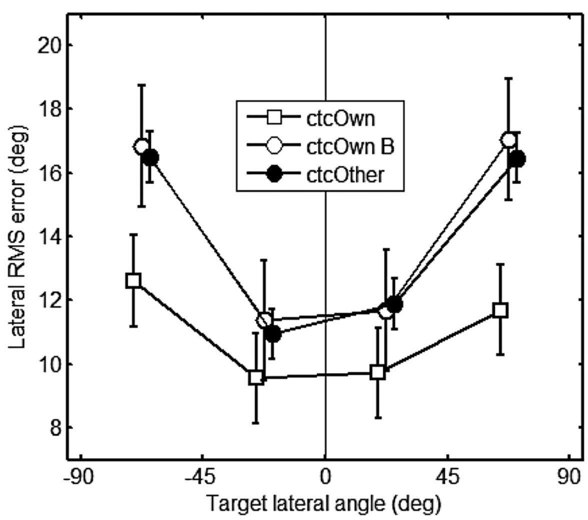

FIG. 6. LE as a function of the lateral target angle. The vertical bars represent ±1 standard errors.

3. Non-individualized CTC systems

The localization performance in the non-individualized and thus mismatched CTC systems is usually assumed to be worse than that in individualized CTC systems (Akeroyd *et al.*, 2007). However, an individualized CTC system can also be matched or mismatched, depending on whether an ideal or a realistic CTC system is under consideration. Thus, in the following we compare the non-individualized CTC systems (ctcOther) with both the ideal, thus matched condition (ctcOwn) and the realistic, but mismatched condition (ctcOwnB).[7]

The LE increased from 10.3° (ctcOwn) to 13.0° (ctcOwnB) but then decreased to 12.6° (ctcOther), indicating a weak impact of the individualization on the horizontal-plane localization performance (see Table III). However, there might have been some differences at particular target directions. Thus, the targets were grouped to those within (central) and those outside (lateral) the loudspeaker span and the LE were calculated as a function of the target lateral angle (Fig. 6). LE was still similar for all the mismatched (individual and non-individual) conditions. Thus, for the horizontal-plane localization, there seems to be no difference between the two types of mismatched CTC systems and an individualized CTC system seems to be of no advantage.

The QE increased from 8.0% (ctcOwn) to 13.6% (ctcOwnB) and then further to 16.9% (ctcOther, see Table I). Also, the PE increased from 31.8° (ctcOwn) to 34.2° (ctcOwnB) and then further to 36.7° (ctcOther; see Table II). The RM ANOVAs were performed with the factor condition at three levels (ctcOwn, ctcOwnB, and ctcOther) on the QEs and PEs. The factor condition significantly affected the QEs ($p < 0.001$) and the PEs ($p = 0.019$). The *post hoc* tests showed that while ctcOther yielded significantly larger errors compared to ctcOwn, the errors were not significantly different when compared to ctcOwnB. On the first sight, this might indicate that for the sagittal-plane localization the

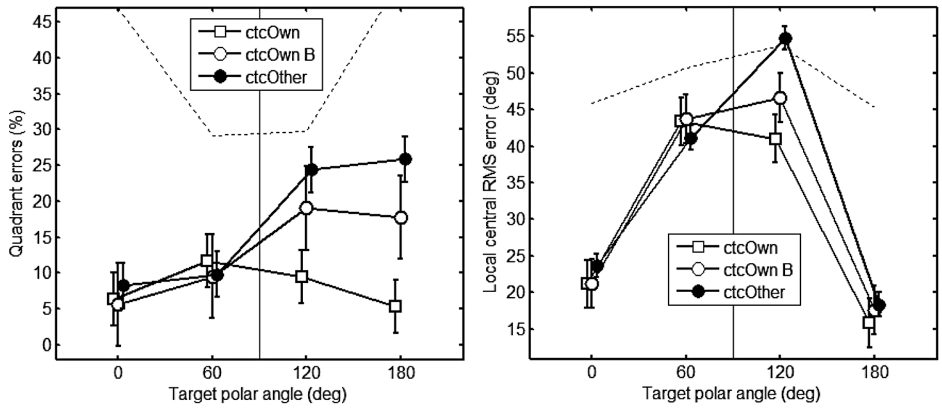

FIG. 7. QE and PE as functions of the polar target angle. The dashed lines show the errors which would result from random responses. The vertical bars represent ±1 standard errors.

