Assessment of Sagittal-Plane Sound Localization Performance in Spatial-Audio Applications

R. Baumgartner, P. Majdak and B. Laback

Acoustics Research Institute, Austrian Academy of Sciences, Vienna, Austria

Summary. Sound localization in sagittal planes, SPs, including front-back discrimination, relies on spectral cues resulting from the filtering of incoming sounds by the torso, head and pinna. While acoustic spectral features are well-described by head-related transfer functions, HRTFs, models for SP localization performance have received little attention. In this article, a model predicting SP localization performance of human listeners is described. Listener-specific calibrations are provided for 17 listeners as a basis to predict localization performance in various applications. In order to demonstrate the potential of this listener-specific model approach, predictions for three applications are provided, namely, the evaluation of non-individualized HRTFs for binaural recordings, the assessment of the quality of spatial cues for the design of hearing-assist devices and the estimation and improvement of the perceived direction of phantom sources in surround-sound systems.

1 Sound Localization in Sagittal Planes

1.1 Salient Cues

Human normal-hearing, NH, listeners are able to localize sounds in space in terms of assigning direction and distance to the perceived auditory image [26]. Multiple mechanisms are used to estimate sound-source direction in the three-dimensional space. While interaural differences in time and intensity are important for sound localization in the lateral dimension, left/right, [53], monaural spectral cues are assumed to be the most salient cues for sound localization in the sagittal planes, SPs, [54, 27]. SPs are planes parallel to the median plane and include points of similar interaural time differences for a given distance. The *monaural* spectral cues are essential for the perception of the source elevation within a hemifield [2, 22, 24] and for front-back discrimination of the perceived auditory event [56, 46]. Note that also the binaural pinna disparities [43], namely, interaural spectral differences, might contribute to SP localization [27].

The mechanisms underlying the perception of lateral displacement are the main topic of other chapters. This chapter focuses on the remaining directional dimension, namely, the one along SPs. Because interaural cues and monaural spectral cues are thought to be processed largely independently of each other [27], the interaural-polar coordinate system is often used to describe their respective contributions in the two dimensions. In the interaural-polar coordinate system the direction of a sound source is described with the lateral angle, $\phi \in [-90°, 90°]$, and the polar angle, $\theta \in [-90°, 270°)$ – see Fig. 1, left panel. SP localization refers to the listener's assignment of the polar angle for a given lateral angle and distance of the sound source.

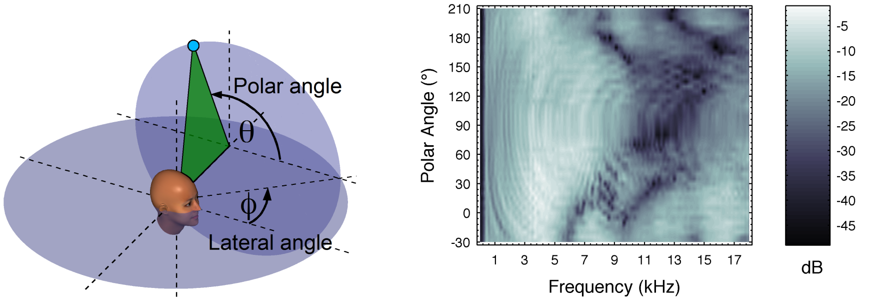


Fig. 1. Left: interaural-polar coordinate system. **Right**: HRTF magnitude spectra of a listener as a function of the polar angle in the median SP – left ear of NH58

Although spectral cues are processed monaurally, the information from both ears affects the perceived location in most cases [39]. The ipsilateral ear, namely, the one closer to the source, dominates and its relative contribution increases monotonically with increasing lateral angle [12]. If the lateral angle exceeds about 60°, the contribution of the contralateral ear becomes negligible. Thus, even for localization in the SPs, the lateral source position, mostly depending on the broadband binaural cues [27], must be known in order to determine the binaural weighting of the monaural spectral cues.

The nature of the spectral features important for sound localization is still subject of investigations. Due to the physical dimensions, the pinna plays a larger role for higher frequencies [36] and the torso for lower frequencies [1]. Some psychoacoustic studies postulated that macroscopic patterns of the spectral features are important rather than fine spectral details [2, 22, 24, 44, 28, 23, 16, 10]. On the other hand, other studies postulated that SP sound localization is possibly mediated by means of only a few local spectral features [52, 37, 17, 56]. Despite a common agreement, according to which the amount of the spectral features can be reduced without substantial reduction of the localization performance, the perceptual relevance of particular features has not been fully clarified yet.

1.2 Head-related Transfer Functions

The effect of the acoustic filtering of torso, head and pinna can be described in terms of a linear time-invariant system by the so-called head-related transfer functions, HRTFs, [4, 45, 38]. The right panel of Fig. 1 shows the HRTF magnitude spectra of an exemplary listener, NH58, left ear[1], along the median SP.

HRTFs depend on the individual geometry of the listener and thus listener-specific HRTFs are required to achieve accurate localization performance for binaural synthesis [6, 35]. Usually, HRTFs are measured in an anechoic chamber by determining the acoustic response characteristics between loudspeakers at various directions and microphones inserted into the ear canals. Currently, much effort is put also into the development of non-contact measurement methods for capturing HRTFs like numerical calculation of HRTFs from optically scanned geometry [20, 21] and on customization of HRTFs basing on psychoacoustic tests [34, 16, 46].

Measured HRTFs contain both direction-dependent and direction-independent features and can be thought of as a series of two acoustic filters. The direction-independent filter, represented by the common transfer function, CTF, can be calculated from an HRTF set comprising many directions [34] by averaging the log-amplitude spectra of all available HRTFs of a listener's ear. The phase spectrum of the CTF is the minimum phase corresponding to the amplitude spectrum of the CTF.

In the current study, the topic of interest is the directional aspect. Thus, the directional features are considered, as represented by the directional transfer functions, DTFs. The DTF for a particular direction is calculated by filtering the corresponding HRTF with the inverse CTF. The CTF usually exhibits a low-pass filter characteristic because the higher frequencies are attenuated for many directions due to the head and pinna shadow – see Fig. 2, left panel.

[1] These and all other HRTFs are from `http://www.kfs.oeaw.ac.at/hrtf`

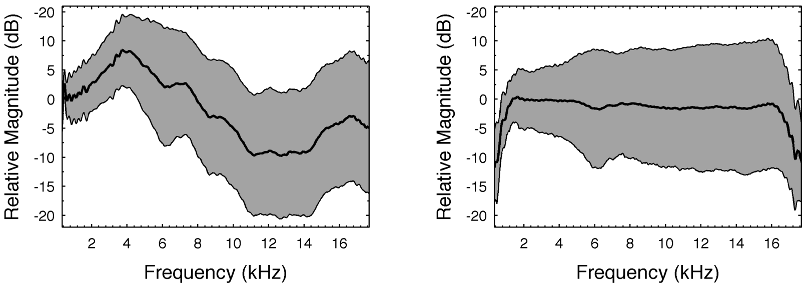


Fig. 2. Left: spatial variation of HRTFs around CTF for listener NH58, left ear. **Right**: corresponding DTFs, i.e. HRTFs with CTF removed. *Solid line*: spatial average of transfer function. *Grey area*: ± 1 standard deviation

Compared to HRTFs, DTFs usually pronounce frequencies and thus spectral features above 4 kHz – see Fig. 2, right panel. DTFs are commonly used to investigate the nature of spectral cues in SP localization experiments with virtual sources [34, 10, 30].

In the following, the proposed model is described in Sect. 2 and the results of its evaluation are presented in Sect. 3, based on recent virtual-acoustics studies that used listener-specific HRTFs. In Sect. 4, the proposed model is applied to predict localization performance for different aspects of spatial-audio applications that involve spectral localization cues. In particular, a focus is put on the evaluation of non-individualized binaural recordings, the assessment of the quality of spatial cues for the design of hearing-assist devices, namely, in-the-ear *vs.* behind-the-ear microphones and the estimation and improvement of the perceived direction of phantom sources in surround-sound systems, namely, 5.1 *vs.* 9.1 *vs.* 10.2 surround. Finally, Sect. 5 concludes with a discussion of the potential of the model for both evaluating audio applications and improving the understanding of human sound-localization mechanisms.

2 Models of Sagittal-plane Localization

This section considers existing models aiming at predicting listener's polar response angle to the incoming sound. These models can help to explain psychoacoustic phenomena or to assess the spatial quality of audio systems while avoiding the running of costly and time-consuming localization experiments.

In general, machine-learning approaches can be used to predict localization performance. Artificial neural networks, ANNs, have been shown to achieve rather accurate predictions when trained with large datasets of a single listener [19]. However, predictions for a larger subpopulation of human listeners would have required much more effort. Also, the interpretation of the ANN parameters is not straight forward. It is difficult to generalize the findings obtained with an ANN-based model to other signals, persons and conditions and thus to better understand the mechanisms underlying spatial hearing.

Hence, the focus is laid on a *functional* model where model parameters should correspond to physiological and/or psychophysical localization parameters. Until now, a functional model considering both spectral and temporal modulations exists only as a general concept [50]. Note that in order to address a particular research question, models dealing with specific types of modulations have been designed. For example, models for narrow-band sounds [37] were provided in order to explain the well-known effect of directional bands [4]. In order to achieve a sufficiently good prediction as an effect of the modification of the spectral cues, it is assumed that the incoming sound is a *stationary broadband* signal, explicitly disregarding spectral and temporal modulations.

