

Proceedings of Meetings on Acoustics

Volume 19, 2013

<http://acousticalsociety.org/>



ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Psychological and Physiological Acoustics
Session 2aPPa: Binaural Hearing and Binaural Techniques II

2aPPa2. Perceptual equalization of artifacts of sound reproduction via multiple loudspeakers

Bernhard U. Seeber* and Ervin R. Hafter

***Corresponding author's address: Audio Information Processing, Technische Universität München, Arcisstrasse 21, Munich, 80333, BY, Germany, seeber@tum.de**

Several techniques for reproducing spatial sounds via multiple loudspeakers have been developed in recent years. A key problem for such techniques are comb filter effects caused by the uncertainty of the receiver position when playing coherent sounds from multiple loudspeakers (spatial aliasing). Here we studied if panning between two closely-spaced loudspeakers can create a virtual source that resembles that of a true source. This requires not only that panned direction and speaker position correspond, but also that source width, loudness, timbre, and temporal aspects are reproduced without perceivable error. A listening experiment in an anechoic chamber showed that panned sources differ primarily in loudness and timbre from a real source at the panned location. The artifacts are caused by effects of the head, and we investigated if they can be compensated by filtering the sounds. Compensation filters were derived from simulations of the sound field at the ears. Listening tests showed that compensation filters reduced panning errors to be nearly inaudible and level roving or reflections in the reproduction room made errors inaudible. We conclude that a simple equalization is sufficient to render panned sources from nearby speakers perceptually equivalent to real sources.

Published by the Acoustical Society of America through the American Institute of Physics

INTRODUCTION

Several techniques for reproducing spatial sounds via multiple loudspeakers have been developed and refined in recent years. Techniques such as wavefield synthesis, ambisonics or cross-talk cancellation aim to simulate the sound field at the listener's ear for sounds from different directions (Vorländer, 2008). With amplitude panning, the same sound is played coherently from two loudspeakers, and a variation in sound pressure level between speakers leads to a shift of the perceived sound direction (Pulkki, 1997). A key problem of amplitude panning and other techniques which play coherent sounds from multiple loudspeakers are comb filter effects (aliasing). Several studies have investigated the perceived direction of amplitude panning (e.g., Pulkki and Karjalainen, 2001; Griesinger, 2002), but we ask if panning can truly replace a loudspeaker at the panned location by not permitting any audible panning errors. This requires not only that the panned direction is accurate, but also that source width, loudness, timbre and temporal aspects are reproduced without perceivable error. For a large loudspeaker spacing the panning error will be audible because the spectrum on both ears will differ from each other, but if both speakers are spaced only 7.5° or 15° apart, which is large compared to what is commonly used for wavefield synthesis (Spors and Ahrens, 2007), spectral differences are smaller and errors might become imperceptible. An informal listening experiment in an anechoic chamber showed that panned sources differ primarily in loudness and timbre from a real source at the panned location. We investigated if these artifacts can be compensated by filtering the sounds and we developed a new approach for equalizing the panning sound with compensation filters derived from simulations of the sound field at the ears – the PEP technique (Perceptually Equalized Panning). In an in-depth evaluation of three different parameterizations of the equalization technique in listening tests we found that compensation filters reduced panning errors to be nearly inaudible. Level roving or sound reflections in the reproduction room made errors inaudible, indicating that errors should be inaudible in practical applications. We conclude that a simple equalization is sufficient to render panned sources from nearby speakers perceptually equivalent to real sources. The simplicity of the equalization approach ensures that results are also valid for listeners wearing hearing devices.