performance is independent of the individualization of the filters. However, differences at particular positions might have been expected. Thus, the targets were grouped to four groups with a polar angle span of 60° and the QE and PE were calculated as a function of the polar angle (Fig. 7). For ctcOther, the QE increased with the increasing distance between the targets and the loudspeakers more than it did for ctcOwnB. An RM ANOVA with factors target hemifield (frontal targets: −30° to 30°, rear targets: 150° to 210°) and condition (ctcOwnB and ctcOther) was performed on the QEs. The main factors condition ($p = 0.048$) and target hemifield ($p < 0.001$) were significant but their interaction was not ($p = 0.23$). For the front targets, the QE was 5.8% and 8.1% for ctcOwnB and ctcOther, respectively. For the rear targets, the QE was 16.8% and 26.1% for ctcOwnB and ctcOther, respectively. While the non-significant interaction shows that none of the hemifield-condition combinations was significantly different from the others, the significance in the factor condition raises evidence that ctcOwnB indeed yielded a significantly better performance than ctcOther—when separately analyzed for the two hemifields. The 50% larger QE in ctcOther for the rear targets further supports this evidence. Similar situation revealed the RM ANOVA performed on the PEs with factors target hemifield (frontal targets: −30° to 90°, rear targets: 90° to 210°), condition (ctcOwnB and ctcOther), and their interaction. The interaction ($p = 0.012$) was significant and the *post hoc* test showed that while for the frontal targets, the difference between the conditions was not significant (32.4° for ctcOther and 33.6° for ctcOwnB), for the rear targets, significantly ($p < 0.01$) larger PE occurred in the ctcOther (40.8°) than in the ctcOwnB (33.7°) condition. Thus, the individualization of the CTC systems was able to substantially reduce the PE for rear targets.

The worse localization performance for the rear targets is consistent with the findings of Takeuchi and Nelson (2007), who used mismatched non-individualized CTC filters to compare between a CTC method comparable to ours and the "optimal source distribution" method (Takeuchi and Nelson, 2002). For the comparable condition, they report more back-to-front than front-to-back confusions, consistent with our findings of more quadrant errors for the rear targets.

In summary, our analysis demonstrates that in mismatched CTC systems, the sagittal-plane localization performance improves when individualized CTC filters are considered, especially for targets placed behind the listener.

C. Channel separation

CS was calculated according to Eq. (6) for all available frequencies between 0.3 and 8 kHz. Similar frequency range was used by Akeroyd *et al*. (2007) and Parodi and Rubak (2011b), and by using this frequency range we are able to

TABLE IV. CS in dB averaged over three frequency ranges. The last two rows show the natural CS, $\hat{CS}$, averaged over both ears. Conditions not tested in the localization experiments are shown italic.

Frequency range	0.3 to 8 kHz									0.3 to 2 kHz	4 to 16 kHz
CS	NH12	NH14	NH15	NH57	NH62	NH64	NH68	NH72	Average ± standard deviation	Average ± standard deviation	Average ± standard deviation
ctcOwn	70.9	67.6	67.9	67.9	68.3	67.0	68.8	68.6	68.4 ± 1.2	50.4 ± 2.2	58.5 ± 2.6
ctcOwnB	16.2	13.5	11.5	15.6	19.2	18.2	11.4	12.3	14.7 ± 3.0	16.5 ± 3.4	14.2 ± 1.2
ctcKemar	17.2	15.6	16.8	15.9	18.1	18.0	12.6	18.9	16.6 ± 2.0	16.5 ± 1.9	12.6 ± 0.6
ctcNH57	13.6	17.4	17.9	—	13.9	16.3	11.9	13.7	14.9 ± 2.3	15.2 ± 2.8	13.2 ± 0.6
ctcNH64	17.2	15.1	16.1	16.2	17.7	—	11.0	17.1	15.8 ± 2.3	16.6 ± 2.3	13.6 ± 1.5
ctcNH68	14.5	11.5	12.0	11.9	12.2	11.0	—	12.9	12.3 ± 1.2	12.4 ± 2.1	11.9 ± 0.7
ctcNH12	—	*13.6*	*13.9*	13.6	*16.3*	17.2	14.5	*18.9*	15.4 ± 2.1	14.1 ± 1.2	14.5 ± 1.1
$\hat{CS}$											
binOwn	16.4	15.3	15.1	14.4	14.3	14.5	13.8	14.6	14.8 ± 0.8	8.1 ± 0.3	17.1 ± 1.1
binOwnB	15.3	14.6	14.1	15.3	14.6	13.9	13.8	14.8	14.5 ± 0.6	8.2 ± 0.3	17.4 ± 1.4