Note that localization models driven by various signal-processing approaches have also been developed [3, 32, 33]. These models are based on general principles of biological auditory systems, they do not, however, attempt to predict human-listener performance – their outcome shows rather the potential of the signal-processing algorithms involved.

In the following, previous developments on modeling SP localization performance are reviewed and a functional model predicting sound localization performance in arbitrary SPs for broadband sounds is proposed. The model is designed to retrieve psychophysical localization performance parameters and can be directly used as a tool to assess localization performance in various applications. An implementation of the model is provided in the auditory modeling toolbox, AMT, as the `baumgartner2013` model [47].

2.1 Template-based Comparison

A common property of existing sound localization models based on spectral cues is that they compare an internal representation of the incoming sound with a template [55, 13, 24] – see Fig. 3. The internal template is assumed to be created by means of learning the correspondence between the spectral features and direction of an acoustic event [14, 49], based on feedback from other modalities. The localization performance is predicted by assuming that in the sound localization task, the comparison yields a distance metric that corresponds to the polar response angle of the listener. Thus, template-based models include a stage modeling the peripheral processing of the auditory system applied to both the template and incoming sound and a stage modeling the comparison process in the brain.

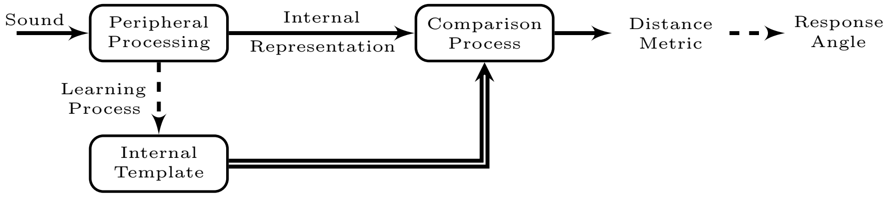


Fig. 3. General structure of a template-based comparison model for predicting localization in SPs

Peripheral processing

The peripheral processing stage is aimed at modeling the effect of human physiology while focusing on directional cues. The effect of the torso, head and outer ear are considered by filtering the incoming sound by an HRTF or a DTF. The effect of ear canal, middle ear and cochlear filtering can be considered by various model approximations. In the early HRTF-based localization models, a parabolic-shaped filter bank was applied [55]. Later, a filter bank averaging magnitude bins of the discrete Fourier transform of the incoming sound was used [24]. Both filter banks, while being computationally efficient, were drastically simplifying the auditory peripheral processing. The Gammatone, GT, filter bank [40] is a more physiology-related linear model of auditory filters and has been used in localization models [13]. A model accounting for the nonlinear effect of the cochlear compression is the dual-resonance nonlinear, DRNL, filter bank [25]. A DRNL filter consists of both a linear and a non-linear processing chain and is implemented by cascading GT filters and Butterworth low-pass filters, respectively. Another non-linear model uses a single main processing chain and accounts for the time-varying effects of the medial-oliviocochlear reflex [57]. All those models account for the contribution of outer hair cells to a different degree and can be used to model the movements of the basilar membrane at a particular frequency. They are implemented in the AMT [47]. In the localization model proposed in this section, the GT filter bank is applied with concentration on applications where the absolute sound level plays a minor role.

The filter bank produces a signal for each center frequency and only the relevant frequency bands are considered in the model. Existing models used frequency bands with constant relative bandwidth on a logarithmic frequency scale [55, 24]. In the model proposed in this section, the frequency spacing of the bands corresponds to one equivalent rectangular bandwidth, ERB, [9]. The lowest frequency is 0.7 kHz, corresponding to the minimum frequency thought to be affected by torso reflections [1]. The highest frequency considered in the model depends on the bandwidth of the incoming sound and is maximally 18 kHz, approximating the upper frequency limit of human hearing.

Further in the auditory system, the movements of the basilar membrane at each frequency band are translated into neural spikes by the inner hair cells, IHCs. An accurate IHC model has not been considered yet and does not seem to be vital for SP localization. Thus, different researches used different approximations. In this model, the IHC is modeled as half-wave rectification followed by a second-order Butterworth low-pass with a cut-off frequency of 1 kHz [8]. Since the temporal effects of SP localization are not considered yet, the output of each band is simply temporally averaged in terms of RMS amplitude, resulting in the internal representation of the sound. The same internal representation and, thus, peripheral processing, is assumed for the template.

Comparison stage

In the comparison stage, the internal representation of the incoming sound is compared with the internal template. Each entry of the template is selected by a polar angle denoted as template angle. A distance metric is calculated as a function of the template angle and can be interpreted as a potential descriptor for the response of the listener.

An early modeling approach proposed to compare the spectral derivatives of various orders in terms of a band-wise subtraction of the derivatives and then averaging over the bands [55]. The comparison of the first-order derivative corresponds to the assumption that the overall sound intensity does not contribute to the localization process. In the comparison of the second-order derivatives, the differences in spectral tilt between the sound and the template do not contribute. Note that the plausibility of these comparison methods had not been investigated at that time. As another approach, Pearson's correlation has been proposed to evaluate the similarity between the sound and the template [37, 13]. Later, the inter-spectral differences, ISDs, namely, the differences between the internal representations of the incoming sound and the template calculated for each template angle and frequency band, were used [34] to show a correspondence between the template angle yielding smallest spectral variance and the actual response of human listeners. All these comparison approaches were tested in [24] who, distinguishing zeroth-, first- and second-order derivatives of the internal representations, found that the standard deviation of ISDs best described their results. This configuration corresponds to an average of the first-order derivative from [55], which is robust against changes in the overall level in the comparison process.

The model proposed in this study also relies on ISDs calculated for a template angle and for each frequency band – see Fig. 4, left panel. Then, the spectral standard deviations of ISDs are calculated for all available template angles – see Fig. 4, right panel. For band-limited sounds, the internal representation results in an abrupt change at the cut-off frequency of the sound. This change affects the standard deviation of the ISDs. Thus, in this model, the ISDs are calculated only within the bandwidth of the incoming sound.

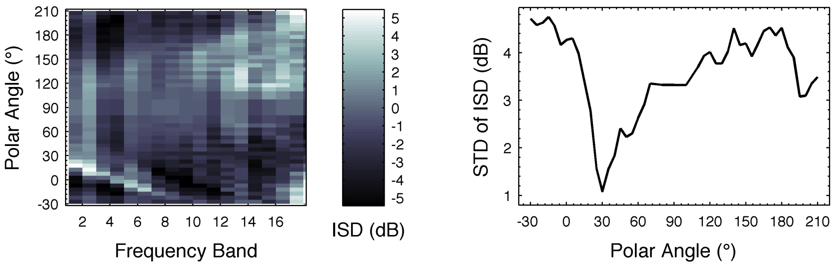


Fig. 4. Example of the comparison process for a target polar angle of 30°. **Left**: ISDs as a function of the template angle. **Right**: spectral standard deviation, STD, of ISDs as a function of the template angle

The result of the comparison stage is a distance metric corresponding to the prediction of the polar response angle. Early modeling approaches used the minimum distance to determine the predicted response angle [55], which would nicely fit the minimum of the distance metric used in the example reported here – see Fig. 4, right panel. Also, the cross-correlation coefficient has been used as a distance metric and its maximum has been interpreted as the prediction of the response angle [37]. Both approaches represent a deterministic interpretation of the distance metric, resulting in exactly the same predictions for the same sounds. This is rather unrealistic. Listeners, repeatedly listening to the same sound, often do not respond to exactly the same direction [7]. The actual responses are known to be scattered and can be even multimodal. The scatter of one mode can be described by the Kent distribution [7], which is an elliptical probability distribution on the two-dimensional unit sphere.

2.2 Response Probability

In order to model the probabilistic response pattern of listeners, a mapping of the distance metric to polar-response probabilities via similarity indices, SIs, has been proposed [24]. For a particular target angle and ear, a monaural SI has been obtained by using the distance metric as the argument of a Gaussian function with a mean of zero and a standard deviation of two – see Fig. 5, $U = 2$. While this choice appears to be somewhat arbitrary, it models the probabilistic relation between the distance metric and the probability of responding to a given direction. Note that the resulting SI is bounded by zero and one and valid for the analysis of the incoming sound at one ear only.

The width of the mapping function, U in Fig. 5, actually reflects a property of an individual listener. A listener being more precise in the response to the same sound would need a more narrow mapping than a less precise listener. Thus, in contrast to the previous approach [24], in the model proposed in this section, the width of the mapping function as a listener-specific uncertainty, U, is considered. It accounts for listener-specific localization precision [34, 42, 56] due to reasons like training and attention [14, 51]. Note that for simplicity, direction-dependent response precision is neglected. The lower the uncertainty, U, the higher the assumed sensitivity of the listener to distinguish spectral features. In the next section, this parameter will be used to calibrate the model to listener-specific performance.

