Audio Engineering Society

Convention Paper 5406

Presented at the 110th Convention
2001 May 12–15 Amsterdam, The Netherlands

The perceptual (ir)relevance of HRTF magnitude and phase spectra

Jeroen Breebaart[1] and Armin Kohlrausch[1,2]

[1]IPO Center for User-System Interaction and [2]Philips Research Laboratories Eindhoven

PO Box 513, NL-5600 MB Eindhoven, The Netherlands

ABSTRACT

This paper discusses the perceptual consequences of smoothing of anechoic HRTF phase and magnitude spectra. The smoothing process is based on a binaural perception model, in which interaural cues in the auditory system are rendered at a limited spectral resolution. This limited resolution is the result of the filterbank present in the peripheral auditory system (i.e., the cochlea). Listening tests with single and multiple virtual sound sources revealed that both the phase and magnitude spectra of HRTFs can be smoothed with a gammatone filter which equals estimates of the spectral resolution of the cochlea without audible artifacts. The amount of smoothing was then increased by decreasing the order of the gammatone filters. If the filter order is reduced by a factor 3, subjects indicate spectral and positional changes in the virtual sound sources. The binaural detection model developed by Breebaart, van de Par and Kohlrausch was used to predict the audibility of the smoothing process. A comparison between model predictions and experimental data showed that the threshold at which subjects start to hear smoothing artifacts can be predicted accurately. Moreover, a high correlation exists between the model output and the amount of stimulus degradation reported by subjects.

INTRODUCTION

Two important features of the waveforms arriving at both ears that determine the lateral location of a sound source are the interaural intensity differences (IIDs) and the interaural time differences (ITDs). Stimuli with specified values of the ITD and IID can be presented over headphones, resulting in a lateralization of the sound source which depends on the magnitude of the ITD or IID [1, 2, 3]. The usual result of these experiments is that the source images are located inside the head, somewhere between the left and right ear. The reason for the fact that these stimuli are not externalized is that the single frequency-independent IID or ITD is a poor representation of the acoustical signals in the real world. The waveforms of sound sources in the real world are filtered by the pinna, head and torso of the listener, resulting in an intricate frequency dependence of the ITD and IID [4]. The filtering can be described by the head-related transfer function (HRTF), which describes the position-dependent change in the phase and magnitude spectra of a sound source. One of the major difficulties in using HRTFs is that these filters are both position and subject dependent [4]. Usually HRTFs are measured as a function of both elevation and azimuth, but there is evidence that spatial cues also depend on the distance of a sound source [5, 6]. If invidivualized HRTFs are used, subjects are not able to discriminate between real and virtual sound sources presented over headphones [7, 8, 9]. If nonindividualized HRTFs are used, however, subjects report poor elevation accuracy and front-back confusions [10, 11]. Some attempts have been made to increase localization performance with nonindividualized HRTFs by emphasizing the pinna effects [12] or the interaural differences [13]. Because of the large amount of data present in individual HRTF sets that is normally required to generate externalized virtual sound sources, researchers have tried to reduce the information in several ways. For example, attempts have been made to only measure HRTF sets for a limited range of source positions and to interpolate HRTF impulse responses for positions in between [14]. Other studies described HRTFs in terms of principal components by deriving a small set of basis spectra with individual, position-dependent weights [15, 16]. Although this method is valid in physical terms, there is a risk that the basis functions that are very important in terms of the least-squares error of the fit are not so relevant in terms of human auditory perception. An other approach consisted of determining the role of spectral and interaural phase cues present in the HRTFs. Wightman and Kistler [17] showed that low-frequency interaural time differences dominate in sound localization, while if the low frequencies are removed from the stimuli, the apparant direction is determined primarily by interaural intensity differences and pinna cues. Hartmann and Wittenberg [8] and Kulkarni *et al.* [18] showed that the frequency-dependent ITD of anechoic HRTFs can be simplified by a frequency-independent delay without perceptual consequences. But also more psychoacoustically-motivated methods of HRTF reduction have been suggested. For example, [19] discussed three methods to reduce HRTF information. The first entailed smoothing of the HRTF magnitude spectra by a rectangular smoothing filter with a bandwidth equal to the equivalent rectangular bandwidth (ERB) [20]. The second embodied weighting of the errors in an HRTF approximation with the inverse of the ERB scale as weighting function. The third method used frequency warping to account for the non-uniform frequency resolution of the auditory system.

