

# Effect of perceptually irrelevant variance in head-related transfer functions on principal component analysis

Jeroen Breebaart<sup>a)</sup>

*Dolby Laboratories, Level 3, 35-51 Mitchell Street, McMahon's Point,  
New South Wales 2060, Australia  
jeroen.breebaart@dolby.com*

**Abstract:** The significant amount of variance in head-related transfer functions (HRTFs) resulting from source location and subject dependencies have led researchers to use principal components analysis (PCA) to approximate HRTFs with a small set of basis functions. PCA minimizes a mean-square error, and consequently may spend modeling effort on perceptually irrelevant properties. To investigate the extent of this effect, PCA performance was studied before and after removal of perceptually irrelevant variance. The results indicate that from the sixth PCA component onward, a substantial amount of perceptually irrelevant variance is being accounted for.

© 2013 Acoustical Society of America

PACS numbers: 43.66.Ba, 43.66.Pn, 43.66.Qp [JL]

Date Received: May 23, 2012 Date Accepted: November 16, 2012

## 1. Introduction

The most common method to simulate one or more virtual sound sources and the associated acoustic pathway from each source to both eardrums comprises a linear, time-invariant system modeling approach. In this method, the acoustic pathway is described by means of an impulse response, referred to as head-related impulse response (HRIR) or binaural room impulse response, for anechoic or echoic conditions, respectively.<sup>1,2</sup> The number of HRIR filter coefficients that is required to synthesize a virtual sound source as a function of subject, azimuth, elevation, and distance can be substantial, as it is often argued that a head-related transfer functions (HRTF) database resolution of  $\sim 5^\circ$  is required to avoid spatial aliasing.<sup>3,4</sup> As a consequence, several studies have addressed methods to reduce the amount of information in an HRTF database. One such method is to decompose HRTF magnitude spectra by means of a Karhunen-Loève transform [also referred to as principal component analysis, (PCA)] to result in a small set of basis functions and position-dependent weights.<sup>5-9</sup> These studies show that typically 90% or more of the HRTF database variance can be accounted for by five basis functions with subject and position-dependent weights, whereas  $\sim 10$ – $20$  basis functions seem to be required for accurate psycho-physical localization performance.<sup>8</sup> Further, results from Hwang and Park<sup>10</sup> suggest that PCA operates less efficiently in the time (HRIR) domain than in the frequency (HRTF) domain.

A potential problem of the PCA method is that it minimizes the mean-square error of the approximated HRTFs, even if certain errors that are important in a least mean-square sense are perceptually irrelevant.<sup>11</sup> The absence of a strong correlation between the mean-square error and psycho-physical detection performance is shown by Scarpaci and Colburn.<sup>9</sup> It is therefore of interest to evaluate the influence of perceptually irrelevant variance on PCA. If the amount of irrelevant variance that is accounted for by PCA is substantial, the data reduction achieved by PCA is likely to be

---

<sup>a)</sup>The work described in this paper represents independent research not affiliated with Dolby Laboratories.

suboptimal, and improved performance can be expected if PCA is forced not to put effort in modeling such irrelevant variance. The goal of this study is therefore to investigate the performance of PCA if irrelevant information is removed from HRTFs.

## 2. Method

PCA models datasets by describing (or reconstructing) the data as a linear combination of a set of orthonormal basis functions (or components). A low-dimensional approximation of the dataset can be obtained by truncating the number of components. PCA minimizes the mean-square error between original and reconstructed dataset, or said differently, maximizes the amount of variance that can be accounted for. The amount or proportion of variance accounted for by one or more components can be computed by an eigenvalue decomposition of the covariance matrix of the data. The reader is referred to other publications<sup>10,12</sup> for more details on this method. To study the influence of variance that is relevant in a least mean-square sense but is irrelevant from a perceptual point of view, the proportion of variance accounted for by PCA is computed before and after *removal* of irrelevant information from HRTFs. Only the magnitude part of HRTF spectra is considered, and (inter-aural) time and/or phase differences are assumed to be treated separately.

The perceptually motivated HRTF parameterization method described by Breebaart *et al.*<sup>13</sup> was used to reduce the perceptually irrelevant variance in HRTF magnitude spectra. The underlying assumption of this method is that the sound source localization cues accessible to the auditory system are averaged in critical bands. In mathematical terms, the parameterization method can be described by a projection of an HRTF magnitude vector  $\vec{x}_i$  with index  $i$  referring to the sound source location, subject and ear of the associated HRTF,

$$\vec{p}_i = \mathbf{M}\vec{x}_i, \quad (1)$$

with  $\vec{p}_i$  a critical-band parameter vector of length  $B=41$ , and matrix  $\mathbf{M}$  comprising the averaging weights to compute the parameter values. The reconstructed (referred to as “processed”) HRTF magnitude vector  $\vec{y}_i$  is then obtained by

$$\vec{y}_i = \mathbf{M}^T \mathbf{M}^+ \vec{p}_i \quad (2)$$

with  $\mathbf{M}^+$  the Moore–Penrose pseudo-inverse of  $\mathbf{M}$ .

