Subjective importance of individual HRTF phase

Eugen Rasumow[1], Matthias Blau[1], Steven van de Par[2], Martin Hansen[1], Simon Doclo[2],
Dirk Püschel[3] and Volker Mellert[2]
[1] *Jade-Hochschule Oldenburg, Institut für Hörtechnik und Audiologie, 26121 Oldenburg, Germany*
[2] *Carl von Ossietzky Universität Oldenburg, 26129 Oldenburg, Germany*
[3] *Akustik Technologie Göttingen, 37073 Göttingen, Germany*

Introduction

Spatial perception of sounds is determined by so-called head related transfer functions (HRTFs), which are known to be highly individual. However, the individual HRTF phase is typically assumed to be irrelevant and/or replaceable by a suitable linear or minimal phase (cf. [Kulkarni et al., 1999]). In this study, the direction-dependent discrimination when gradually fading the individual HRTF phase into a suitable linear phase was investigated using three-AFC listening tests with eight normal-hearing subjects. Analogically, in a second experiment the frequency range of relevance for (HRTF) phase discrimination was investigated by varying the crossover frequency between original phase for the lower and linear phase for the higher frequencies.

Motivation

Individual HRTFs contain characteristic spatial information, which encompasses interaural time and level difference cues but also, for instance, monaural spectral cues due to the shape of the pinna. For many applications (e.g. the virtual artificial head, cf. [Rasumow et al., 2011]), it is desirable to reduce the resolution of individual HRTFs or to smooth these HRTFs as long as this smoothing operation does not affect the perception. In the frequency domain, it is generally assumed that incoming sounds are grouped into relative frequency bands associated with so-called critical bands (cf. [Fletcher, 1940]). Disadvantageously, a straightforward smoothing of complex-valued HRTFs in the frequency domain in relative bands is complicated by the fact that measured HRTFs often exhibit too noisy and/or steep phases. A complex-valued smoothing of such HRTFs with noisy phases would result in very small magnitudes for certain frequencies, as depicted in Figure 1 (cf. [Rasumow et al., 2012b] and [Rasumow et al., 2012a]). These small magnitudes may result from destructive interferences when smoothing/averaging complex-valued data in frequency bands. Moreover, we assume that the detection of such a smoothing operation would be rather dominated by the notches for higher frequencies than by the reduction of the spectral fine structure. However, these implausible notches at higher frequencies may be avoided when the HRTF phase is substituted with an appropriate deterministic phase prior to the smoothing operation. Thus, one could take advantage of the hypothesized insensitivity to the exact phase spectrum, typically assumed with human listeners at higher frequencies, and indiscriminably preprocess HRTFs as to enable an adequate smoothing in the frequency domain. In the following the discriminability towards a broadband phase linearization as well as a crossover frequency for linearizing individual HRTF phases only at higher frequencies are investigated and discussed in two subsequent experiments.

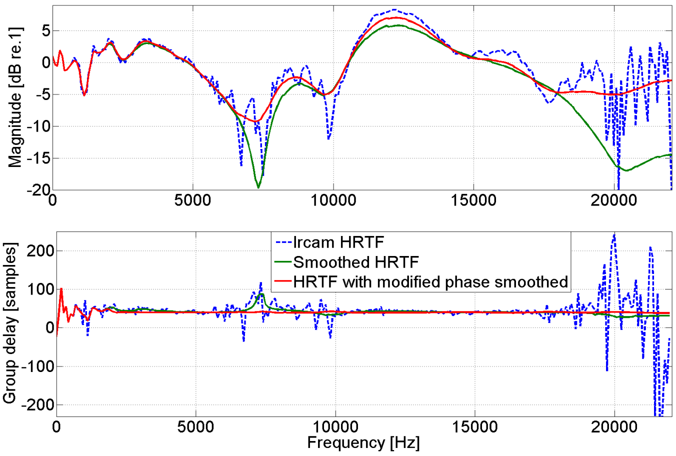

Figure 1: Exemplary HRTF from the IRCAM database (dashed blue graph) and the same HRTF smoothed in relative bandwidths of one-third octave according [Hatziantoniou and Mourjopoulos, 2000] with (red) and without (green) substituting the HRTF phase with a linear phase for frequencies $f \geq 2$ kHz are depicted as the magnitude in dB (top) and the group delay in samples (bottom) as a function of frequency in Hz.