From many of the studies described above, it can be concluded that although a single IID and/or ITD does not result in an externalized image, the complex magnitude and phase spectra which are present in HRTFs can be simplified to some extent without deteriorating the externalization. In this paper, we will investigate the relaxation of anechoic HRTF accuracy based on smoothing of the phase and magnitude spectra. The method of smoothing is derived from a binaural detection model described by Breebaart et al. [21, 22, 23]. Although smoothing has been proposed before, our efforts differ in two aspects from other studies:

- the method of smoothing aims at a minimized *perceptual degradation* of the sound image. This is achieved by minimizing the changes in the internal representation of a binaural detection model rather than minimizing a norm of the HRTF impulse response errors.
- because we are interested in a generalized theory of describing HRTF data, we do not discuss any filter structure that may achieve the desired smoothing, because we do not want to be limited by implementation issues.

HRTF SMOOTHING

It is generally accepted that the auditory system splits the incoming waveforms in several band-limited signals. The bandwidth of these filters depends on the center frequency [20] and can be seen as a limit of the spectral accuracy of (binaural) processing. We hypothesize that *the HRTF phase and magnitude spectra do not need a higher resolution than the spectral resolution of the filterbank in the peripheral auditory system.* This hypothesis is supported by the binaural detection model presented by [21, 22, 23]. This model consists of 3 consecutive stages which are described in more detail in section 3. The first stage comprises a peripheral preprocessing model, which among other things simulates the spectral filtering of the cochlea by applying a gammatone filterbank. It has been shown that with the correct choice of its bandwidth parameter, the spectral resolution of a 4th order gammatone filter closely matches the spectral resolution of the human cochlea [24, 25]. The consecutive stages explore monaural properties (such as spectral content) and binaural properties (ITDs and IIDs) *after* the gammatone filterbank. Hence this filterbank limits the spectral resolution for the binaural auditory system. We will therefore use the same gammatone filter to explore the perceptual consequences of HRTF phase and magnitude smoothing.

HRTF magnitude smoothing

The gammatone filter has an impulse response for $t \geq 0$ given by [25]

$$h(t) = t^{n-1} e^{-2\pi bt} \cos(2\pi f_c t + \phi), \tag{1}$$

where n denotes the order of the filter, b determines the bandwidth, f_c is the center frequency of the filter and ϕ the initial phase. The resulting transfer function $H(f, f_c)$ for ϕ=0 can be approximated by [25]

$$H(f, f_c) = \left(\frac{1}{1 + j(f - f_c)/b}\right)^n, \tag{2}$$

and the 3-dB bandwidth B_{3dB} is given by

$$B_{3dB} = 2b\sqrt{2^{1/n} - 1}. \tag{3}$$

[1] The equivalent rectangular bandwidth of a bandpass gammatone filter is always larger than the 3-dB bandwidth. For a filter order of 3, the ERB is about 13% larger than the 3-dB bandwidth. Hence the ERB of our smoothing filters is a bit larger than the ERB estimate of the auditory filters, even at a filter order of 3. Our method of smoothing encompasses decrements of the filter order and hence increments of the ERB of the smoothing filter. However, the 3-dB bandwidth is always kept constant and was equal to the ERB-estimate of the human auditory filters.

This 3-dB bandwidth was set to the equivalent rectangular bandwidth (ERB) estimate [1] of the human auditory filters given by [20], resulting in

$$b(f_c) = \frac{24.7(0.00437 f_c + 1)}{2\sqrt{2^{1/n} - 1}}. \qquad (4)$$

Then the smoothed magnitude $|Y(f_c)|$ of HRTF $X(f)$ is given by

$$|Y(f_c)| = \sqrt{\frac{\int_0^\infty |X(f)|^2 |H(f, f_c)|^2 df}{\int_0^\infty |H(f, f_c)|^2 df}}. \qquad (5)$$

The numerator denotes the product of the original magnitude spectrum $|X(f)|$ with the smoothing function $|H(f, f_c)|$, while the denominator compensates for a spectral tilt resulting from the changing bandwidth with center frequency. The explicit form given in Eq. 5 has some important advantages:

1. because a gammatone filter is used, the amount of crosstalk between adjacent filters is closer to that in the human auditory system than for a rectangular smoothing window, as suggested by [19].
2. the binaural detection model described in [21] uses the *energy* of the difference signal of the waveforms arriving at the two ears after the peripheral filterbank as a decision variable to detect interaural differences. It can be shown that smoothing of the power spectrum of the HRTF magnitude spectra instead of smoothing the linear magnitude spectra gives a better fit in terms of the binaural model.

The parameter that was used to change the effect of the smoothing process is the order of the filter n. If n is decreased, the skirts of the smoothing filter become less steep while keeping the 3-dB bandwidth constant. Hence processing an HRTF with a *lower* filter order leads to *more* smoothing. This is demonstrated in Fig. 1. The left panel shows the magnitude of the smoothing filter $|H(f, 500)|$ at a center frequency of 500 Hz for different values of the filter order, ranging from 0.5 to 3. The right panel corresponds to a center frequency of 2000 Hz.