The model stages described so far are monaural. Thus, they do not consider binaural cues and have been designed for the median SP where the interaural differences are zero and thus binaural cues do not contribute. In order to take into account the contribution of both ears, the monaural model results for both ears are combined. Previous approaches averaged the monaural SIs for both ears [24] and thus were able to consider the contribution of both ears for targets placed in the median SP. In the model proposed in this section, the lateral target range is extended to arbitrary SPs by applying a binaural weighting function [12, 29], which reduces the contribution of the contralateral ear depending on the perceived lateral direction of the target sound. Thus, the binaural weighting function is applied to each monaural SI, and the sum of the weighted monaural SIs yields the binaural SI.

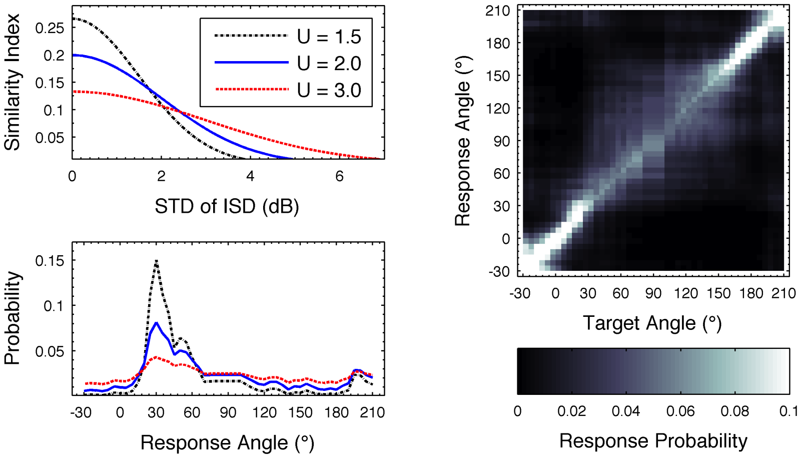


Fig. 5. Left: mapping function of SI, *top*, for various uncertainties, U, and the resulting PMVs, *bottom* – corresponding to the example shown in Fig. 4. **Right**: predicted response PMV of the localization model as a function of the target angle, i.e. prediction matrix, for the baseline condition in the median SP for listener NH58. Response probabilities are encoded by brightness

For an incoming sound, the binaural SIs are calculated for all template entries selected by the template angle. Such a binaural SI as a function of the template angle is related to the listener's response probability as a function of the response angle. It can be interpreted as a discrete version of a probability density function, namely, a probability mass vector, PMV, showing the probability of responding at an angle to a particular target. In order to obtain a PMV, the binaural SI is normalized to have a sum of one. Note that this normalization assumes that the template angles regularly sample an SP. If this is not the case, regularization by spline interpolation is applied before the normalization.

The PMVs, calculated separately for each target under consideration, are represented in a prediction matrix. This matrix describes the probability of responding at a polar angle given a target placed at a specific angle. The right panel of Fig. 5 shows the prediction matrix resulting for the exemplary listener, NH58, in a baseline condition where the listener uses his/her own DTFs, and all available listener-specific DTFs are used as targets. The abscissa shows the target angle, the ordinate shows the response angle and the brightness represents the response probability. This representation is used throughout the following sections and it also allows for a visual comparison between the model predictions and the responses obtained from actual localization experiments.

2.3 Interpretation of the Probabilistic Model Predictions

In order to compare the probabilistic results from the model with the experimental results, likelihood statistics, calculated for actual responses from sound localization experiments and for responses resulting from virtual experiments driven by the model prediction, can be used – see equation (1) in [24]. The comparison between the two likelihoods allows one to evaluate the validity of the model, because only for similar likelihoods the model is assumed to yield valid predictions. The likelihood has, however, a weak correspondence with localization performance parameters commonly used in psychophysics.

Localization performance in the polar dimension usually considers local errors and hemifield confusions [35]. Although these errors derived by geometrical aspects cannot sufficiently represent the current understanding of human hearing, they are frequently used and thus enable comparison of results between studies. Quadrant errors, QEs, that is the percentage of polar errors larger or equal to 90°, represent the confusions between hemifields – for instance, front/back or up/down – without considering the local response pattern. Unimodal local responses can be represented as a Kent distribution [7], which, considering the polar dimension only, can be approximated by the polar bias and polar variance. Thus, the local errors are calculated only for local responses within the correct hemifield, namely, without the responses yielding the QEs. A single representation of the local errors is the local polar RMS error, PE, which combines localization bias and variance in a single metric.

In the proposed model, QEs and PEs are calculated from the PMVs. The QE is the sum of the PMV entries outside the local polar range for which the response-target difference is greater or equal to 90°. The PE is the discrete expectancy value within the local polar range. In the visualization of prediction matrices – see for example right column of Fig. 5 – bright areas in the upper left and bottom right corners would indicate large QEs, a strong concentration of the brightness at the diagonal would indicate small PEs. Both

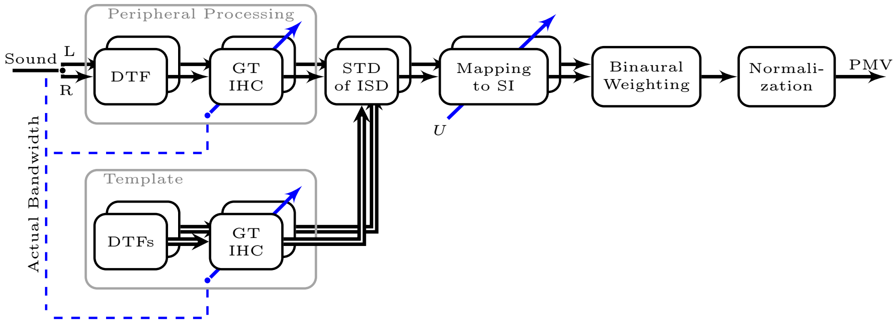


Fig. 6. Structure of the proposed SP localization model – see text for the description of the stages

errors can be calculated either for a specific target angle or as the arithmetic average across all target angles considered in the prediction matrix.

Figure 6 summarizes the final structure of the model. It requires the incoming signal from a sound source as the input and results in the response probability as a function of response angle, namely, PMV, for given template DTFs. Then, from PMVs calculated for the available target angles, QEs and PEs are calculated for a direct comparison with the outcome of a sound-localization experiment.

3 Listener-specific Calibration and Evaluation

Listeners show an individual localization performance even when localizing broadband sounds in free field [31]. While the listener-specific differences in the HRTFs may play a role, also other factors like experience, attention, or utilization of auditory cues might be responsible for differences in the localization performance. Thus, this section is concerned with the calibration of the model for each particular listener. By creating calibrations for 17 listeners, a pool of listener-specific models is provided. In order to estimate the use of this pool in future applications, the performance of this pool is evaluated in two experiments. In Sect. 4, the pool is applied to various applications.

3.1 Calibration: Pool of Listener-specific Models

The SP localization model is calibrated to the baseline performance of a listener in terms of finding an optimal uncertainty, U. Recall that the lower the uncertainty, U, the higher the assumed efficiency of the listener in evaluating spectral features. An optimal U minimizes the difference between the predicted and the listener's actual baseline performance in terms of a joint metric of QE and PE, namely, the $\mathcal{L}^2$-norm.

The actual baseline performance was obtained in localization experiments where a listener was localizing sounds using his/her own DTFs presented via headphones. Gaussian white noise bursts with a duration of 500 ms and a fade-in/out of 10 ms were used as stimuli. The acoustic targets were available for elevations from $-30°$ to $80°$ in the lateral range of at least $\pm 30°$ around the median SP. Listeners responded by manually pointing to the perceived direction of a target. For more details on the experimental methods see [30, 10, 51].

The model predictions were calculated considering SPs within the lateral range of $\pm 30°$. The targets were clustered to SPs with a width of 20° each. For the peripheral processing, the lower and upper corner frequency was 0.7 and 18 kHz, respectively, resulting in 18 frequency bands with a spacing of one ERB.

Table 1 shows the values of the uncertainty, U, for the pool of 17 listeners. The impact of the calibration becomes striking by comparing the predictions

Table 1. Values of the uncertainty U for the pool of listener-specific models identified by listener IDs, NHn.

NHn	12	15	21	22	33	39	41	42	43	46	55	58	62	64	69	71	72
U	1.6	2.0	1.8	2.0	2.3	2.3	3.0	1.8	1.9	1.8	2.0	1.4	2.2	2.1	2.1	2.1	2.2

based on the listener-specific, calibrated pool with the predictions basing on the pool using $U = 2$ for all listeners as in [24]. Figure 7 shows the actual and predicted performance as a comparison with a pool calibrated to $U = 2$ for all listeners and a listener-specific calibrated pool. Note the substantially higher correlation between the prediction with the actual results in the case of the listener-specific calibration. The correlation coefficients in the order of $r = 0.85$ provide evidence for sufficient power in the predictions for the pool.