ANALYSIS OF PANNING ERRORS

Figure 1 shows the general layout used in deriving and assessing the panning technique. Sounds were either played directly from the loudspeaker in the middle (LS2) between LS1 and LS3, or a virtual, panned source was generated at the LS2 position by playing a coherent sound from both outer speakers (LS1 and LS3). Loudspeakers of the Simulated Open Field Environment (SOFE) were used (Seeber *et al.*, 2010) which were carefully equalized in amplitude and phase such that the theoretical 6 dB summation was obtained at the listener's head position when playing a coherent signal from two speakers. For panned stimuli, the 6 dB phase-coherent summation likewise occurs at the ears at low frequencies. However, the common sinusoidal panning law (Vector-based amplitude panning) corrects for only 3 dB summation, equal to intensity addition ($\sin(45^\circ) = 0.707$) (Pulkki, 1997; Hafter and Seeber, 2004), resulting in a 3 dB error. The estimated panning error in the ear signals is depicted in Figure 2. At high frequencies, interference effects lead to pronounced spectral differences between the panned (virtual) and the direct source as well as across the ears. Listening tests revealed audible differences between panned and true sources, primarily in timbre. Here we follow the idea that a single equalization filter applied identically to both loudspeaker signals might alleviate these errors and render them inaudible.

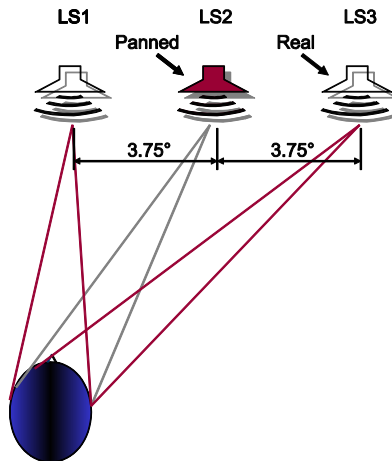


FIGURE 1. Sketch of the panning and test situations. A virtual source at position LS2 is created by playing the sound from loudspeakers LS1 and LS3. In the experiment, the virtual source at LS2 position is compared to the unequalized sound played from the loudspeaker LS2, i.e. a real source. Panning errors stem from interference effects and differences in HRTFs on the ears.

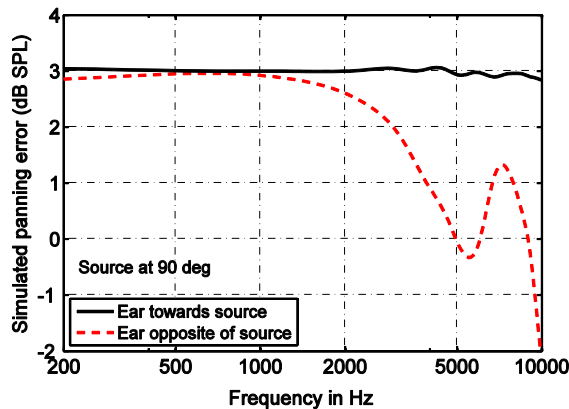


FIGURE 2. Panning error in dB SPL derived for each ear for 90° sound incidence. Error was computed with KEMAR-HRTFs as the difference in level when panning between two loudspeakers (LS1 and LS3) with 5° spacing relative to level when playing from a single loudspeaker (LS2) at the midpoint between the panning loudspeakers (LS1 and LS3).

PERCEPTUALLY EQUALIZED PANNING – THE PEP ALGORITHM

Equalization filters were derived for different source directions to spectrally compensate the panning error. The problem is that a single filter for both panning loudspeakers cannot correct the spectrum at both ears independently, but only a certain, unknown average spectrum. Three approaches to gain the filters and a perceptually-optimal method to average the spectral error across the ears were investigated:

1. The compensation filters were determined in a listening experiment in which the level of narrow band noise was found to yield equal loudness with and without panning. Seven normal hearing listeners (< 20 dB HL in 300Hz - 10 kHz) participated in the loudness comparison experiment. 1 Bark-wide white noises (500 ms duration, 60 dB SPL comparison) of various center frequencies (4.5/8.5/14.5/18.5/19.5/20.5/21.5 Bark) were used as the stimulus. Listening tests were done for directions 0, 45, 90, ..., 315°. The filter functions were computed from the level corresponding to the midpoints of the psychometric functions of response likelihood when the panned sound was perceived louder than the true source from LS2.
2. Panning errors were simulated with KEMAR-HRTFs (Gardner and Martin, 1994), see above. A generalized filter was obtained from smoothing across frequency and directions and large level differences were compressed.