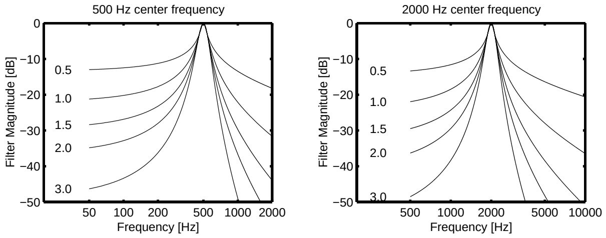

Fig. 1: Examples of the smoothing functions $|H(f, f_c)|$ as a function of frequency for two center frequencies (500 Hz for the left panel and 2000 Hz for the right panel) and different values for the filter order, ranging from 0.5 to 3.

The result of the smoothing process upon the magnitude of the HRTF can be observed in Fig. 2. Here, the magnitude of the HRTFs for a sound source at an elevation of 0° and an azimuth of 30° is shown. The left panel corresponds to the ipsilateral ear, the right panel to the contralateral ear. The solid line denotes the original (i.e., unprocessed) HRTF. The dashed line is a smoothed magnitude spectrum for n=1. Clearly, sharp peaks and dips disappear through the smoothing operation.

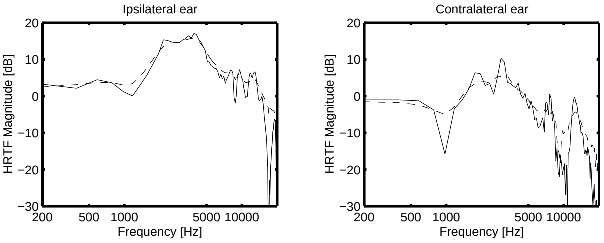

Fig. 2: Unprocessed (solid lines) and smoothed (dashed lines) HRTF magnitude spectra for a sound source at 0° elevation and 30° azimuth. The order of the smoothing filter n equals 1. The left panel shows the spectra for the ipsilateral ear, the right panel for the contraleral ear.

HRTF phase smoothing

The phase spectra of HRTF pairs usually consist of interaural phase differences that are not linear with frequency (i.e., an overall delay of the contralateral ear). From an engineering point of view, however, it would be very attractive if linear phase or minimum phase filters could be used for the generation of virtual sound sources due to their lower complexity. We therefore decided not to smooth the phase spectra themselves, but to use a smoothing that eventually (for low n) results in linear phase HRTFs, and hence in a frequency-independent ITD. This time smoothing is obtained by first dividing the phase spectrum by $2\pi f$. Given a HRTF $X(f_c)$, the smoothed phase spectrum of $Y(f_c)$ is given by

$$\arg\{Y(f_c)\} = 2\pi f_c \frac{\int_0^\infty \frac{\arg\{X(f)\}}{2\pi f} |H(f, f_c)| df}{\int_0^\infty |H(f)| df} \qquad (6)$$

An example of the resulting ITD is given in Fig. 3. The left panel shows the ITD for a sound source at 30° azimuth, the right panel for 120° azimuth. The solid line denotes the original (i.e., unprocessed) ITD, the dashed line the smoothed ITD for n=1. The following sections describe psycho-acoustic listening tests to reveal the audibility of the smoothing operation described above.

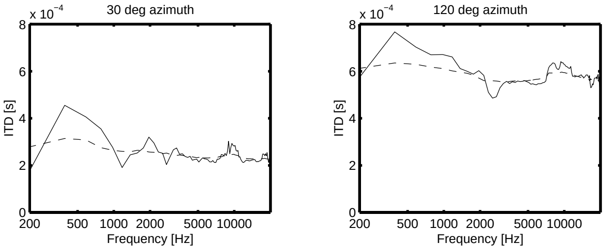

Fig. 3: ITD as a function of frequency for an unprocessed HRTF pair (solid line) and a smoothed HRTF pair (dashed line). The left panel corresponds to a sound source at 30° azimuth, the right panel to 120° azimuth.

PERCEPTUAL EVALUATION

Stimuli

Six different wideband CD-quality stereo musical fragments were used to create virtual loudspeakers. These fragments had a duration of about 2.5 seconds. The fragments cover a wide variety of musical genres and sonic attributes: some fragments only contained one instrument (voice or piano), while other fragments created various phantom sources when played through a stereo sound set (orchestra and rock band). The normalized cross-correlation between the left and right channels ranged from 0.05 to 0.94.