Test preparation

At the very beginning of the test session, individual HRTFs were measured in an anechoic room with a circular loudspeaker array (radius 1.1 m). The HRTF measurements were carried out using the blocked ear method according to [Hammershoi and Moller, 1996] with custom made ear shells and embedded Knowles FG-23329 miniature electret microphones. The transfer functions were estimated using a 8192-point Hann window, 50% overlap and 52 averages. All measured transfer functions were equalized by the free field transfer function measured at the position of the center of the head without the acoustic influence of the subject, using a free-field microphone (G.R.A.S. 40AF). The resulting impulse responses were truncated to 512 samples ($\approx$ 11,6 ms at a sampling rate of $f_s = 44,1$ kHz, cf. [Rasumow et al., 2012b]) and the tails of the impulse responses were flanked with tapered Hann windows with a descending flank of 50 samples. As to provide a rough overview of a potential direction-dependency of the tested variables and to limit the number of experiments to a manageable amount, four different source positions were chosen in the horizontal plane with azimuth angles 0° (front), 90° (left), 225° (back right) and 315° (front right).

General procedure and stimuli

As to determine the respective threshold values, a three-alternative forced-choice paradigm in conjunction with the one up- one down method, converging to a 50%-correct value on the psychometric function was applied. This threshold was chosen to represent the just noticeable difference (JND). Three intervals of filtered test signals (each separated by 0.3 seconds silence) were presented to the subjects in randomized order. One of these signals was filtered with HRTFs with modified phases and two signals were filtered with the original HRTF (reference). The subjects were instructed to detect the odd interval out of the three presented. The two different modifications of the HRTF phase are described further below.

The test signal consisted of bursts of white noise with a spectral content of 150 Hz $< f <$ 18050 Hz. As to avoid a pitch cue due to sharp edges in the frequency domain, the spectral amplitude of the noise bursts was faded in between 150 and 200 Hz and faded out between 18000 and 18050 Hz using tapered Hann windows. Each test signal itself consisted of three noise bursts of 0,15 seconds noise and 0,15 seconds silence, gated with 0,01 seconds onset-offset ramps. All signals were individually headphone-equalized and presented via headphones (AKG K 240 Studio) at an overall sound pressure level of 78 dB SPL (calibrated with a G.R.A.S. type 43AA artificial ear). The experiments were designed using psylab (cf. [Hansen, 2006]), a set of MATLAB-scripts for the implementation of psychoacoustical detection and discrimination experiments.

Experiment I

In the first experiment the discrimination of a broadband manipulation of the individual HRTF phase was investigated. To do so, the individual HRTF phase was substituted by a mixture of the original phase $\phi_{\text{orig}}(f)$ and a linear phase $\phi_{\text{lin}}(f)$, i.e.

$$\phi_{\text{test}}(f) = L_\phi \cdot \phi_{\text{lin}}(f) + (1 - L_\phi) \cdot \phi_{\text{orig}}(f). \qquad (1)$$

Note that the length of the impulse response was kept constant (512 samples) throughout all experiments. The blending factor L_ϕ ranged between $L_\phi = 0$ (original phase) and $L_\phi = 1$ (no original phase present) and was kept constant for all frequencies. The slope of $\phi_{\text{lin}}(f)$ was calculated by determining the delay of the maximum of the Hilbert envelope of the impulse response in the time domain. The initial value was set to $L_\phi = 0,5$ and the initial step size was set to $\Delta L_\phi = 0,2$. L_ϕ was varied in the familiarization phase until a minimal step size of $\Delta L_\phi = 0,05$ was reached.

Experiment II

In the second experiment, only the phase for frequencies $f > f_c$ was substituted with a linear phase while the phase for frequencies $f \leq f_c$ remained unchanged. By varying f_c, the frequency range for a non-discriminable phase linearization of HRTFs is investigated. Note that in contrast to [Rasumow et al., 2012a] and [Rasumow et al., 2012b] no blending between the original phase for the lower and the linear phase for higher frequencies was applied. This aspect was changed since unpublished investigations suggest that any blending operation between phases could introduce additional cues for the discrimination task. In this experiment the linear phase also was calculated by determining the delay of the maximum of the Hilbert transform of the impulse response in the time domain.

The initial value of the crossover frequency was set to $f_c = 300$ Hz and the initial step size to $\Delta\ f_c = 160$ Hz. f_c was varied in the familiarization phase until a minimal step size of $\Delta\ f_c = 20$ Hz was reached.