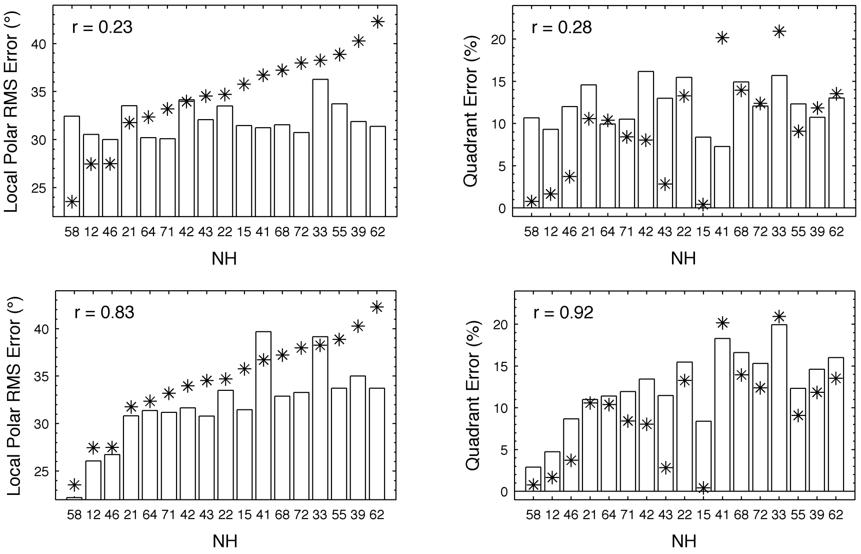


Fig. 7. Localization performance, PE, QE. *Bars*: predicted by the model. *Asterisks*: actual performance obtained in sound localization experiments. **Top**: model predictions for $U = 2$ as in [24]. **Bottom**: model predictions for listener-specific calibration. r... Pearson's correlation coefficient with respect to actual and predicted performance

3.2 Evaluation

In order to evaluate the SP localization model, the experimental data from two studies investigating stationary broadband sounds are modeled and compared to the experimental results. Only two studies were available because both the listener-specific HRTFs and the corresponding responses are necessary for the evaluation. For each of these studies, two predictions are calculated, namely, one for the listeners who actually participated in that experiment and one for the whole pool of listener-specific, calibrated models. For the participants, the predictions are done on the basis of the actual targets, whereas for the pool, all targets are considered by randomly drawing from the available DTFs.

Effect of the number of spectral channels

A previous study tested the effect of the number of spectral channels on the localization performance in the median SP [10]. While that study was focused on cochlear-implant processing, the localization experiments were done on listeners with normal hearing using a Gaussian-envelope tone vocoder – for more details see [10]. The frequency range of 0.3–16 kHz was divided into 3,

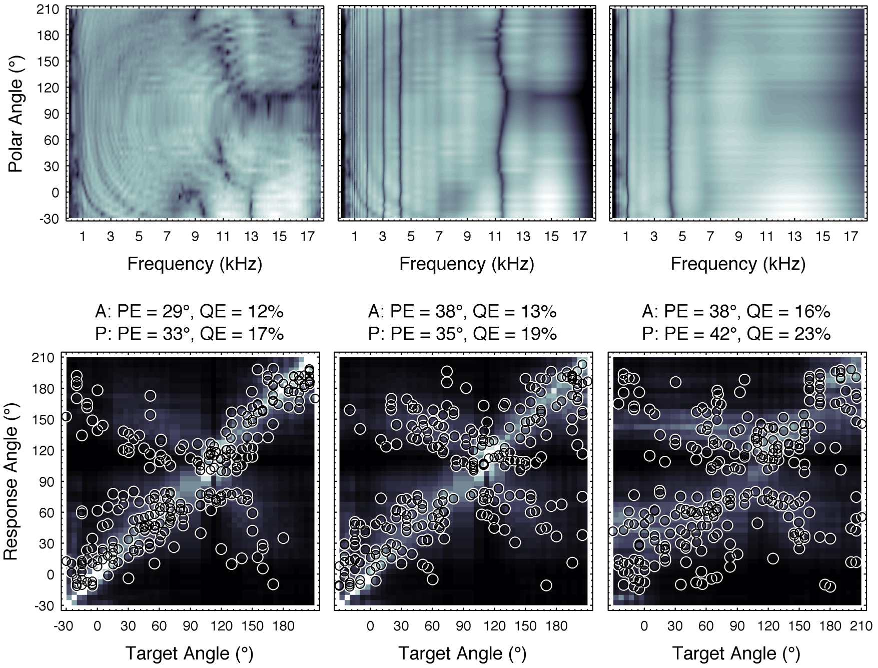


Fig. 8. Effect of the number of spectral channels for NH42. **Top**: channelized DTFs of median SP, left ear, brightness-encoded magnitude as in Fig. 1, right panel – from [10]. **Bottom**: prediction matrices with brightness-encoded probability as in Fig. 5, right panel, and actual responses, *open circles*. **Left**: unlimited number of channels. **Center**: 24 spectral channels. **Right**: 9 spectral channels. A. . . actual performance from [10], P. . . predicted performance

6, 9, 12, 18, or 24 channels, equally spaced on the logarithmic frequency scale. The top row of Fig. 8 shows the channelized DTFs from an exemplary listener.

The bottom row of Fig. 8 shows the corresponding prediction matrices including the actual responses – open circles – for this particular listener. Note the correspondence of the localization performance for that particular listener between the actual responses, A, and the model predictions, P. Good correspondence between the actual responses and prediction matrices was found for most of the tested listeners, which is supported by the overall response-prediction-correlation coefficients of 0.62 and 0.74 for PE and QE, respectively.

Figure 9 shows the predicted and the actual performance as averages over the listeners. In comparison to the actual performance, the models underestimated the PEs for 12 and 18 channels and overestimated them for 3 channels. The predictions for the pool seem to follow the predictions for the actually tested listeners showing generally similar QEs but slightly smaller PEs. While the analysis of the nature of these errors is outside of the focus of this chapter, both predictions, those for the actual listeners and those for the pool, seem to well represent the actual performance in this localization experiment.

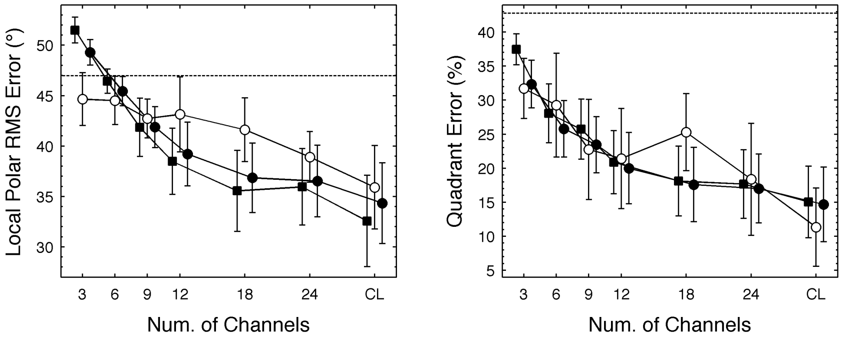


Fig. 9. Localization performance, namely, PE and QE, for listener groups as functions of the number of spectral channels. *Open circles*: actual performance of the listeners replotted from [10]. *Filled circles*: performance predicted for the listeners tested in [10] using the targets from [10]. *Filled squares*: performance predicted for the listener pool, using randomly chosen targets. *Error bars*: ±1 standard deviations of the average over the listeners. *Dashed line*: chance performance corresponding to guessing the direction of the sound. CL... unlimited number of channels, broadband clicks

Effect of band limitation and spectral warping

In another previous study, localization performance was tested in listeners using their original DTFs, band-limited DTFs and spectrally warped DTFs

[51]. The band limitation was done at 8.5 kHz. The spectral warping compressed the spectral features in each DTF from the range 2.8–16 kHz to the range 2.8–8.5 kHz. While the focus of that study was to estimate the potential of re-learning sound localization with drastically modified spectral cues in a training paradigm, the experimental *ad-hoc* results from the pre-experiment are used to evaluate the proposed model. Note that, for this purpose, the upper frequency of the peripheral processing stage was configured to 8.5 kHz for the band-limited and warped conditions.

The top row of Fig. 10 shows the DTFs and the bottom row the prediction matrices for the original, band-limited and warped conditions for the exemplary listener, NH12. The actual responses – open circles – show a good correspondence to the prediction matrices. Figure 11 shows group averages of the experimental results and the corresponding predictions. The group averages show a good correspondence between the actual and predicted performance. The correlation coefficient between the actual responses and predictions was 0.81 and 0.85 for PE and QE, respectively. The predictions of the pool well reflect the group averages of the actual responses.