3. Generalization was obtained by using HRTFs from a spherical head model which do not incorporate pinna effects (Duda and Martens, 1998). Filters were computed as for KEMAR-HRTFs.

Filters were implemented as zero-phase FIR filters of 21-61 taps at $f_s = 44.1$ kHz. Filtering was done identically to both panning loudspeakers. Loudspeaker signals thus differed only in level according to the sinusoidal panning law.

EVALUATION OF THE PEP ALGORITHM

Listening Test Methods

Five normal hearing listeners participated in each experiment to evaluate PEP-equalization filters. Stimuli were bursts of white noise (30ms duration, 70ms pause) with various upper cut-off frequencies (300Hz - 510/1080/2000/2700/3700/5300/7700/10000 Hz), Fastl noise (Fastl, 1987), speech babble, and the words 'shape' and 'wide' spoken by a female speaker and taken from the CASPA speech test (Mackersie *et al.*, 2001). Level was roved in 3 dB-steps within ± 3 dB from a base level of 60 dB(A) (55 dB(A) for the words). Subjects were asked to assess and rate the difference ("same"/"different") between two successive presentations of the stimuli where all combinations of panning (P) and no panning (N) were presented once (NN, NP, PN, PP). Tests were done at 24 center directions (every $45^\circ + [0, \pm 7.5]^\circ$) for panning over 7.5° and at further 8 directions for panning over 15° (not shown). An experiment consisted of (12 sounds)*(3 levels)*(32 directions)*(4 conditions) = 4608 trials in randomized order, broken into 36 runs. The experiment was done each for the no-equalization baseline and for the 3 filter generation approaches. Feedback was given on each trial and subjects were trained before each experiment.

Results

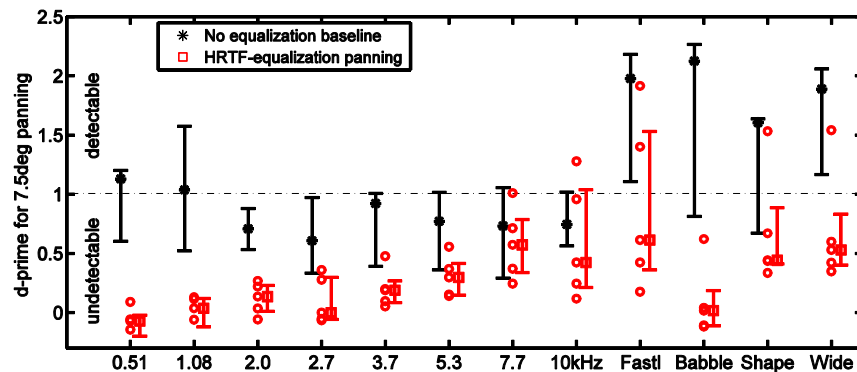


FIGURE 3. Detectability of differences between panned and direct sources without equalization of the panning errors (*, black) and with the PEP algorithm for panning equalization derived from KEMAR-HRTFs (\square , red). Medians and quartiles are given. Despite the direct A-B comparison between panned and true sources and despite using highly trained listeners, equalization with the PEP algorithm rendered panning errors undetectable in almost all conditions for almost all listeners.

Figure 3 shows results for the no-equalization baseline (*, black) and for equalization using filters derived from KEMAR-HRTFs (\square , red). The results were evaluated according to signal detection theory for a yes/no-task. In figure 3 responses were collapsed over levels, directions and listeners. Without equalization, panning errors were detectable for most sounds ($d' > 1$). However, detectability was smaller for all wide-band noises compared to the speech-related sounds. The reasons for this are not clear at present, since all stimuli were temporally modulated and wide-band noises contained more energy at high frequencies than the speech-related sounds.