Results - Experiment I

The results from the first experiment are depicted as the individual mean and standard deviation of L_ϕ at threshold in figure 2. We assume that smaller L_ϕ correspond to a higher sensitivity to broadband modifications of the individual HRTF phase, since for these conditions the subject is able to detect already relative small changes of the original phase ϕ_{orig}. On the other hand, with larger L_ϕ the perceptual difference between ϕ_{orig} and ϕ_{lin} is assumed to be less salient.

Considering the assumed role of L_ϕ, the results in figure 2 indicate a rather individual and apparently unpredictable distribution of L_ϕ at threshold. Also, the standard deviation of the measured L_ϕ at threshold does not seem to follow any trend. None of the tested subjects showed a complete inability to discriminate ϕ_{orig} from ϕ_{lin}, which would have resulted in very large L_ϕ. Hence, these results indicate that a broadband linearization of the HRTF phase is at least partly discriminable, which to some extent is in contrast to the findings from [Kulkarni et al., 1999]. For many subjects, the broadband phase manipulation was perceived least salient for the frontal direction (0°). Furthermore, for many subjects the phase manipulation is best discriminable with rather lateral directions. On the other hand, in both cases there are subjects who do not follow this trend.

We suppose that the various L_ϕ at threshold correlate with the individually differing phase changes introduced by ϕ_{lin}. From a perceptual point of view, one may assume that a binaural aspect, namely the interaural phase difference (IPD) for lower frequencies, dominates the discrimination associated with phase manipulations. In detail, we suppose the IPD for lower frequencies to be the primal aspect for the perception of phase changes (cf. [Perrott and Nelson, 1969]). In order to quantify the individual perceptual changes associated with ϕ_{lin}, we introduce the modeled individual discrimination ability

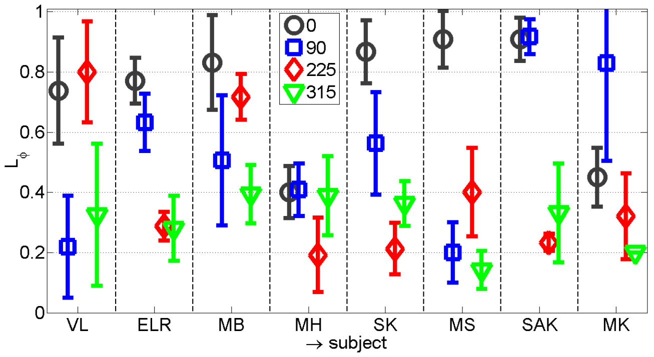

Figure 2: The symbols and whiskers represent the individual mean and standard deviation of L_ϕ (cf. equation 1) at threshold as a function of subjects (x-axis) and direction (legend).

L_{mod} as

$$L_{\mathrm{mod}} = \nu \cdot \sum_{\mathrm{f}=150\text{ Hz}}^{1500\text{ Hz}} \frac{\Delta \mathrm{f}}{\mathrm{ERB(f)}} \cdot \frac{|\mathrm{IPD_{orig}(f)} - \mathrm{IPD_{lin}(f)}|}{\phi_{\mathrm{JND}}(\mathrm{f})}$$

$$\text{with} \sum_{\mathrm{f}=150\text{ Hz}}^{1500\text{ Hz}} \left(\frac{\Delta \mathrm{f}}{\mathrm{ERB(f)}}\right) \cdot \nu = 1, \tag{2}$$

where ν is a normalization constant, $\Delta f = \frac{f_s}{\mathrm{NFFT}}$ is the frequency resolution, IPD(f) is the frequency-dependent interaural phase difference for linear and original phase, ERB(f) is the frequency-dependent equivalent rectangular bandwidth according to [Glasberg and Moore, 1990] and ϕ_{JND} is the JND for interaural phase differences (taken from [Klumpp and Eady, 1956]).

We suppose that larger L_{mod} correlate with a rather clear discrimination of the phase manipulation and thus with smaller L_ϕ at threshold. This relationship between L_ϕ and L_{mod} is illustrated graphically in figure 3. The rather high linear correlation coefficient of $\rho(L_{\mathrm{mod}}, L_\phi) = -0.83$ substantiates that L_{mod} is well suited to predict the individual sensitivity to HRTF phase modifications. Based on this correlation between L_{mod} and L_ϕ, we in turn hypothesize that the linearization of the HRTF phase may be by far less discriminable if ϕ_{orig} is preserved for the lower frequencies. This assumption, however, was to be verified by the second experiment.