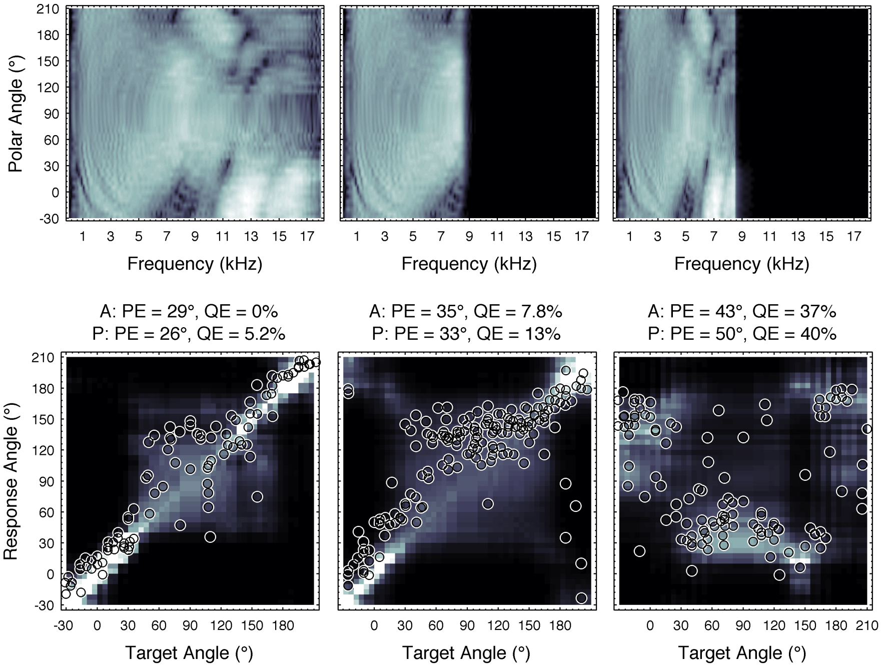


Fig. 10. Localization with the different DTFs, namely, original, *left column*, band-limited, *center column*, and spectrally warped, *right column*. **Top**: DTFs from NH12, left ear, in the median SP. **Bottom**: prediction matrices for NH12. *Open circles*: actual responses for NH12 from [51]. All other conventions are as in Fig. 8

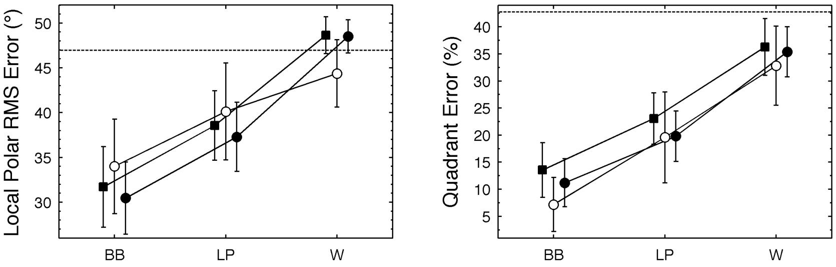


Fig. 11. Localization performance for listener groups in conditions broadband, BB, band-limited, LP, and spectrally warped, W. *Open circles*: Actual performance of the tested listeners from [51]. All other conventions are as in Fig. 9

4 Applications

The evaluation from the previous section shows response-prediction correlation coefficients in the order of 0.75. This indicates that the proposed model is reliable in predicting localization performance when applied with the listener-specific calibrations. Thus, in this section, the calibrated models are applied to predict localization performance in order to address issues potentially interesting in spatial-audio applications.

4.1 Non-individualized Binaural Recordings

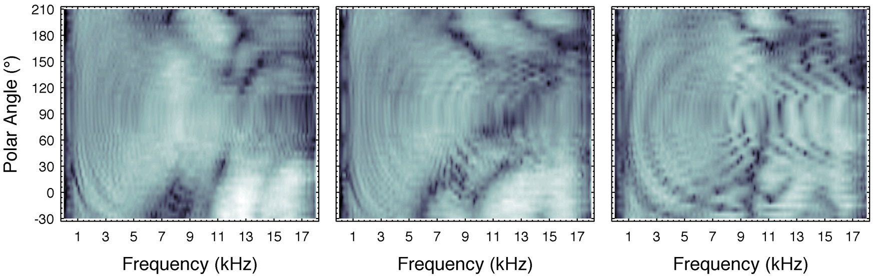


Fig. 12. DTFs of median SP, left ear. **Left**: NH12. **Center**: NH58. **Right**: NH33. *Brightness*: Spectral magnitude – for code see Fig. 1, right panel

Binaural recordings aim at creating a spatial impression when listening via headphones. They are usually created using either an artificial head or mounting microphones into the ear canal of a listener and, thus, implicitly use HRTFs. When listening to binaural recordings, the HRTFs of the listener do not necessarily correspond to those used in the recordings. HRTFs are, however, generally highly listener-specific and the relevant spectral features differ across listeners – see Fig. 12. Usually, SP localization performance degrades when listening to binaural signals created with non-individualized HRTFs [34]. The degree of the performance deterioration can be expected to depend on the similarity of the listener's DTFs with those actually applied. Here, the proposed model is used to estimate the localization performance for non-individualized binaural recordings. Figure 13 compares the performance when listening to individualized recordings with the average performance when listening to non-individualized recordings created from all other 16 listeners. It is evident that, on average, listening with other ears results in an increase of predicted localization errors.

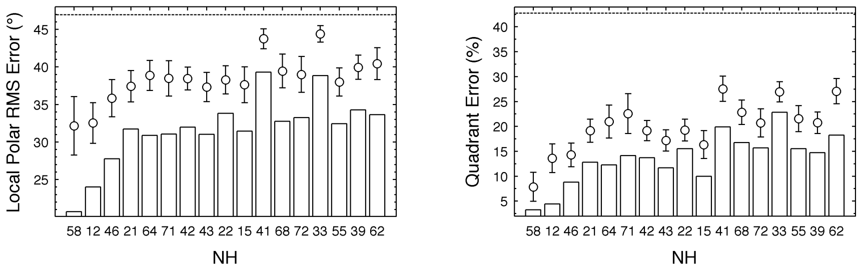


Fig. 13. Listeners' localization performance for non-individualized versus individualized DTFs. *Bars*: individualized DTFs. *Circles*: non-individualized DTFs averaged over 16 DTF sets. *Error bars*: ±1 standard deviation of the average. *Dashed line*: chance performance corresponding to guessing the direction of the sound

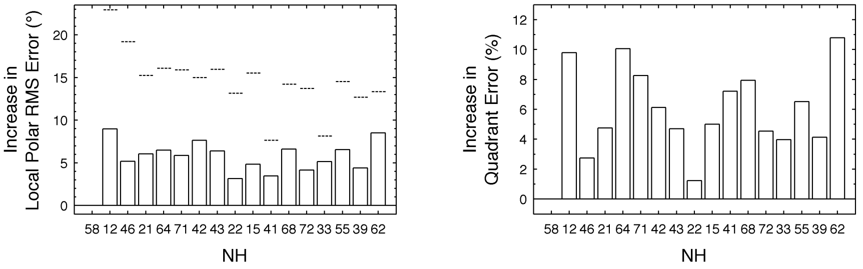


Fig. 14. *Bars*: increase in predicted localization errors when listening to the DTFs from NH58 with respect to the errors predicted when listening to individualized DTFs. *Dashed lines*: chance performance, not shown if too large

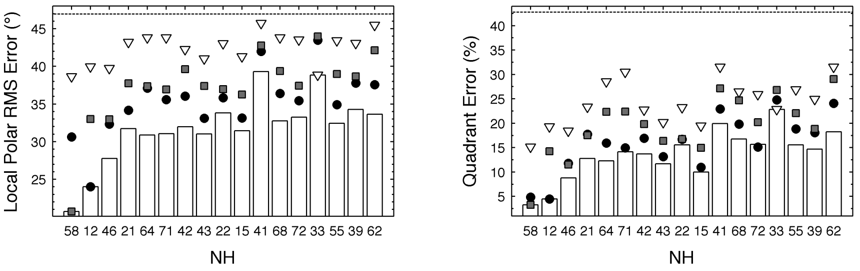


Fig. 15. *Bars*: localization performance of the pool listening to selected DTFs. *Circles*: DTFs from NH12. *Squares*: DTFs from NH58. *Triangles*: DTFs from NH33. *Dashed line*: chance performance

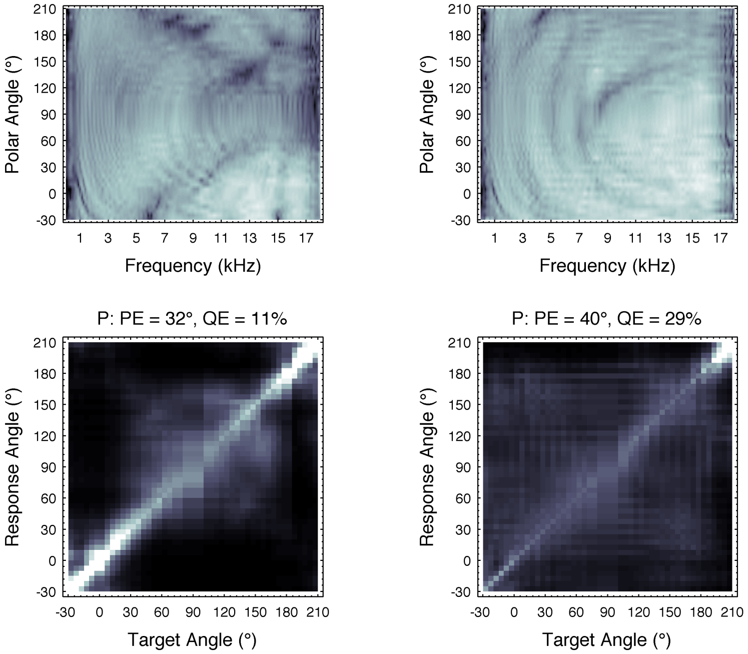
Thus, the question arises of how a pool of listeners would localize a binaural recording from a particular listener, for instance, NH58. Figure 14 shows the listener-specific *increase* in the predicted localization errors when listening to a binaural recording spatially encoded using the DTFs from NH58 with respect to the errors predicted for using individualized DTFs. Some of the listeners like NH22 show only little increase in errors, while others like NH12 show large increase.