Anechoic HRTFs were taken from the AUDIS CDROM [26] for one person only. Each subject in our experiments listened to the same (non-individual) HRTF set. The six audio fragments were filtered with original (unprocessed) HRTFs and smoothed HRTFs. Smoothing was applied for HRTF phase only, HRTF magnitude only, or both. The smoothing order n ranged from 3 (little smoothing) to 0.1 (very close to linear phase or a flat spectrum). Two distinct cases were investigated. The first comprised only one virtual sound source at an elevation of 0° and an azimuth of 0°, 30° or 120°. The second condition comprised 2 virtual loudspeakers, at ± 30° or at ± 120° azimuth. In this condition, the signals consisted of the left channel and the right channel of the original stereo fragment, respectively. All stimuli were presented to the subjects over headphones (Beyerdynamic DT990) in an isolated listening booth at a level between 70 and 80 dB SPL (depending on the fragment).

Procedure

Three trained subjects participated in the experiments. Each trial consisted of the presentation of a fragment filtered through unprocessed HRTFs, followed by 300 ms silence and the same fragment filtered through processed HRTFs. All combinations of fragment, smoothing parameters and number of virtual loudspeakers were presented once in random order. Subjects had to judge the difference between the two fragments by giving one out of three possible answers: no audible differences, small audible differences (subjects could hear some subtle changes) or large audible differences (subjects could clearly hear the effect of the smoothing operation).

Results

The subjects' responses averaged over fragment and subject are shown in Fig. 4. The left panels show data for phase smoothing only, the middle panels for magnitude smoothing only. The right panels correspond to combined magnitude and phase smoothing. The upper panels correspond to a single virtual loudspeaker, the lower panels correspond to two simultaneous virtual loudspeakers. If the data for a single virtual loudspeaker are considered, filter orders of 1 and higher do not result in audible aftifacts for phase smoothing or magnitude smoothing, independent of the position of the virtual loudspeaker. Below filter order 1, subjects indicate very small audible differences for phase smoothing, small audible artifacts for magnitude smoothing and clear effects of the combined smoothing.

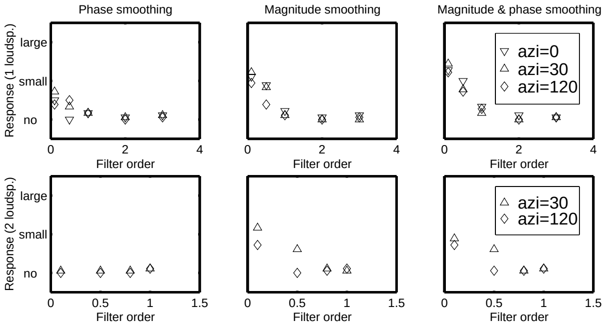

Fig. 4: Subjects' average responses as a function of the smoothing filter order for a single virtual loudspeaker (upper panels) and for two simultaneous virtual loudspeakers (lower panels). The left panels show data for phase smoothing only, the middle panels for magnitude smoothing only, and the right panels for combined magnitude and phase smoothing. The downward triangles correspond to virtual loudspeakers at an azimuth of 0°, and the upward triangles and diamonds to 30 and 120° azimuth, respectively.

The lower panels show the data for 2 simultaneous loudspeakers, at ± 30 or ± 120° azimuth (triangles and diamonds, respectively). Clearly, in these conditions, phase smoothing does not result in audible changes (see lower-left panel). Magnitude smoothing results in audible artifacts if the filter order is at or below 0.5 for sources at ± 30° azimuth. For sources at ± 120° azimuth audible artefacts are only reported for filter orders below 0.5. The combined magnitude and phase smoothing results in very similar data as for magnitude smoothing alone. Overall, it can be concluded that for two virtual loudspeakers more smoothing can be allowed to result in similar perceptual stimulus degradation as for a single virtual loudspeaker.

MODEL PREDICTIONS

The binaural detection model of [21, 22, 23] was used to simulate the perceptual consequences of HRTF smoothing. The reader is referred to the references above for the complete details of the model. Only the general model setup will be discussed here (see Fig. 5).