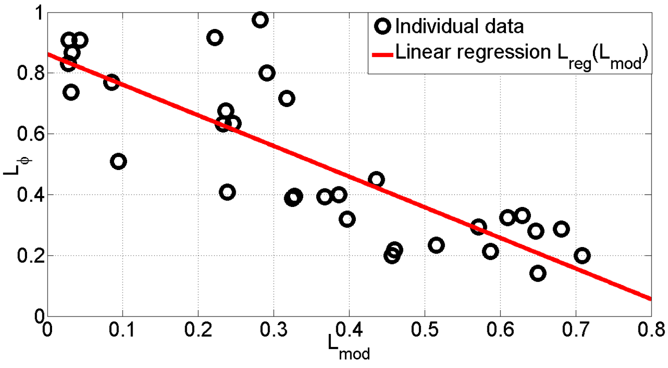

Figure 3: Mean L_ϕ as a function of the (modeled) individual discrimination ability L_{mod} (cf. equation 2)

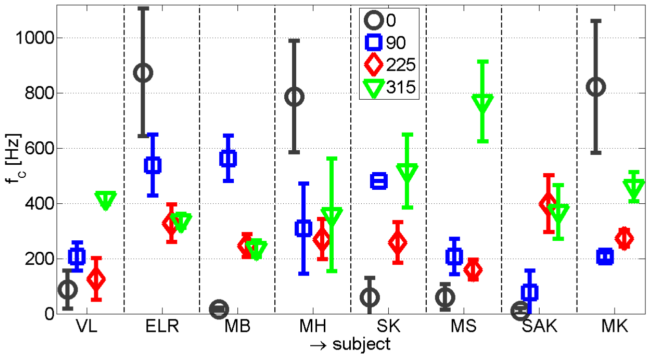

Figure 4: Mean crossover frequencies f_c at threshold and corresponding standard deviations are shown on the y-axis in Hz as a function of subjects (x-axis) and direction (legend).

Results - Experiment II

The results from the second experiment, examining the crossover frequency f_c between ϕ_{orig} for the lower and ϕ_{lin} for the higher frequencies are depicted in figure 4. There, f_c in Hz is plotted on the y-axis in separated boxes and colors, indicating the various subjects and directions, respectively. As in the results from the first experiment, the f_c at threshold do not seem to follow a deterministic direction- and/or subject-dependent pattern. The mean f_c at threshold reach from 50 Hz to approximately 900 Hz. However, there is a slight trend towards larger standard deviations for higher f_c.

In general, we assume that lower f_c correspond to a more salient discrimination since here a larger proportion of the original phase ϕ_{orig} is substituted with ϕ_{lin}. Higher f_c, on the other hand, imply that the original phase is preserved for a wider frequency range and thus are supposed to result in less salient discriminations. Hence, the most critical or sensitive subjects/cases when substituting the original phase with a linear phase at higher frequencies are those with higher f_c in figure 4. The highest f_c ($f_c \approx 900$ Hz) mostly occur with the frontal direction (0°, subject ELR, MH and MK). Yet, for some subjects the highest intraindividual f_c also occur with lateral directions (90° and 315°, subject VL, MB, SK, MS and SAK). Considering the results from the two experiments, one would expect that the conditions associated with a poor sensitivity to broadband phase modification (large L_ϕ in figure 2) would also correspond to small crossover frequencies f_c in the second experiment. When comparing the results in figure 2 and 4, it can be seen that the lowest f_c values (reaching down to $f_c \approx 0$ Hz) for subjects MB (0°), SK (0°), MS (0°) and SAK (0° and 90°) indeed correspond to L_ϕ close to $L_\phi \approx 1$. Large L_ϕ, e.g. for VL (225°) and MK (90°), imply a lacking discriminability for broadband phase modifications. Consistently, for these conditions the according f_c reach down to $f_c = 120$ Hz and $f_c = 205$ Hz, which is approximately the lower limit of the test signal. In sum the observed results imply that a non-discriminable linearization of HRTF phases is possible if the original phase is maintained for lower frequencies up to $\approx$ 1 kHz.

Remarkably, the f_c from figure 4 are constantly smaller compared with f_c in a previous studies ([Rasumow et al., 2012a] and [Rasumow et al., 2012b]). We assume that in the presented experiments the omission of fading between ϕ_{orig} and ϕ_{lin} resulted in less discriminable cues and thus in smaller f_c.