provide a direct comparison to those studies. The frequency-dependent CS was then spectrally averaged over the frequency range resulting in a single-valued CS for each condition, listener, and frequency range. Note that, for the sake of clarity, we skip the word "average" in this section, despite the fact that the CS is actually a spectral average. The CS is shown in Table IV for each listener and condition. The last two rows of Table IV show the corresponding $\hat{CS}$ calculated according to Eq. (7) for the left and by analogy for the right ear, and averaged over the ears and the same frequency range as done for the CS.

For ctcOwn, the CS was large, on average 68.4 dB, and in the range of those reported previously for matched CTC systems (Bai and Lee, 2006; Akeroyd *et al.*, 2007). For ctcOwnB, the CS was substantially lower, on average 14.7 dB, and in the range of $\hat{CS}$ for both measured HRTF sets (14.9 and 14.5 dB). This indicates that individualized but mismatched CTC systems and reproduction systems without any CTC yield similar CS.

For ctcOther, the CS was on average 15.0 dB and in the range of those reported previously for non-individualized CTC systems (Akeroyd *et al.*, 2007, average of 17.1 dB). It was also similar to that for ctcOwnB (14.7 dB), which might lead to the conclusion that an individualization of the CTC systems is not necessary at all. Note that such a conclusion would be inconsistent with our results from the localization experiments.

One reason for the discrepancy between the CS and the localization performance might be the frequency range used for averaging the frequency-dependent CS. The choice of our frequency range was based on previous studies (Akeroyd *et al.*, 2007; Parodi and Rubak, 2011b). While this choice might appear arbitrary, the frequency range between 0.3 and 8 kHz provides indeed some advantages as explained in the following. The CS is calculated between the two ears and is, in principle, an interaural metric. Therefore, it might indeed have a potential to describe the horizontal-plane localization performance, which also depends on interaural cues. According to the duplex theory (Strutt, 1907; Macpherson and Middlebrooks, 2002), the horizontal-plane localization depends on the ITDs in the lower (below 2 kHz) frequencies and on the ILDs in the higher (above 2 kHz) frequencies. Frequencies above 8 kHz do not seem to further contribute to the horizontal-plane localization. Thus, the mid-frequency CS, calculated for the frequency range from 0.3 to 8 kHz, seems to capture all the ITDs and ILDs, both relevant features for the horizontal-plane localization.

The CS calculated for other frequency ranges, however, might better describe other aspects of sound localization. The low-frequency CS (0.3 to 2 kHz) might better reflect the sound localization based on the ITDs, which actually are assumed to be the most salient cues for sound localization in the horizontal planes (Wightman and Kistler, 1992). The high-frequency CS (4 to 16 kHz) might better reflect the sound localization in the sagittal planes where the spectral cues are assumed to most contribute in the frequency range from 4 to 16 kHz (Carlile and Pralong, 1994; Perrett and Noble, 1997; Middlebrooks, 1999). Note that even though the CS is an interaural metric, it is commonly used to describe the quality of CTC systems and its relation to the sagittal-plane localization performance is of interest. Both low- and high-frequency CSs averaged over the listeners are shown in the right-most column of Table IV.

As averages over listeners, low- and high-frequency CSs showed a similar trend to the mid-frequency CS (0.3 to 8 kHz) when compared across the conditions. A further comparison of the CS and $\hat{CS}$ revealed that for the low-frequency range, the average CS was 15.0 dB and thus, larger than the corresponding $\hat{CS}$ of 8.15 dB. This indicates that the CTC indeed increased the CS in the frequencies below 2 kHz. For the high-frequency range, the CS was 13.2 dB and thus *smaller* than the corresponding $\hat{CS}$ of 17.25 dB. This means that the CTC actually *decreased* the CS in the frequencies above 4 kHz for the tested CTC setup. Thus, without any CTC in the frequency range above 4 kHz, our mismatched CTC systems would show a larger CS.