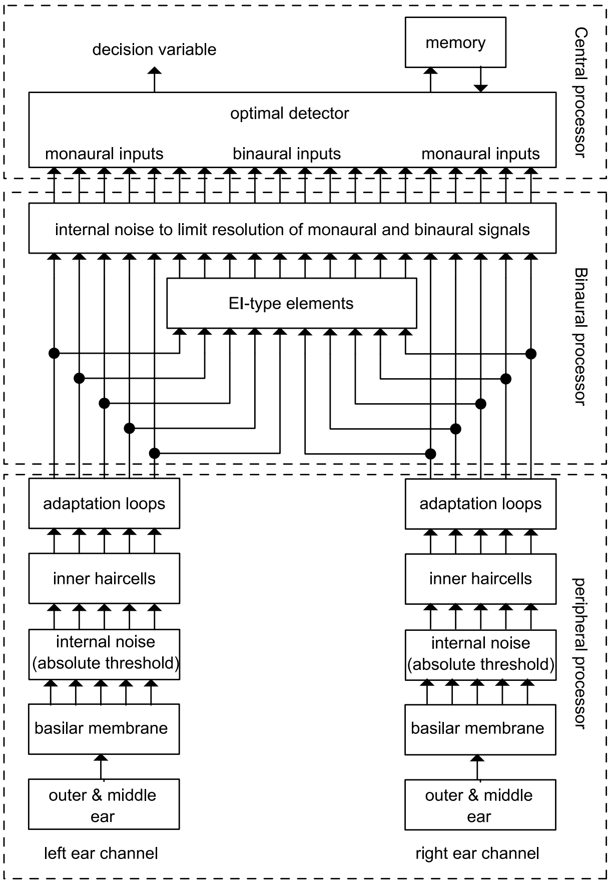

Fig. 5: General model setup. The model consists of three stages: a peripheral preprocessor, a binaural processor and a central processor.

The model consists of three stages. The first stage comprises a peripheral preprocessing stage. The three most promiment features of this stage are:

- Filtering of the gammatone filterbank. The filterbank present in the peripheral processing stage determines the spectral resolution of the model, in line with the ERB estimates published by [20].
- Inner haircell model. This stage consists of a half-wave rectifier followed by a fifth-order lowpass filter with a cutoff frequency (-3dB) of 770 Hz. Hence below 770 Hz, both the ITDs and IIDs are preserved at the output of this stage. However, above 2 kHz, the output approximates the envelope of the incoming signals and hence only IIDs and ITDs present in the envelope are preserved. For frequencies in between, the ITD in the fine structure waveforms is gradually lost.
- Adaptation loops. The chain of adaptation loops in the peripheral processor has an almost logarithmic input-output characteristic in steady state and is a non-linear device. It has been shown frequently that for both monaural and binaural detection of signals added to a wideband masker with a variable level, the threshold *signal-to-masker* ratio is approximately constant, as long as the masker level is well above the absolute threshold [cf. 27, 28]. If it is assumed that a certain constant *change* at the output of the adaptation loops is needed to detect a signal, the signal must be equal to a certain *fraction* of the masker level due to the logarithmic compression. Hence the signal-to-masker ratio will be approximately constant at threshold.

The second stage comprises binaural interaction based on an Equalization-Cancellation (EC) mechanism [29, 30]. For each frequency channel, the squared difference between the waveforms from the left and right peripheral preprocessors is computed *as a function of an internal interaural delay τ (in seconds) and an internal interaural level adjustment α (in dB)* by so-called EI-type (Excitation-Inhibition) elements. These squared-difference signals are then fed through a temporal averager to account for a limited binaural temporal resolution. This process is performed for all center frequencies of the gammatone filterbank, resulting in a set of 3-dimensional activity patterns which usually have a minimum somewhere in these patterns. At this minimum, the externally presented IID and ITD at that frequency are compensated optimally by the internal delay and level adjustments. Hence the position of the minimum depends on the interaural time and intensity difference that was present in the stimulus at that center frequency.

The third stage, the central processor, receives both binaural (from the binaural processor) as well as monaural (directly from the adaptation loops) information. These inputs are all corrupted by additive internal noise to limit their resolution. The task of this stage is to compute an overall difference measure between two different internal representations.

The model predictions were obtained in the following way. First, an audio fragment which was filtered by an unprocessed HRTF set was fed through the model. This stimulus resulted in a certain internal representation. Such an internal representation can be divided into a monaural component and a binaural component, shown by the direct connections from peripheral preprocessor to the central processor and the outputs of the binaural processor to the central processor, respectively. These channels represent monaural cues, such as timbre or overall power changes. On the other hand, the outputs of the binaural processor supply binaural properties of the presented waveforms, such as the IID or ITD present in the stimulus. An example of the computed binaural activity pattern generated by the binaural processor is shown in the left panel of Fig. 6 as a function of the internal ITD and IID. The picture was generated for a virtual loudspeaker radiating white noise at 30° azimuth and 0° elevation. The center frequency of the model was set to 250 Hz. There is a clear minimum at a certain small internal delay, which just compensates for the ITD present in the stimulus due to HRTF filtering.

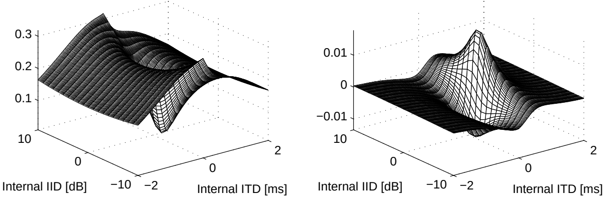

Fig. 6: Example of the effect of the combined phase and magnitude smoothing on the internal representation of the model. The left panel shows the internal activity pattern at 250-Hz center frequency as a function of the internal delay and internal intensity difference for an unprocessed HRTF pair corresponding to 30° azimuth. The right panel shows the *change* in the pattern due to HRTF smoothing.