Panning equalization with averaged KEMAR-HRTFs rendered panning errors inaudible for all but one listener who detected the audible differences in the panning condition with speech sounds and with wide-band stimuli when the upper frequency was 7.7 kHz or higher. Median d' , however, was clearly below detection threshold for all stimuli. The other two equalization approaches lead to similar results. This is particularly interesting for equalization on the basis of the spherical head model which does not make assumptions about pinna cues. Apparently, even

without equalizing for cues originating from the pinna panning errors were inaudible. PEP should thus be perceptually correct for people wearing hearing devices which place microphones outside the pinna.

An additional listening test evaluated if remaining panning errors stemmed from spectrotemporal or overall level effects (data not shown). Level roving between the first and the second comparison stimulus made panning errors undetectable for all subjects ($d' = 0$). This shows that subjects used minute level differences as the detection criterion. Timbral, temporal, or spatial cues could not be used, which is important for practical applications in the music industry.

A further test assessed the detection of panning errors in the primary sound in the presence of simulated room reflections. The introduction of reflections also rendered panning errors inaudible, thus demonstrating that perceptual equalization with the PEP algorithm is a viable approach even in real listening rooms with sound reflections.

CONCLUSIONS

An algorithm to equalize perceptual errors caused by playing coherent sounds from two nearby loudspeakers was presented and evaluated (PEP algorithm – Perceptually Equalized Panning). The PEP algorithm is lightweight to implement and sufficient to render panned sources from nearby speakers perceptually equivalent to real sources placed at the panned location. This is particularly important in practical applications where reflections are present where the algorithm likewise creates an accurate reproduction of timbral cues.

ACKNOWLEDGMENTS

We thankfully acknowledge the support by NIH RO1 DCD 00087.

REFERENCES

- Duda, R. O., and Martens, W. L. (1998). "Range dependence of the response of a spherical head model," *J. Acoust. Soc. Am.* **104**, 3048-3058.
- Fastl, H. (1987). "Ein Störgeräusch für die Sprachaudiometrie (A background noise for speech audiometry)," *Audiologische Akustik* **26**, 2-13.
- Gardner, B., and Martin, K. (1994). "HRTF Measurements of a KEMAR Dummy-Head Microphone," (MIT Media Lab).
- Griesinger, D. (2002). "Stereo and Surround Panning in Practice," in *Proceedings 112th AES Convention, Munich*, edited by B. C. J. Moore (Audio Eng. Soc.).
- Hafter, E., and Seeber, B. (2004). "The Simulated Open Field Environment for auditory localization research," in *Proc. ICA 2004, 18th Int. Congress on Acoustics, Kyoto, Japan, 4.-9.04.2004* (Int. Commission on Acoustics), pp. 3751-3754.
- Mackersie, C. L., Boothroyd, A., and Minniear, D. (2001). "Evaluation of the Computer-assisted Speech Perception Assessment Test (CASPA)," *J Am Acad Audiol* **12**, 390-396.
- Pulkki, V. (1997). "Virtual Sound Source Positioning Using Vector Base Amplitude Panning," *J. Audio Eng. Soc.* **45**, 456-466.
- Pulkki, V., and Karjalainen, M. (2001). "Localization of Amplitude-Panned Virtual Sources I: Stereophonic Panning," *J. Audio Eng. Soc.* **49**, 739-752.
- Seeber, B. U., Kerber, S., and Hafter, E. R. (2010). "A System to Simulate and Reproduce Audio-Visual Environments for Spatial Hearing Research," *Hearing Research* **260**, 1-10.
- Spors, S., and Ahrens, J. (2007). "Comparison of higher-order ambisonics and wave field synthesis with respect to spatial aliasing artifacts," in *Proc. ICA 2007, 19th Int. Congress on Acoustics, Madrid, Spain, 2.-7.09.2007*, edited by A. Calvo-Manzano, A. Pérez-López, and S. Santiago (Int. Commission on Acoustics).
- Vorländer, M. (2008). *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality* (335 pages, Springer, Berlin).