Discussion

Considering the results in [Kulkarni et al., 1999, figure 8], one may assume that the ability of human listeners to discriminate between linearized and unmodified HRTF phases is approximately at chance. Our results in figure 2, however, suggest a more pronounced sensitivity to broadband phase linearizations. In detail, none of the eight subjects of the current investigation showed a complete inability to detect a broadband phase linearization. A reasonable explanation for this difference may lie in the use of different testing paradigms in both investigations. It is worth noting that fading the measured phase into a minimum phase plus delay condition (as also tested in [Kulkarni et al., 1999]) resulted in even better discriminable cues and thus was not investigated further. Furthermore, the results in [Kulkarni et al., 1999, figure 9] indicate that phase manipulations are less discriminable for higher frequencies ($f > 2000$ Hz). This finding is in line with our current results from figure 4. More precisely, our results suggest that the frequency range which is relevant for phase discrimination reaches up to approximately 1000 Hz.

Conclusion

The results from the first experiment indicate an individual and apparently unpredictable sensitivity towards a broadband linearization of the individual HRTF phase. Furthermore, these results imply that a broadband phase linearization is at least party discriminable. Consequently such a broadband phase linearization seems to be an inappropriate preprocessing for a complex-valued smoothing in the frequency domain. The modeling according to equation 2 suggests that the discrimination of the phase manipulation is primarily bound to the IPD for lower frequencies. In the second experiment the results indicate that the original phase needs to be preserved for approximately $f < 1000$ Hz, while the phase for $f > 1000$ Hz may be substituted with a suitable linear phase without yielding discriminable artifacts compared with the original HRTF. Based on these findings it seems appropriate to linearize the HRTF phase above 1000 Hz before smoothing, as for instance shown in figure 1.

Outlook

The described parameters enable a non-discriminable phase manipulation for an appropriate smoothing in the frequency domain. On this basis we will investigate the impact of smoothing HRTFs in the frequency domain and its potential advantages with regard to the implementation of the virtual artificial head.

Acknowledgment

This project was partially funded by Bundesministerium für Bildung und Forschung under grant no. 17080X10.

References

[Fletcher, 1940] Fletcher, H. (1940). Auditory patterns. *Rev. Mod. Phys.*, 12:47–65.

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

[Hammershoi and Moller, 1996] Hammershoi, D. and Moller, H. (1996). Sound transmission to and within the human ear canal. *The Journal of the Acoustical Society of America*, 100(1):408–427.

[Hansen, 2006] Hansen, M. (2006). Lehre und Ausbildung in Psychoakustik mit psylab: Freie Software für psychoakustische Experimente. In *Fortschritte der Akustik – DAGA '06*, pages 591–592, Braunschweig. Dega.

[Hatziantoniou and Mourjopoulos, 2000] Hatziantoniou, D. P. and Mourjopoulos, J. N. (2000). Generalized fractional-octave smoothing of audio and acoustic responses. *J. Audio Eng. Soc.*, 48(4):259–280.

[Klumpp and Eady, 1956] Klumpp, R. G. and Eady, H. R. (1956). Some measurements of interaural time difference thresholds. *The Journal of the Acoustical Society of America*, 28(5):859–860.

[Kulkarni et al., 1999] Kulkarni, A., Isabelle, S. K., and Colburn, H. S. (1999). Sensitivity of human subjects to head-related transfer-function phase spectra. *The Journal of the Acoustical Society of America*, 105(5):2821–2840.

[Perrott and Nelson, 1969] Perrott, D. R. and Nelson, M. A. (1969). Limits for the detection of binaural beats. *The Journal of the Acoustical Society of America*, 46(6B):1477–1481.

[Rasumow et al., 2012a] Rasumow, E., Blau, M., Doclo, S., Hansen, M., Mellert, V., Püschel, D., and van de Par, S. (2012a). Psychoakustisch motivierte Glättung von kopfbezogenen Übertragungsfunktionen: Hörbarkeit der Linearisierung von Phasengängen. In *Fortschritte der Akustik - DAGA 2012*, pages 633–634. DEGA.

[Rasumow et al., 2011] Rasumow, E., Blau, M., Hansen, M., Doclo, S., van de Par, S., Mellert, V., and Püschel, D. (2011). Robustness of virtual artificial head topologies with respect to microphone positioning errors. In *Proc. Forum Acusticum, Aalborg*, pages 2251–2256, Aalborg.

[Rasumow et al., 2012b] Rasumow, E., Blau, M., Hansen, M., Doclo, S., van de Par, S., Püschel, D., and Mellert, V. (2012b). Smoothing head-related transfer functions for a virtual artificial head. In *Acoustics 2012*, pages 1019–1024, Nantes, France.