If the stimuli are filtered by smoothed HRTFs and subsequently fed through the model, the internal representation of that stimulus will be slightly different from the internal representation of stimuli filtered by the original HRTFs. This change is the cue for detection by the model. This cue can be purely monaural, for example the smoothing of peaks and dips in the magnitude spectrum, but may also be binaural, such as changes in the interaural phase spectrum. The change in activity between the two internal representations was computed for each frequency and as a function of time, and both for monaural and binaural channels. As an example, the right panel shows the *change* in the binaural activity pattern due to combined phase and magnitude HRTF smoothing with order n=1, again at a center frequency of 250 Hz.

To result in a single measure of distance or difference between two internal representations, the changes in the internal representations (for both monaural and binaural cues at all center frequencies) are combined into one difference measure. This difference measure is obtained as follows. For each combination of sound source position and audio fragment, the internal representation of the corresponding stimulus was computed. This resulted in a time-varying model activity for each frequency channel and α,τ combination. However, only two monaural (from the left and right ears) and one binaural output (i.e., one α,τ combination) per frequency channel was used in the detection process. The α,τ values were obtained by computing the time-averaged binaural output per frequency channel and selecting α and τ that corresponded to the minimum average activity. The difference in internal representations between stimuli filtered by smoothed and unprocessed HRTFs was subsequently computed for each filter. To obtain a single time-varying distance measure, these differences were combined across frequency channels according to an optimal criterion [see 21, for details]. To account for the (limited) temporal integration ability of human listeners, this output was smoothed by a double-exponential averaging window with an equivalent rectangular duration of 300 ms. The maximum of this smoothed output was used as the overall distance measure. Because all internal channels are corrupted by internal noise, the overall distance measure will also be corrupted by internal errors. We therefore use the detectability index d', defined as the mean value of the difference (i.e., without noise), devided by the standard deviation of the noise on the decision variable. Thus, d' serves as a measure of detectability of the change in the internal representation due to HRTF smoothing. A low value of d' ($\leq$1) denotes inaudible changes or changes near threshold, while large values of d' >1 correspond to clearly audible artifacts. Values for d' were computed for each stimulus and virtual source position. The results are given in Fig. 7. The format is the same as in Fig. 4, except for the fact that the subject responses are replaced by the output of the model.

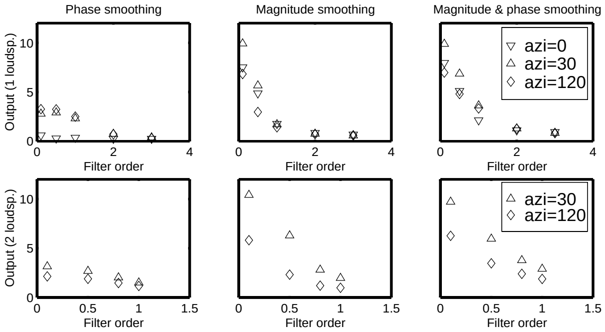

Fig. 7: Model output as a function of the smoothing filter order. The format is the same as in Fig. 4.

As expected, the model output increases monotonically with a decrease in filter order. Similar to the results obtained with human listeners, phase smoothing results in less audible artifacts than magnitude smoothing. Furthermore, more smoothing is necessary to obtain a similar model output for two virtual loudspeakers than for one virtual loudspeaker. To make a more quantitative comparison between model output and subject responses, we compared the responses of the listeners with the model output for corresponding stimuli. The results of this comparison are shown in Fig. 8. The abscissa gives the model output (d'), the ordinate the corresponding response of the subjects. The values for d' were averaged across audio fragments and plotted for each parameter combination (i.e., smoothing order, number of virtual loudspeakers and position of the virtual loudspeakers). The left panel corresponds to a single virtual loudspeaker, the right panel for 2 virtual loudspeakers. The solid lines are a linear fit to the data. Several remarks can be made with respect to Fig. 8.

- A high correlation exists between model output and subjects' response, both for single and multiple virtual sources.
- Predictions for phase smoothing, magnitude smoothing and the combined processing results in a similar relationship between model output and subjects' response. This relation can be successfully described by a linear fit. This indicates that the model can predict the perceptual consequences of different types of smoothing and transform differences in stimuli into one single difference measure.