Generally, one might assume that the different anatomical shapes of ears produce more or less distinct directional features. Thus, the quality of the HRTFs might vary, having effect on the ability to localize sounds in the SPs. Figure 15 shows the performance of the pool, using the DTFs from NH12, NH58 and NH33. The DTFs from these three listeners provided best, moderate and worst performance, respectively, predicted for the pool listening to binaural signals created with one of those DTF sets.

This analysis demonstrates how to evaluate across-listener compatibility of binaural recordings. Such an analysis can also be applied for other purposes like the evaluation of HRTFs of artificial heads for providing sufficient spatial cues for binaural recordings.

4.2 Assessing the Quality of Spatial Cues in Hearing-assist Devices

In the development of hearing-assist devices, the casing, its placement on the head and the placement of the microphone in the casing play an important role for the effective directional cues. The proposed SP localization model can be used to assess the quality of the directional cues picked up by the microphone in a given device. Figure 16 shows DTFs resulting from behind-the-ear, BTE, compared to in-the-ear, ITE, placement of the microphone for the same listener. The BTE microphone was placed above the pinna, pointing to the front, a position commonly used by the BTE processors in cochlear-implant systems. The bottom row of Fig. 16 shows the corresponding prediction matrices and the predicted localization performance, namely, PE and QE. For


P: PE = 32°, QE = 11%

P: PE = 40°, QE = 29%

Fig. 16. Impact of the microphone placement. **Top**: DTFs of median SP from NH10, left ear. **Bottom**: prediction matrices. **Left**: ITE microphone. **Right**: BTE microphone. All other conventions are as in Fig. 8

this particular listener, the model predicts that if NH10 were listening with the BTE DTFs, his/her QE and PE would increase from 12% to 30% and from 32° to 40°, respectively. This can be clearly related to the impact of degraded spatial cues. Note that in this analysis it was assumed that NH10 fully adapted to the particular HRTFs. This was realized by using the same set of DTFs for the targets and the template in the model.

The impact of using BTE DTFs was also modeled for the pool of listeners using the calibrated models. Two cases are considered, namely, *ad-hoc* listening where the listeners are confronted with the DTF set without any experience in using it, and trained listening where the listeners are fully adapted to the respective DTF set. Figure 17 shows the predictions for the pool. The BTE DTFs result in performances close to guessing and the ITE DTFs from the same listener substantially improve the performance. In trained listening, the performance for the ITE DTFs is at the level of the individualized DTFs, consistent with the potential of the plasticity of the spectral-to-spatial mapping [13]. The BTE DTFs, however, do not allow performance at the same level as the ITE DTFs, even when full adaptation is allowed for.

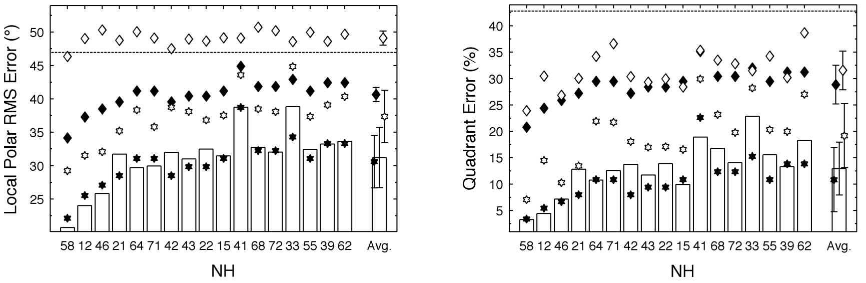


Fig. 17. Localization performance of the pool listening to different DTFs. *Bars*: individualized DTFs. *Open symbols*: *ad-hoc* listening. *Filled symbols*: trained listening. *Hexagrams*: ITE DTFs from NH10. *Diamonds*: BTE DTFs from NH10. *Avg.*: average performance over all listeners. *Error bars*: ±1 standard deviation. *Dashed line*: chance performance

This analysis shows a model-based method to optimize the microphone placement with respect to the salience of directional cues. Such an analysis might be advantageous in the development of future hearing-assist devices.

4.3 Phantom Sources in Surround-sound Systems

Sound synthesis systems for spatial audio have to deal with a limited number of loudspeakers surrounding the listener. In a system with a small number of loudspeakers, vector-base amplitude panning, VBAP [41], is commonly applied in order to create phantom sources perceived between the loudspeakers. In a surround setup, this method is also commonly used to position the phantom source along SPs, namely, to pan the source from the front to the back [11] or from the eye level to an elevated level [41]. In this section, the proposed model is applied to investigate the use of VBAP within SPs.

Amplitude panning along a sagittal plane

Now a VBAP setup with two loudspeakers is assumed – placed at the same distance, in the horizontal plane at the eye level, and in the same SP. Thus, the loudspeakers are in the front and in the back of the listener, corresponding to polar angles of 0° and 180°, respectively. While driving the loudspeakers with the same signal, the amplitude panning ratio can be varied from 0, front speaker only, to 1, rear speaker only, with the goal of panning the phantom source between the two loudspeakers.
Figure 18 shows the predicted listener-specific response probabilities in terms of the PMV as a function of the panning ratio for two loudspeakers placed at the lateral angle of 30°. The PMVs are shown for two individual listeners and also for the pool of listeners. The directional stability of phantom

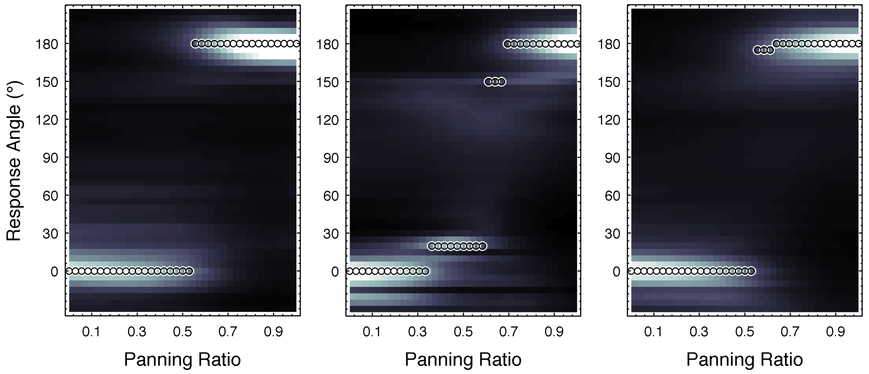


Fig. 18. Predicted response probabilities, PMVs, as a function of the amplitude panning ratio. **Left**: results for NH22. **Center**: results for NH64. **Right**: results for the pool of listeners. *Circle*: maximum of a PMV. Panning ratio of 0: Only front loudspeaker active. Panning ratio of 1: Only rear loudspeaker active. All other conventions are as in Fig. 5, right panel

sources varies across listeners. For NH22, the prediction of perceived location abruptly changes from front to back, being bimodal only around the ratio of 0.6. For NH64, the transition seems to be generally smoother, with a blur in the perceived sound direction. Note that for NH64 and a ratio of 0.5, the predicted direction is elevated even though the loudspeakers were placed in the horizontal plane. On average, the results for the pool predict an abrupt change in the perceived direction from front to back, with a blur indicating a listener-specific unstable representation of the phantom source for ratios between 0.5 and 0.7.

Effect of loudspeaker span

The unstable synthesis of phantom sources might be reduced by using a more adequate distance in the SP between the loudspeakers. Thus, it is shown how to investigate the polar span between two loudspeakers required to create a stable phantom source in the synthesis. To this end, a VBAP setup of two loudspeakers placed in the median SP, separated by a polar angle and driven with the panning ratio of 0.5, is used. Note that a span of 0° corresponds to a synthesis with a single loudspeaker and thus to the baseline condition. In the proposed SP localization model, the target angle describes the average of the polar angles of both loudspeakers, which, in VBAP, is thought to correspond to the direction of the phantom source. The model was run for all available target angles resulting in the prediction of the localization performance.

Figure 19 shows prediction matrices and predicted localization performance for NH12 and three different loudspeaker spans. Note the large increase of the errors from the span of 30°–60°, consistent with the results from [5].

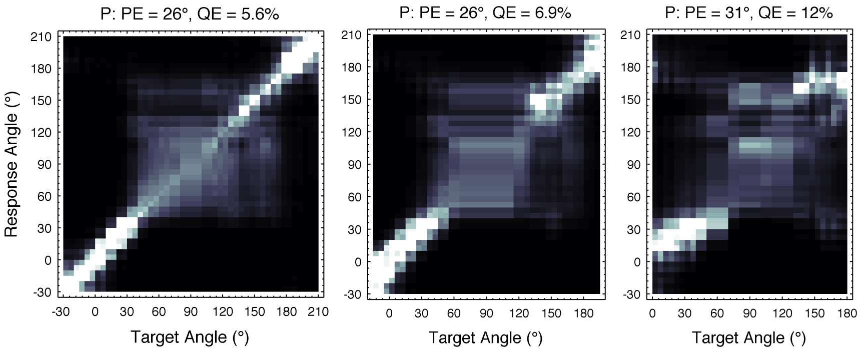


Fig. 19. Predictions for different loudspeaker spans and NH12. **Left**: span of 0°, single-loudspeaker synthesis, baseline condition. **Center**: span of 30°. **Right**: span of 60°. All other conventions are as in Fig. 8

Figure 20 shows the average increase in localization error predicted for the pool of listeners as a function of the span. The increase is shown relative to the listener-specific localization performance in the baseline condition. Note that not only the localization errors but also the variance across the listeners increase with increasing span.