This finding may provide an explanation for the results from Bai and Lee (2007), who band-limited their CTC to 6 kHz and obtained similar horizontal-plane localization performance as for the full-bandwidth CTC. While their choice for the band limitation was based on computational issues, the lack of the mismatched CTC at higher frequencies, and thus, no decrease in the CS at these frequencies might also have contributed to their findings. Generally, it seems like a frequency-dependent amount of CTC might be useful in order to avoid a decrease of the CS in mismatched CTC systems. Note that such a procedure, however, might yield a drastic reduction of directional cues in the concerned frequencies.

D. Channel separation and localization performance

One quality aspect of a CTC system is the localization performance the system is able to provide. On the other hand, the CS is usually employed to describe the general quality of a CTC system. However, not much is known about the correspondence between the CS and localization performance. On the first sight, the relation between the mid-frequency CS and the performance seems to be weak. For example, for ctcOwn, ctcOwnB, and ctcOther, we found QE of 8.0%, 13.6%, and 16.9%, respectively, which corresponds to the mid-frequency CS of 68.4, 14.7, and 15.0 dB, respectively. While from ctcOwn to ctcOwnB, the increase in QE is well represented by the decrease in CS, the further increase in QE from ctcOwnB to ctcOther is not. Generally, mid-frequency CS in the range of 50 dB corresponded to a good localization performance. However, smaller CS (in the range between 13 to 18 dB) did not provide any statement on the localization performance. One example is NH72, who for quite different CSs (12.3 and 17.1 dB) showed nearly the same QE (23.2% and 23.9%). Another example is NH15, who for similar CSs (11.5 and 12.0 dB) showed completely different QEs (5% and 27.1%). Note that the same listener also showed QE of 6.8% in the condition having CS of 67.9 dB. This demonstrates the rather complex correspondence between the mid-frequency CS and the localization performance in terms of QEs. Similar pictures can be drawn for the low- and high-frequency CS.

TABLE V. Correlation coefficients for the correlation between the localization errors and the CS. Coefficients significantly ($p < 0.05$) different from zero are shown bold. The matched condition was represented by ctcOwn. The mismatched condition was represented by the conditions ctcOwnB and ctcOther.

Frequency range	0.3 to 8 kHz (mid)			0.3 to 2 kHz (low)			4 to 16 kHz (high)		
Condition	QE	PE	LE	QE	PE	LE	QE	PE	LE
All tested	**−0.35**	**−0.32**	**−0.43**	**−0.35**	**−0.33**	**−0.43**	**−0.32**	**−0.31**	**−0.38**
All tested, relative to binOwn	**−0.37**	**−0.36**	**−0.45**	**−0.39**	**−0.38**	**−0.49**	**−0.35**	**−0.35**	**−0.42**
Matched	0.00	−0.34	−0.31	0.21	0.03	0.51	0.30	−0.40	0.26
Matched, relative to binOwn	−0.25	−0.35	−0.27	−0.57	−0.36	−0.43	−0.25	−0.33	**−0.79**
Mismatched	**−0.50**	−0.28	**−0.60**	**−0.35**	**−0.33**	**−0.42**	**−0.33**	−0.16	−0.27
Mismatched, relative to binOwn	**−0.48**	−0.20	**−0.55**	**−0.38**	−0.24	**−0.53**	−0.33	−0.02	**−0.25**

In order to estimate the statistical relation between the CS and the localization performance, the correlation between the CS and the localization errors was analyzed. The correlation coefficients calculated for the mid-, low-, and high-frequency CS are shown in Table V. For all tested conditions, the correlation coefficient between mid-frequency CS and QE, PE, and LE was −0.35, −0.32, −0.43, respectively (all significant, $p < 0.025$). Similar correlations resulted for the low- and high-frequency CS. Such weak correlations suggest that, generally, CS is not a good predictor for the localization performance.

The correlation between the CS and localization performance might be confounded by the listener-individual performance in the localization task. In order to compensate for the listener-individual performance, the correlation coefficients were calculated between the CS and the performance relative to those obtained from the binOwn condition. All correlation coefficients increased (see Table V), with the largest coefficient being at −0.49. Hence, CS seems to be a poor predictor for the localization performance even when compensated for the listener-individual localization performance.