- For single and multiple sources, the linear relation is not exactly equal, but very similar. In Fig. 3, it has been shown that subjects tolerate more smoothing for multiple sources than for a single virtual source. A single relation for the model was demonstrated in Fig. 7. Thus, the model correctly predicts that more smoothing is allowed for multiple sources compared to a single source to result in similar stimulus degradation.
- The values for d' which correspond to small audible artefacts amount about 8. This is significantly higher than the d' value of about +1 which is used to define the threshold of detectability in critical listening tests [see 22, 23]. This difference may be the result of the experimental procedure that was used in our experiments. Obviously, the procedure in our experiments does not yield maximum detection performance. Pilot studies revealed that experiments with noise instead of musical fragments combined with many repetitions of the same stimulus resulted in somewhat higher reported artefacts. Furthermore, subjects were asked to rate the audibility of the artefacts and were not instructed to optimally detect the presence of artefacts. A third rationale for high d' values may be related to the fact that the subjects did not listen through their own ears, i.e., nonindividual HRTFs were used to generate virtual speakers. The use of individualized HRTFs may result in larger reported smoothing artefacts. A fourth reason for high d' values may be related to the method of deriving the model predictions. The maximum d' value in time occuring for each audio fragment was used as model prediction. It may well be the case that listeners do not base their response on the maximum audible artefact that occurs during one interval but that they form some kind of average distortion measure. If the model predictions were based on an average output rather than the maximum output, lower model outputs are expected.

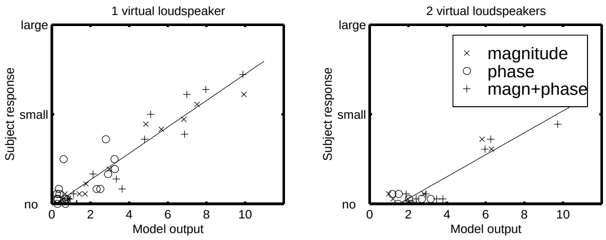

Fig. 8: Subjects' responses of the experiments described in section II as a function of the model output in terms of d'. The left panel shows data for a single virtual loudspeaker, the right panel for two simultaneous virtual loudspeakers. The crosses denote magnitude smoothing, the circles phase smoothing and the plus signs combined magnitude and phase smoothing. The solid lines are a linear fit.

DISCUSSION AND CONCLUSIONS

The results of the experiments demonstrate that the complex phase and magnitude spectra present in anechoic HRTFs can be simplified by the assumption that their spectral resolution does not have to exceed the spectral resolution of the cochlea. Specifically, a first-order gammatone filterbank with bandwidths of 1 ERB is sufficient to describe the frequency dependence of both the phase and magnitude spectra. The amount of crosstalk between filters of a first-order filter is substantially more than for a 4th-order filterbank, which is usually used to describe the spectral resolution of the auditory system. Furthermore, when the smoothing is strong enough to lead to audible differences, these are stronger for mangitude than for phase smoothing. Even if the phase spectra are almost linear (n=0.1), subjects indicate only very small differences, or do not report differences at all. This result is in line with the data from [18] where it is shown that linear phase HRTFs are not discriminable from unprocessed HRTFs. Our results also suggest that anechoic HRTFs can successfully be described by a linear-phase filter with a magnitude spectrum with a limited resolution, as long as the interaural delay matches the average delay found in the original HRTF set.

If more virtual loudspeakers are combined which have a partially overlapping spectrum, even more smoothing is allowed than for a single virtual source. This is due to masking across channels which results in a decreased sensitivity for interaural parameters of the different sources.

The model of [21] can successfully describe the perceptual degradation of the smoothing process, both for phase and magnitude smoothing. It can also account for the difference between one and two virtual loudspeakers. The transformation between physical differences between stimuli to the distance measure provided by the model results in a metric which is highly correlated with the scaled perceptual differences reported by subjects.

ACKNOWLEDGMENTS

The investigations were supported by the Research Council for Earth and Lifesciences (ALW) with financial aid from the Netherlands Organization for Scientific Research (NWO). We would like to thank Ronald Aarts and Daniël Schobben of Philips Research Laboratories Eindhoven for their valuable contributions and support.

Bibliography

[1] B. McA. Sayers. Acoustic image lateralization judgments with binaural tones. *J. Acoust. Soc. Am.*, 36:923–926, 1964.

[2] E. R. Hafter and S. C. Carrier. Masking-level differences obtained with pulsed tonal maskers. *J. Acoust. Soc. Am.*, 47:1041–1047, 1970.

[3] W. A. Yost. Lateral position of sinusoids presented with interaural intensive and temporal differences. *J. Acoust. Soc. Am.*, 70:397–409, 1981.

[4] F. L. Wightman and D. J. Kistler. Headphone simulation of free-field listening. I. Stimulus synthesis. *J. Acoust. Soc. Am.*, 85:858–867, 1989.