Breebaart *et al.*<sup>13</sup> demonstrated that one spectral magnitude parameter for each critical band, in combination with triangularly shaped, 50% overlapping parameter analysis and interpolation does not result in audible differences between the original and reconstructed HRTFs. In the remainder of this paper, we therefore use this configuration for the removal of perceptually irrelevant variance.

The Center for Image Processing and Integrated Computing HRIR database that is publicly available from the University of California<sup>14</sup> was used to evaluate PCA performance. A total of 110 000 HRIRs present in the database were converted to the frequency domain using a discrete Fourier transform. To study the effect of the domain PCA is operating on, PCA was applied on the linear magnitude spectrum as well as on the logarithmic spectrum.

## 3. Results

### 3.1 Proportion of variance

The proportion of variance not accounted for as a function of the number of PCA components is shown in Fig. 1. The left-hand panel represents PCA operating on linear magnitude spectra; the right-hand panel corresponds to PCA operating on a logarithmic spectrum representation. The crosses represent the original HRTF spectra, whereas the circles denote the performance of the processed HRTFs (e.g., after irrelevancy removal). For all conditions tested, the proportion of variance that is not

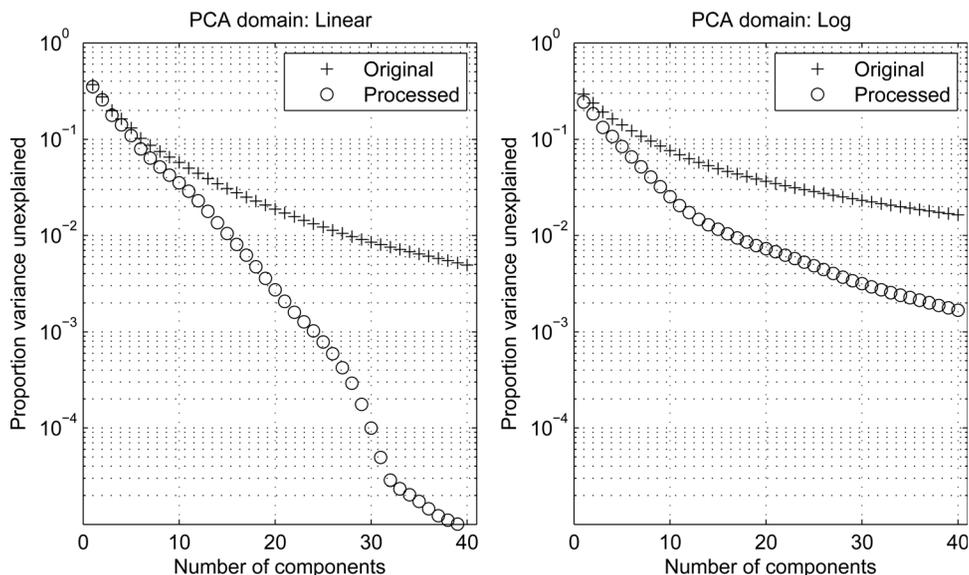


Fig. 1. Unexplained proportion of variance as a function of the number of PCA components. The left-hand panel represents PCA operating in the linear domain; the right-hand panel corresponds to PCA in the logarithmic domain. The crosses represent the performance for the original HRTF spectra; the circles represent the HRTF spectra after irrelevancy removal.

accounted for decreases with the number of components. Irrespective of some minor deviations, for all four conditions a proportion of variance accounted for of  $\sim 90\%$  is achieved for five components, which is in line with results from other studies.<sup>8,9</sup>

The differences between the original and processed HRTF spectra become articulated for more than five components. In that case, both the linear and logarithmic HRTF spectra of the processed HRTF database show a stronger convergence and smaller residual variance than for the original HRTFs. This difference is stronger for PCA in the linear domain compared to the logarithmic domain. In the linear domain, 30 components account for 99% and 99.99% of the variance, for the original and processed HRTFs, respectively. In the logarithmic domain, these proportions of variance amount to 98% and 99.7%. The observation that PCA in the linear domain converges faster than in the logarithmic domain is also in line with results reported by Leung and Carlile.<sup>8</sup>

### 3.2 Root mean-square error

The proportion of variance that is (un)explained for a given number of components does not reveal information on the magnitude of the errors, such as the root mean-square error (RMSE). Therefore, RMSEs were calculated in the logarithmic domain in the same manner as employed by Scarpaci and Colburn.<sup>9</sup> The PCA analysis and synthesis was performed in the linear and in the logarithmic domain. The results are shown in Fig. 2. The left-hand panel represents the RMSE for PCA applied in the linear magnitude spectrum domain; the right-hand panel denotes the RMSE for PCA in the logarithmic domain.