Since there is a large difference between the matched and mismatched CTC systems in the localization performance, CS might better correlate with the performance when compared separately for the matched and mismatched CTC conditions. For the matched condition (ctcOwn in Table V), most correlation coefficients were low and not significantly different from zero (= uncorrelated). The only significant ($p = 0.018$) correlation coefficient (−0.79) was found between the high-frequency CS and the LE relative to that obtained for binOwn. This might suggest that in matched CTC systems, the high-frequency CS is able to predict the horizontal-plane localization performance relative to the listener-individual performance. Note however, that despite the statistical support, this correlation is based on a small sample size ($n = 8$) and thus such a conclusion is to be treated with caution.

For the mismatched CTC systems, the largest significant correlation coefficients were −0.50 (QE and mid-frequency CS), −0.33 (PE and low-frequency CS), and −0.60 (LE and mid- frequency CS). These correlations did not improve when the relative localization performance was considered and show the extent to which CS might act as a predictor for the localization performance. Especially for the horizontal-plane localization performance, the correlation of −0.6 might be useful in further evaluations of CTC systems, depending on the application criteria. For the sagittal-plane localization performance, the CS seems to be a poor predictor. This is not surprising, considering that monaural, not interaural, cues are the most salient cues for the sagittal-plane localization.

IV. CONCLUSIONS

In this study, sound-localization performance was investigated in CTC reproduction systems tested under matched, individualized but mismatched, and non-individualized conditions. The performance was compared to the baseline binaural condition. CS, an objective measure for the quality of a CTC system, was calculated for the tested conditions and compared to the localization performance in a correlation analysis.

In the binaural conditions, the localization performance in terms of quadrant errors, polar errors, and lateral errors was within the range of previously reported performance. It was also the case when tested with HRTFs obtained from a measurement repeated approximately five years later, even though training was conducted only with the first HRTF set. This suggests that the human auditory localization system is robust to HRTF-measurement variability, at least for the measurement setup used in this study.

With the matched CTC systems, the performance was similar to that from the binaural conditions. With the individualized but mismatched CTC systems, where CTC filters were based on the repeated HRTF measurements, the listeners showed a degraded localization performance in terms of larger lateral, polar, and quadrant errors. This shows that the propagation paths from the loudspeakers to the ears must *exactly* match the filters in a CTC system in order to provide localization performance at the similar level as the binaural reproduction.

The direction-dependent analysis of the localization performance showed that in the mismatched CTC systems, the performance deteriorated especially for targets placed outside the loudspeaker span and/or behind the listener. With the non-individualized CTC systems, the quadrant errors further increased for the rear targets, but the performance for the frontal targets was in the range of that for the individualized but mismatched CTC system.

These findings show evidence that for targets placed within the loudspeaker span, the mismatch of the CTC system is not critical and the amount of CTC can be reduced in order to provide a better timbre reproduction. Much attention, however, should be given to the CTC systems for targets placed at other directions, in particular for the rear targets. As a work around to the currently unachievable matched CTC system, a second CTC system with loudspeakers placed behind the listener (Parodi and Rubak, 2011a) appears indeed to be a promising approach. For the more lateral targets, additional loudspeakers at lateral positions might help, which when combined with the loudspeakers in the rear, would form a ring of loudspeakers around the listener. Such a system would need to consider all available loudspeakers to choose the most adequate CTC filter for each source to be reproduced and might thus be seen as a combination of the wave-field synthesis (for a recent review, see Ahrens and Spors, 2012) and CTC. To our knowledge, such a combination of these systems has not been scientifically investigated yet.

A common quality metric for CTC systems is the CS. Our results show a substantial difference in CS between the matched and the mismatched CTC systems. However, CS was similar in both individualized and non-individualized mismatched CTC systems, even though the sagittal-plane localization performance was not. For the mismatched CTC systems, the CS was in the range of the natural CS provided by a stereophonic reproduction. The mismatched CTC systems were able to improve the CS in frequency range below 2 kHz, but they degraded the CS in the frequency range above 4 kHz, suggesting that mismatched CTC should be avoided at higher frequency regions. The matching had only little impact on the low-frequency CS. Hence, future efforts in the matching should focus on the mid- and high-frequency regions, at least for the tested loudspeaker span.

We observed a generally weak correlation between the CS and the sagittal-plane localization performance, even when compensating for the listener-individual localization performance in the binaural condition. Although the correlation improved slightly (-0.5) when only mismatched CTC systems were considered, CS does not seem to be an appropriate predictor for the sagittal-plane localization performance. A purely monaural metric like the equalization performance might be of advantage when describing the quality of CTC systems with respect to the sagittal-plane localization.