[5] D. S. Brungart, N. I. Durlach, and W. M. Rabinowitz. Auditory localization of nearby sources. II. Localization of a broadband source. *J. Acoust. Soc. Am.*, 106:1956–1968, 1999.

[6] B. G. Shinn-Cunningham, S. Santarelli, and N. Kopco. Tori of confusion: binaural localization cues for sources within reach of a listener. *J. Acoust. Soc. Am.*, 107:1627–1636, 2000.

[7] F. L. Wightman and D. J. Kistler. Headphone simulation of free-field listening. II: psychophysical validation. *J. Acoust. Soc. Am.*, 85:868–878, 1989.

[8] W. M. Hartmann and A. Wittenberg. On the externalization of sound images. *J. Acoust. Soc. Am.*, 99:3678–3688, 1996.

[9] E. H. A. Langendijk and A. W. Bronkhorst. Fidelity of three-dimensional-sound reproduction using a virtual auditory display. *J. Acoust. Soc. Am.*, 107:528–537, 2000.

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

[11] F. L. Wightman and D. J. Kistler. Resolution of front-back ambiguity in spatial hearing by listener and source movement. *J. Acoust. Soc. Am.*, 105:2841–2853, 1999.

[12] M. Zhang, K. Tan, and M. H. Er. Three-dimensional sound synthesis based on head-related transfer functions. *J. Audio. Eng. Soc.*, 146:836–844, 1998.

[13] N. I. Durlach and X. D. Pang. Interaural magnification. *J. Acoust. Soc. Am.*, 80:1849–1850, 1986.

[14] E. M. Wenzel and S. H. Foster. Perceptual consequences of interpolating head-related transfer functions during spatial synthesis. In *Proceedings of the 1993 workshop on applications of signal processing to audio and acoustics*, New York, 1993.

[15] D. J. Kistler and F. L. Wightman. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *J. Acoust. Soc. Am.*, 91:1637–1647, 1992.

[16] N. Cheung, S. Trautmann, and A. Horner. Head-related transfer function modeling in 3-D sound systems with genetic algorithms. *J. Audio Eng. Soc.*, 46:531–539, 1998.

[17] 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.

[18] A. Kulkarni, S. K. Isabelle, and H. S. Colburn. Sensitivity of human subjects to head-related transfer-function phase spectra. *J. Acoust. Soc. Am.*, 105:2821–2840, 1999.

[19] J. Huopaniemi and N. Zacharov. Objective and subjective evaluation of head-related transfer function filter design. *J. Audio. Eng. Soc.*, 47:218–239, 1999.

[20] B. R. Glasberg and B. C. J. Moore. Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47:103–138, 1990.

[21] J. Breebaart, S. van de Par, and A. Kohlrausch. Binaural processing model based on contralateral inhibition. I. Model setup. *J. Acoust. Soc. Am.*, Under review, 2001.

[22] J. Breebaart, S. van de Par, and A. Kohlrausch. Binaural processing model based on contralateral inhibition. II. Dependence on spectral parameters. *J. Acoust. Soc. Am.*, Under review, 2001.

[23] D. J. Breebaart, S. van de Par, and A. Kohlrausch. Binaural processing model based on contralateral inhibition. III. Dependence on temporal parameters. *J. Acoust. Soc. Am.*, Under review, 2001.

[24] P. I. M. Johannesma. The pre-response stimulus ensemble of neurons in the cochlear nucleus. In *Proceedings of the Symposium of Hearing Theory*, IPO, Eindhoven, The Netherlands, 1972.

[25] R. D. Patterson, J. Holdsworth, I Nimmo-Smith, and P. Rice. Svos final report: The auditory filterbank. Technical Report APU report 2341, 1988.

[26] J. Blauert, M. Brüggen, K. Hartung, A.W. Bronkhorst, R. Drullmann, G. Reynaud, L. Pellieux, W. Krebber, and R. Sotteck. The audis catalog of human hrtfs. In *Proc. 16th Int. Congr. Acoust.* ICA, Inst. of Physics, USA-NY, CD-ROM, 1998.

[27] D. McFadden. Masking-level differences determined with and without interaural disparities in masker intensity. *J. Acoust. Soc. Am.*, 44:212–223, 1968.

[28] J. W. Hall and A. D. G. Harvey. NoSo and NoSπ thresholds as a function of masker level for narrow-band and wideband masking noise. *J. Acoust. Soc. Am.*, 76:1699–1703, 1984.

[29] N. I. Durlach. Equalization and cancellation theory of binaural masking-level differences. *J. Acoust. Soc. Am.*, 35:1206–1218, 1963.

[30] N. I. Durlach. Binaural signal detection: Equalization and cancellation theory. In J.V. Tobias, editor, *Foundations of modern auditory theory*, volume II, pages 369–462. Academic Press, New York, London, 1972.