In line with the results for unexplained variance, the processed magnitude spectra have a smaller RMSE than the original magnitude spectra, indicating a benefit from irrelevancy removal. This benefit amounts to 1 and 2 dB, for 20 components and PCA applied in the linear and logarithmic domain, respectively. In contrast to the results for the unexplained proportion of variance, PCA in the linear domain results in significantly larger errors expressed in decibels compared to PCA applied in the logarithmic domain. This is likely to be caused by the fact that both PCA and the RMSE

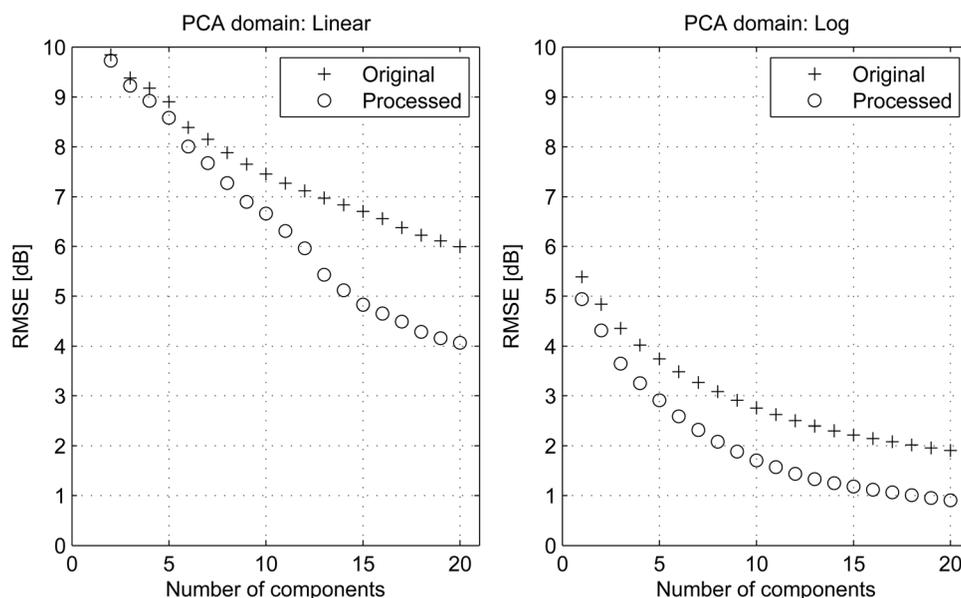


Fig. 2. HRTF root mean-square error (RMSE) as function of the number of PCA components. The left-hand panel represents PCA operating in the linear domain; the right-hand panel corresponds to PCA in the logarithmic domain. The crosses represent the performance for the original HRTF spectra; the circles represent the HRTF spectra after irrelevancy removal. The RMSE was calculated in the logarithmic (dB) domain.

operate in the same (logarithmic) domain, as opposed to the case in which PCA operates in the linear domain and errors are formulated in the logarithmic domain.

#### 4. Discussion

When PCA is applied to linear HRTF magnitude spectra, the results obtained in this study are very much in line with those published earlier in that  $\sim 90\%$  of the HRTF variance can be explained by the first five PCA components.<sup>5,7-9</sup> Further, the observed slower convergence of PCA applied in the logarithmic domain compared to the linear domain is also in line with data obtained by Leung and Carlile.<sup>8</sup> When comparing the proportion of unexplained variance between the original and processed HRTF spectra for up to five PCA components, only small differences are observed. This result suggests that PCA applied to HRTFs with or without perceptual irrelevancies removed is practically equivalent for the first five components. An important question is therefore whether five components are sufficient for perceptual transparency. Kistler and Wightman<sup>7</sup> stated that “Subjects’ judgments of the apparent directions of headphone-presented sounds that had been synthesized from the modeled HRTFs were nearly identical to their judgments of sounds synthesized from measured HRTFs.” More recent studies indicate however, that localization performance continues to improve when the number of components is increased from 5 to 10 or 20.<sup>8,9</sup> The RMSEs reported in Fig. 2 are also significant for just five components, especially if one assumes that the just-noticeable difference for spectral changes in HRTFs can be less than 1 dB in critical conditions.<sup>15</sup> When more than five PCA components are employed, the benefits of perceptual irrelevancy removal in HRTFs become most apparent. For 20 PCA components, the proportion of unexplained variance is almost 1 order of magnitude smaller when irrelevancy is removed prior to PCA analysis. Further, when PCA operates in the logarithmic domain, the RMSE is reduced by up to 50% when perceptual irrelevancy is removed.