For the horizontal-plane, we expected a better correlation. It was -0.49, in general, and increased to -0.79 when only matched CTC systems were considered. This correlation was between the high-frequency CS and the lateral errors relative to the baseline performance; it was significantly different from uncorrelated, however, it was based on only eight samples.

For mismatched CTC systems only, the correlation of -0.6 was found and it was based on a more convincing sample size (40 samples). Such a correlation indicates that the mid-frequency CS might be indeed useful in evaluating mismatched CTC systems with respect to the horizontal-plane localization only.

ACKNOWLEDGMENTS

We thank Michael Mihocic for the help with the experiments and Bernhard Laback for fruitful discussions on this topic. We are also grateful to our subjects for their patience during the experiments. Further, we thank the two anonymous reviewers for their useful comments. B.M. is a scholarship holder from CNPq, Brazil.

[1]http://www.kfs.oeaw.ac.at/hrtf (Last viewed 6/9/2012).

[2]Note that 85.33 ms correspond to 4096 samples since we used a sampling rate of 48 kHz for all calculations.

[3]Note that a frequency-dependent regularization would offer more degrees of freedom in the CTC filter design, however, our intention was to be consistent with Akeroyd *et al*. (2007).

[4]http://www.ita-toolbox.org (Last viewed 6/9/2012).

[5]http://sf.net/projects/expsuite (Last viewed 6/9/2012).

[6]The choice of using either binOwn or binOwnB for the playback HRTFs was arbitrary. In realistic CTC reproduction systems, a single HRTF set is measured and assumed to represent the propagation paths, which actually remain unknown.

[7]The performance varied also across particular non-individual conditions (compare ctcKEMAR versus ctcNH68), however, in order to increase the statistical power, we pooled all non-individual CTC conditions in our analysis. An analysis of each non-individual CTC condition and its impact on the localization performance would be outside the scope of this study.

Ahrens, J., and Spors, S. (**2012**). "Wave field synthesis of a sound field described by spherical harmonics expansion coefficients," J. Acoust. Soc. Am. **131**, 2190–2199.

Akeroyd, M. A., Chambers, J., Bullock, D., Palmer, A. R., Summerfield, A. Q., Nelson, P. A., and Gatehouse, S. (**2007**). "The binaural performance of a cross-talk cancellation system with matched or mismatched setup and playback acoustics," J. Acoust. Soc. Am. **121**, 1056–1069.

Atal, B. S., Hill, M., and Schroeder, M. R. (**1966**). "Apparent sound source translator," U.S. patent code no. 3236949, granted on 1966 Feb. 22.

Bai, M. R., and Lee, C. (**2006**). "Objective and subjective analysis of effects of listening angle on crosstalk cancellation in spatial sound reproduction," J. Acoust. Soc. Am. **120**, 1976–1989.

Bai, M. R., and Lee, C. (**2007**). "Subband approach to bandlimited crosstalk cancellation system in spatial sound reproduction," EURASIP J. Adv. Signal Process. **2007**, 071948.

Bauck, J., and Cooper, D. H. (**1996**). "Generalized transaural stereo and applications," J. Audio Eng. Soc. **44**, 683–705.

Bauer, B. B. (**1961**). "Stereophonic earphones and binaural loudspeakers," J. Audio Eng. Soc. **9**, 148–151.

Best, V., Carlile, S., Jin, C., and van Schaik, A. (**2005**). "The role of high frequencies in speech localization," J. Acoust. Soc. Am. **118**, 353–363.

Blauert, J. (**1969/70**). "Sound localization in the median plane," Acustica **22**, 205–213.

Blauert, J. (**1997**). *Spatial Hearing*, 2nd ed. (MIT Press, Cambridge, MA), p. 497.

Carlile, S., and Pralong, D. (**1994**). "The location-dependent nature of perceptually salient features of the human head-related transfer functions," J. Acoust. Soc. Am. **95**, 3445–3459.

Gardner, W. G. (**1997**). *3-D Audio Using Loudspeakers* (Springer, New York), p. 168.

Gardner, W. G., and Martin, K. D. (**1995**). "HRTF measurements of a KEMAR," J. Acoust. Soc. Am. **97**, 3907–3908.

Goupell, M. J., Majdak, P., and Laback, B. (**2010**). "Median-plane sound localization as a function of the number of spectral channels using a channel vocoder," J. Acoust. Soc. Am. **127**, 990–1001.