This analysis shows how the model may help in choosing the adequate loudspeaker span when amplitude panning is applied to create phantom sources. Such an analysis can also be applied when more sophisticated sound-field reproduction approaches like Ambisonics or wave-field synthesis are involved.

Results for typical surround-sound setups

The most common standardized surround-sound setup is known as the 5.1 setup [18]. In this setup, all loudspeakers are placed in the horizontal plane

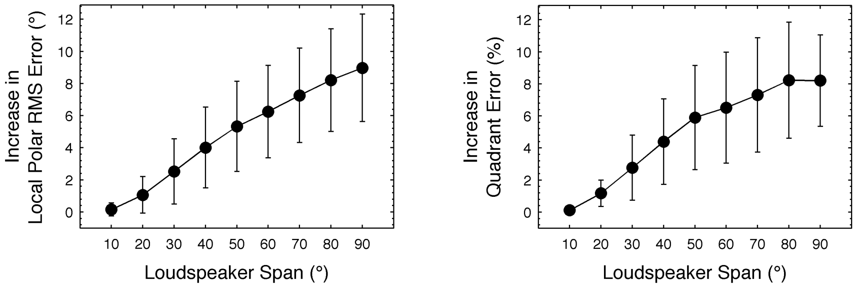


Fig. 20. Increase in localization errors as a function of the loudspeaker span. *Circles*: averages over all listeners from the pool. *Error bars*: ±1 standard deviation

at a constant distance around the listener. Recently, other schemes have been proposed to include elevated speakers in the synthesis systems. The 10.2 setup, known as *Audyssey DSX* [15] and the 9.1 setup, known as *Auro-3D* [48], consider two and four elevated loudspeakers, respectively. Figure 21 shows the positions of the loudspeakers in those three surround-sound setups. The model was applied to evaluate the localization performance when VBAP is used to pan a phantom source at the left hand side from front, L, to back, LS. While in the 5.1 setup only loudspeakers L and LS are available, in 10.2 and 9.1 the loudspeakers LH2 and LH1 & LSH, respectively, may also contribute even to create an elevated phantom source.

VBAP was applied between the closest two loudspeakers using the law of tangents [41]. For a desired polar angle of the phantom source, the panning ratio was $R = \frac{1}{2} - \frac{\tan(\delta)}{2\tan(0.5\beta)}$ with β denoting the loudspeaker span in polar dimension and δ denoting the difference between the desired polar angle and the polar center angle of the span. The contributing loudspeakers were not always in the same SP, thus, the lateral angle of the phantom source was considered for the choice of the SP in the modeling by applying the law of tangents on the lateral angles of the loudspeakers for the particular panning ratio, R.

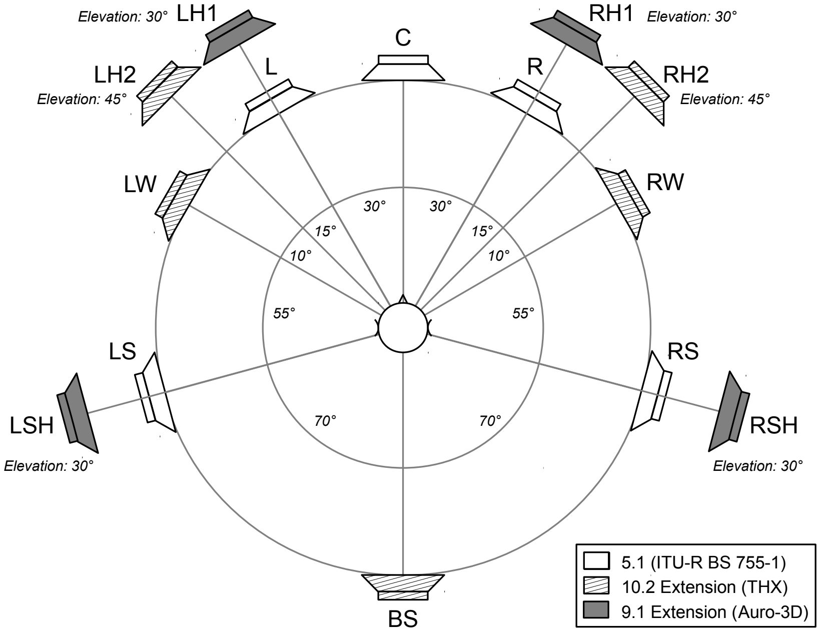


Fig. 21. Loudspeaker positions of three typical surround-sound systems. Drivers for the low-frequency effect, LFE, channels not shown

Figure 22 shows the predicted pool averages of the PMVs as a function of the desired polar angle of the phantom source. The improvements due to the additional elevated loudspeakers in the 10.2 and 9.1 setups are evident. Nevertheless, the predicted phantom sources are far from perfectly following the desired angle. Especially for the 9.1 setup, in the rear hemifield, the increase in the desired polar angle, namely, *decrease* in the elevation, resulted in a decrease in the predicted polar angle, namely, *increase* in the elevation.

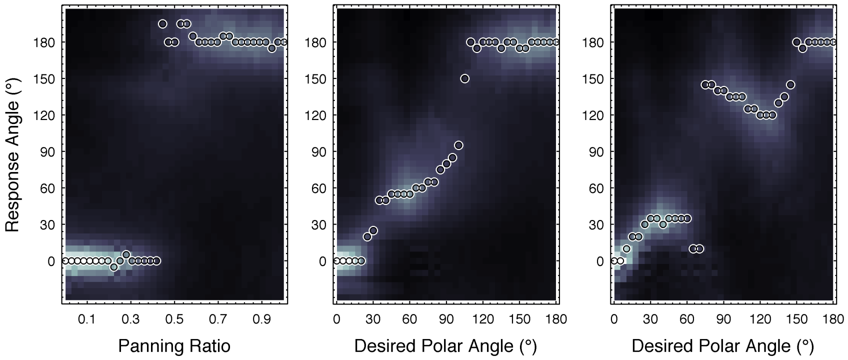


Fig. 22. Predictions for the surround setups in the VBAP configuration. **Left**: 5.1 setup, panning between the loudspeakers L and LS. **Center**: 10.2 setup, DSX, panning from L , polar angle of 0°, via LH2, 55°, to LS, 180°. **Right**: 9.1 setup, Auro-3D, panning from L, 0°, via LH1, 34°, and LSH, 121°, to LS, 180°. *Desired polar angle*: Continuous scale representing the VBAP across pair-wise contributing loudspeakers. All other conventions are as in Fig. 18

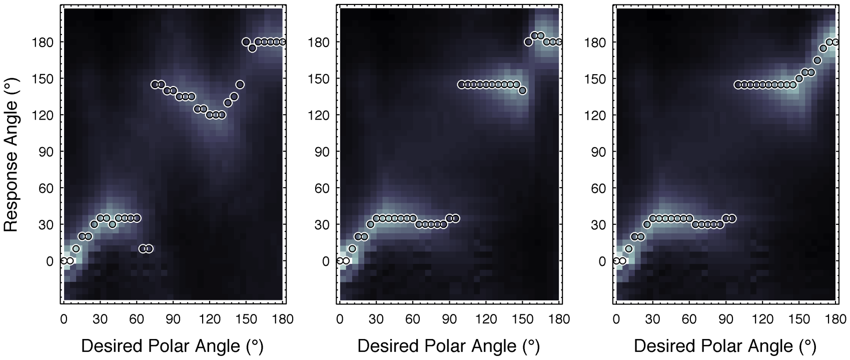


Fig. 23. Predictions for two modifications to the 9.1 setup, Auro 3D. **Left**: original setup, loudspeakers LS and LSH at azimuth of 110°. **Center**: LSH at azimuth of 140°. **Right**: LS and LSH at azimuth of 140°. All other conventions are as in Fig. 22

The proposed model seems to be well-suited for addressing such a problem. It is easy to show how modifications of the loudspeaker setup would affect the perceived angle of the phantom source. As an example, the positions of the elevated loudspeakers in the 9.1 setup were modified in two ways. First, the lateral distance between the loudspeakers, LH1 and LSH, was decreased by modifying the azimuth of LSH from 110° to 140°. Second, both loudspeakers, LSH and LS, were placed to the azimuth of 140°. Figure 23 shows the predictions for the modified setups. Compared to the original setup, the first modification clearly resolves the problem described above. The second modification – right panel – while only slightly limiting the lateral range, provides an even better representation of the phantom source along the SP.

5 Conclusions

Sound localization in SPs refers to the ability to estimate the sound-source elevation and to distinguish between front and back. The SP localization performance is usually measured in time-consuming experiments. In order to address this disadvantage, a model predicting SP localization performance of individual listeners has been proposed. Listener-specific calibration was performed for a pool of 17 listeners, and the calibrated models were evaluated using results from psychoacoustic localization experiments. The potential of the calibrated models was demonstrated for three applications, namely,

1. The evaluation of the spatial quality of binaural recordings
2. The assessment of the spatial quality of directional cues provided by the microphone placement in hearing-assist devices
3. The evaluation and improvement of the loudspeaker position in surround-sound systems

These applications are thought to be examples of situations where SP localization cues, namely, spectral cues, likely play a role. The model is, however, not limited to those applications and it hopefully will help in assessing spatial quality in other applications as well.