The presence of variance that is perceptually irrelevant is not the only caveat for PCA applied to HRTF spectra. The RMSEs shown in Fig. 2 also indicate the

importance of the domain in which the variance or errors are evaluated. The results obtained for the proportion of variance accounted for (Fig. 1) seem to suggest that PCA operating in the linear magnitude domain is more efficient than in the logarithmic domain. On the contrary, if the RMSE is evaluated in the logarithmic domain, PCA operating in that same domain results in significantly smaller RMSEs than PCA operating in the linear magnitude domain. These observations stress the importance of a well-chosen optimization criterion that is indicative of perceptual quality. An accurate model of perceptual differences given a modified HRTF would be very helpful and possibly even a necessity to optimize the performance of PCA, or any other HRTF data reduction method for its ability to efficiently model HRTF spectra.

Although the results presented in this paper suggest significant benefits of perceptual irrelevancy removal in terms of PCA data compaction, there exists a risk that the changes imposed by irrelevancy removal and PCA individually do not result in perceptual differences, whereas the combined effect may, in fact, be audible. The existence and/or severity of such combined effect can only be demonstrated through psycho-physical validation or by means of an accurate perceptual model.

## 5. Conclusions

Principal component analysis is regularly used as a means for data compaction or efficient approximation of HRTF magnitude spectra, despite the risk of spending modeling effort on variance that is perceptually irrelevant. By comparison of PCA performance before and after perceptual irrelevancy removal, it can be concluded that the first five PCA components seem to capture predominantly perceptually relevant properties of HRTFs, whereas subsequent components are less effective from a perceptual viewpoint. Further, for PCA to be most effective, it seems important that the PCA optimization criterion (or residual error that is minimized) is formulated in a domain that correlates well with perceptual quality.

## References and links

- <sup>1</sup>F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. I. Stimulus synthesis," *J. Acoust. Soc. Am.* **85**, 858–867 (1989).
- <sup>2</sup>F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. II: Psychophysical validation," *J. Acoust. Soc. Am.* **85**, 868–878 (1989).
- <sup>3</sup>T. Ajdler, L. Sbaiz, and M. Vetterli, in "The plenacoustic function on the circle with application to HRTF interpolation," *Proceedings of the ICASSP* (IEEE, New York, 2005), pp. 273–276.
- <sup>4</sup>T. Ajdler, L. Sbaiz, and M. Vetterli, "Head related transfer functions interpolation considering acoustics," in *Proceedings of the 118th AES Convention*, Barcelona, Spain (May 2005).
- <sup>5</sup>J. Chen, B. D. Van Veen, and K. E. Hecox, "A spatial feature extraction and regularization model for the head-related transfer function," *J. Acoust. Soc. Am.* **97**, 439–452 (1995).
- <sup>6</sup>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).
- <sup>7</sup>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).
- <sup>8</sup>J. Leung and C. Carlile, "PCA compression of HRTFs and localization performance," in *Proceedings of the International Workshop on the Principles and Applications of Spatial Hearing* (November 2009).
- <sup>9</sup>J. Scarpaci and H. Colburn, "Principal components analysis interpolation of head related transfer functions using locally-chosen basis functions," *J. Acoust. Soc. Am.* **117**, 2561–2562 (2005).
- <sup>10</sup>S. Hwang and Y. Park, "Interpretations on principal components analysis of head-related impulse responses in the median plane," *J. Acoust. Soc. Am.* **123**, EL65–EL71 (2008).
- <sup>11</sup>J. Breebaart and A. Kohlrausch, "The perceptual (ir)relevance of HRTF magnitude and phase spectra," in *Preprint No. 5406, 110th AES Convention*, Amsterdam, The Netherlands (2001).
- <sup>12</sup>W. Wahab Hugeng and D. Gunawan, "Improved method for individualization of head-related transfer functions on horizontal plane using reduced number of anthropometric measurements," *J. Telecommun.* **2**(2), 31–41 (2010).
- <sup>13</sup>J. Breebaart, F. Nater, and A. Kohlrausch, "Spectral and spatial parameter resolution requirements for parametric, filter-bank-based HRTF processing," *J. Audio. Eng. Soc.* **58**(3), 126–140 (2010).

<sup>14</sup>V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in *Proceedings of the Workshop on Applications of Signal Processes to Audio and Acoustics*, Mohonk Mountain House, New Paltz, NY (October 2001) (IEEE, New York, 2001).

<sup>15</sup>P. F. Hoffmann and H. Møller, "Some observations on sensitivity to HRTF magnitude," *J. Audio Eng. Soc.* **56**(11), 972–982 (2008).