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

Langendijk, E. H. A., and Bronkhorst, A. W. (**2002**). "Contribution of spectral cues to human sound localization," J. Acoust. Soc. Am. **112**, 1583–1596.

Lentz, T. (**2006**). "Dynamic crosstalk cancellation for binaural synthesis in virtual reality environments," J. Audio Eng. Soc. **54**, 283–294.

Macpherson, E. A., and Middlebrooks, J. C. (**2002**). "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," J. Acoust. Soc. Am. **111**, 2219–2236.

Majdak, P., Balazs, P., and Laback, B. (**2007**). "Multiple exponential sweep method for fast measurement of head-related transfer functions," J. Audio Eng. Soc. **55**, 623–637.

Majdak, P., Goupell, M. J., and Laback, B. (**2010**). "3-D localization of virtual sound sources: Effects of visual environment, pointing method, and training," Atten. Percept. Psychophys. **72**, 454–469.

Majdak, P., Goupell, M. J., and Laback, B. (**2011**). "Two-dimensional localization of virtual sound sources in cochlear-implant listeners," Ear Hear. **32**, 198–208.

Middlebrooks, J. C. (**1999**). "Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency," J. Acoust. Soc. Am. **106**, 1493–1510.

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.

Morimoto, M., and Aokata, H. (**1984**). "Localization cues in the upper hemisphere," J. Acoust. Soc. Jpn. (E) **5**, 165–173.

Parodi, Y. L., and Rubak, P. (**2010**). "Objective evaluation of the sweet spot size in spatial sound reproduction using elevated loudspeakers," J. Acoust. Soc. Am. **128**, 1045–1055.

Parodi, Y. L., and Rubak, P. (**2011a**). "Sweet spot size in virtual sound reproduction: A temporal analysis," in *Principles And Applications of Spatial Hearing*, edited by Y. Suzuki, D. Brungart, Y. Iwaya, K. Iida, D. Cabrera, and H. Kato (World Scientific, Singapore), pp. 292–302.

Parodi, Y. L., and Rubak, P. (**2011b**). "A subjective evaluation of the minimum channel separation for reproducing binaural signals over loudspeakers," J. Audio Eng. Soc. **59**, 487–497.

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

Qiu, X., Masiero, B., and Vorländer, M. (**2009**). "Channel separation of crosstalk cancellation systems with mismatched and misaligned sound sources," J. Acoust. Soc. Am. **126**, 1796–1806.

Rose, J., Nelson, P., Rafaely, B., and Takeuchi, T. (**2002**). "Sweet spot size of virtual acoustic imaging systems at asymmetric listener locations," J. Acoust. Soc. Am. **112**, 1992–2002.

Strutt, J. W. (**1907**). "On our perception of sound direction," Philos. Mag. **13**, 214–232.

Takemoto, H., Mokhtari, P., Kato, H., Nishimura, R., and Iida, K. (**2012**). "Mechanism for generating peaks and notches of head-related transfer functions in the median plane," J. Acoust. Soc. Am. **132**, 3832–3841.

Takeuchi, T., and Nelson, P. A. (**2002**). "Optimal source distribution for binaural synthesis over loudspeakers," J. Acoust. Soc. Am. **112**, 2786–2797.

Takeuchi, T., and Nelson, P. A. (**2007**). "Subjective and objective evaluation of the optimal source distribution for virtual acoustic imaging," J. Audio Eng. Soc. **55**, 981–997.

Takeuchi, T., Nelson, P. A., and Hamada, H. (**2001**). "Robustness to head misalignment of virtual sound imaging systems," J. Acoust. Soc. Am. **109**, 958–971.

Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (**1993**). "Localization using nonindividualized head-related transfer functions," J. Acoust. Soc. Am. **94**, 111–123.

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.

Wightman, F. L., and Kistler, D. J. (**1992**). "The dominant role of low-frequency interaural time differences in sound localization," J. Acoust. Soc. Am. **91**, 1648–1661.

Wightman, F. L., and Kistler, D. J. (**1997a**). "Factors affecting the relative salience of sound localization cues," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Lawrence Erlbaum Associates, Mahwah, NJ), pp. 1–23.

Wightman, F. L., and Kistler, D. J. (**1997b**). "Monaural sound localization revisited," J. Acoust. Soc. Am. **101**, 1050–1063.