Acknowledgement

The authors would like to thank H. S. Colburn and two anonymous reviewers for constructive comments on earlier versions of this manuscript. This research was supported by the Austrian Science Fund, FWF, contract P 24124–N13.

References

[1] V. R. Algazi, C. Avendano, and R. O. Duda. Elevation localization and head-related transfer function analysis at low frequencies. *J Acoust Soc Am*, 109:1110–1122, 2001.

[2] F. Asano, Y. Suzuki, and T. Sone. Role of spectral cues in median plane localization. *J Acoust Soc Am*, 88:159–168, 1990.

[3] E. Blanco-Martin, F. J. Casajus-Quiros, J. J. Gomez-Alfageme, and L. I. Ortiz-Berenguer. Estimation of the direction of auditory events in the median plane. *Appl Acoust*, 71:1211–1216, 2010.

[4] J. Blauert. *Räumliches Hören (Spatial Hearing).* S. Hirzel Verlag S-tuttgart, 1974.

[5] P. Bremen, M. M. van Wanrooij, and A. J. van Opstal. Pinna cues determine orientation response modes to synchronous sounds in elevation. *J Neurosci*, 30:194–204, 2010.

[6] A. W. Bronkhorst. Localization of real and virtual sound sources. *J Acoust Soc Am*, 98:2542–2553, 1995.

[7] S. Carlile, P. Leong, and S. Hyams. The nature and distribution of errors in sound localization by human listeners. *Hear Res*, 114:179–196, 1997.

[8] T. Dau, D. Püschel, and A. Kohlrausch. A quantitative model of the "effective" signal processing in the auditory system. I. Model structure. *J Acoust Soc Am*, 99:3615–3622, 1996.

[9] B. R. Glasberg and B. C. J. Moore. Derivation of auditory filter shapes form notched-noise data. *Hear Res*, 47:103–138, 1990.

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

[11] J. Hilson, D. Gray, and M. DiCosimo. *Dolby Surround Mixing Manual*. Dolby Laboratories, Inc, San Francisco, CA, 2005. chapter 5 - Mixing techniques.

[12] M. Hofman and J. Van Opstal. Binaural weighting of pinna cues in human sound localization. *Exp Brain Res*, 148:458–470, 2003.

[13] P. M. Hofman and A. J. V. Opstal. Spectro-temporal factors in two-dimensional human sound localization. *J Acoust Soc Am*, 103:2634–2648, 1998.

[14] P. M. Hofman, J. G. A. van Riswick, and A. J. van Opstal. Relearning sound localization with new ears. *Nature Neurosci*, 1:417–421, 1998.

[15] T. Holman. *Surround Sound: Up and Running*. Focal Press, 2008.

[16] S. Hwang and Y. Park. Interpretations on pricipal components analysis of head-related impulse responses in the median plane. *J Acoust Soc Am*, 123:EL65–EL71, 2008.

[17] K. Iida, M. Itoh, A. Itagaki, and M. Morimoto. Median plane localization using a parametric model of the head-related transfer function based on spectral cues. *Appl Acoust*, 68:835–850, 2007.

[18] Int Telecommunication Union, Geneva, Switzerland. *Multichannel stereophonic sound system with and without accompanying picture*, 2012. Recommendation ITU-R BS.775-3.

[19] C. Jin, M. Schenkel, and S. Carlile. Neural system identification model of human sound localization. *J Acoust Soc Am*, 108:1215–1235, 2000.

[20] B. F. Katz. Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation. *J Acoust Soc Am*, 110:2440–2448, 2001.

[21] W. Kreuzer, P. Majdak, and Z. Chen. Fast multipole boundary element method to calculate head-related transfer functions for a wide frequency range. *J Acoust Soc Am*, 126:1280–1290, 2009.

[22] A. Kulkarni and H. S. Colburn. Role of spectral detail in sound-source localization. *Nature*, 396:747–749, 1998.

[23] A. Kulkarni and H. S. Colburn. Infinite-impulse-response models of the head-related transfer function. *J Acoust Soc Am*, 115:1714–1728, 2004.

[24] E. H. A. Langendijk and A. W. Bronkhorst. Contribution of spectral cues to human sound localization. *J Acoust Soc Am*, 112:1583–1596, 2002.

[25] E. A. Lopez-Poveda and R. Meddis. A human nonlinear cochlear filterbank. *J Acoust Soc Am*, 110:3107–3118, 2001.

[26] F. R. S. Lord Rayleigh. On our perception of sound direction. *Philos Mag*, 13:214–232, 1907.

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

[28] E. A. Macpherson and J. C. Middlebrooks. Vertical-plane sound localization probed with ripple-spectrum noise. *J Acoust Soc Am*, 114:430–445, 2003.

[29] E. A. Macpherson and A. T. Sabin. Binaural weighting of monaural spectral cues for sound localization. *J Acoust Soc Am*, 121:3677–3688, 2007.

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

[31] J. C. Makous and J. C. Middlebrooks. Two-dimensional sound localization by human listeners. *J Acoust Soc Am*, 87:2188–2200, 1990.

[32] M. I. Mandel, R. J. Weiss, and D. P. W. Ellis. Model-based expectation-maximization source separation and localization. *IEEE Trans Audio Speech Proc*, 18:382–394, 2010.

[33] T. May, S. van de Par, and A. Kohlrausch. A probabilistic model for robust localization based on a binaural auditory front-end. *IEEE Trans Audio Speech Lang Proc*, 19:1–13, 2011.

[34] J. C. Middlebrooks. Individual differences in external-ear transfer functions reduced by scaling in frequency. *J Acoust Soc Am*, 106:1480–1492, 1999.

[35] J. C. Middlebrooks. Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. *J Acoust Soc Am*, 106:1493–1510, 1999.

[36] J. C. Middlebrooks and D. M. Green. Sound localization by human listeners. *Annu Rev Psychol*, 42:135–159, 1991.

[37] J. C. Middlebrooks and D. M. Green. Observations on a principal components analysis of head-related transfer functions. *J Acoust Soc Am*, 92:597–599, 1992.

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

[39] M. Morimoto. The contribution of two ears to the perception of vertical angle in sagittal planes. *J Acoust Soc Am*, 109:1596–1603, 2001.

[40] R. Patterson, I. Nimmo-Smith, J. Holdsworth, and P. Rice. *An efficient auditory filterbank based on the gammatone function*. APU, Cambridge, UK, 1988.

[41] V. Pulkki. Virtual sound source positioning using vector base amplitude panning. *J Audio Eng Soc*, 45:456–466, 1997.

[42] B. Rakerd, W. M. Hartmann, and T. L. McCaskey. Identification and localization of sound sources in the median sagittal plane. *J Acoust Soc Am*, 106:2812–2820, 1999.

[43] C. L. Searle and I. Aleksandrovsky. Binaural pinna disparity: Another auditory localization cue. *J Acoust Soc Am*, 57:448–455, 1975.

[44] M. A. Senova, K. I. McAnally, and R. L. Martin. Localization of virtual sound as a function of head-related impulse response duration. *J Audio Eng Soc*, 50:57–66, 2002.

[45] E. A. Shaw. Transformation of sound pressure level from the free field to the eardrum in the horizontal plane. *J Acoust Soc Am*, 56:1848–1861, 1974.

[46] R. H. Y. So, B. Ngan, A. Horner, J. Braasch, J. Blauert, and K. L. Leung. Toward orthogonal non-individualised head-related transfer functions for forward and backward directional sound: cluster analysis and an experimental study. *Ergonomics*, 53:767–781, 2010.

[47] P. Søndergaard and P. Majdak. The auditory modeling toolbox. In J. Blauert, editor, *The technology of binaural listening*, chapter 2. Springer, Berlin–Heidelberg–New York NY, 2013.

[48] G. Theile and H. Wittek. Principles in surround recordings with height. In *Proceedings of the 130th AES Convention*, page Convention Paper 8403, London, UK, 2011.

[49] M. M. van Wanrooij and A. J. van Opstal. Relearning sound localization with a new ear. *J Neurosci*, 25:5413–5424, 2005.

[50] J. Vliegen and A. J. V. Opstal. The influence of duration and level on human sound localization. *J Acoust Soc Am*, 115:1705–1703, 2004.

[51] T. Walder. Schallquellenlokalisation mittels Frequenzbereich-Kompression der Außenohrübertragungsfunktionen (sound-source localization through warped head-related transfer functions). Master's thesis, University of Music and Performing Arts, Graz, Austria, 2010.

[52] A. J. Watkins. Psychoacoustical aspects of synthesized vertical locale cues. *J Acoust Soc Am*, 63:1152–1165, 1978.

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

[54] F. L. Wightman and D. J. Kistler. Monaural sound localization revisited. *J Acoust Soc Am*, 101:1050–1063, 1997.

[55] P. Zakarauskas and M. S. Cynader. A computational theory of spectral cue localization. *J Acoust Soc Am*, 94:1323–1331, 1993.

[56] P. X. Zhang and W. M. Hartmann. On the ability of human listeners to distinguish between front and back. *Hear Res*, 260:30–46, 2010.

[57] M. S. A. Zilany and I. C. Bruce. Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery. *J Acoust Soc Am*, 120:1446–1466, 2006.

Index