Perceptual equalization in near-speaker panning

Bernhard U. Seeber* and Ervin Hafter

Auditory Perception Lab, UC Berkeley, 3210 Tolman Hall #1650, Berkeley, CA 94720-1650

**now at: MRC Institute of Hearing Research, University Park, Nottingham, NG7 2RD, UK, bernhard@ihr.mrc.ac.uk*

1 Introduction

In recent years, extensive work has been done on spatial sound reproduction techniques involving multiple loudspeakers. Techniques like wavefield synthesis or cross-talk cancellation aim to simulate the soundfield at the listener's ear for sounds from different directions. In the simple approach of amplitude panning sounds are played simultaneously from two loudspeakers while variations in level between speakers alter perceived direction. Previous studies investigated the perceived direction (e.g. [5]), but we ask if panning can truly replace a loudspeaker at the panned location by not permitting audible panning errors. This requires not only that panned direction is accurate, but also that source width, loudness, timbre, and temporal aspects are equally reproduced. Obviously, for large loudspeaker spacings panning errors will be audible, but if both speakers are spaced only 7.5° or 15° apart, which is large compared to requirements for wavefield synthesis, errors might be very small. Informal listening convinced us that errors are nevertheless clearly audible. In the first part of the manuscript these errors are characterized, while the second suggests three methods for panning equalization. An in-depth evaluation of the equalization techniques found that all three approaches lead to similar results, that a filter derived from a spherical head model is sufficient to render panning errors inaudible, and that errors will be inaudible in all practical applications. The simplicity of the equalization approach ensures that results are also valid for listeners wearing hearing devices.

2 Panning assessment

Figure 1 shows the general layout for panning. Sounds are either played directly from the speaker in the middle or a virtual, panned source is heared at the center between both outer speakers for synchronous playback from them. Speakers were equalized in our setup in amplitude and phase which results in a 6 dB summation from both speakers at the listener's head position while the common sinusoidal panning law corrects for only 3 dB summation equal to intensity addition ($\sin(45°) = 0.707$) [3, 4]. Figure 2 shows the estimated panning error at the ears, i.e. when the head is placed in the soundfield. The 3 dB excess remains present at low frequencies while at high frequencies interference effects lead to pronounced differences between panned and direct sources as well as across ears. Listening tests revealed small perceptual differences, primarily in timbre. Equalization filters that can be applied identically to both loudspeaker signals might alleviate those errors.

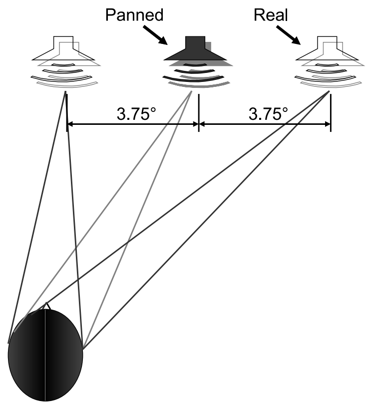

Figure 1: Sketch of the panning situation. Panning errors stem from interference effects and differences in HRTFs.

3 Panning equalization

Equalization filters were derived with three approaches:

1. Compensation filters were determined from a listening experiment in which the level of narrow band noise was found to yield equal loudness with and without panning. Seven normal hearing subjects (< 20 dB HL in 300 Hz - 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 $0, 45, 90, \ldots, 315°$. The filter functions were computed from the level at the psychometric function midpoints of the responses for panning louder/softer than the direct sound. Filters for alternate directions were interpolated.

2. Panning errors were simulated with KEMAR-HRTFs [2]. 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 [1]. 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.

4 Equalization evaluation – Methods

Five normal hearing subjects participated in each experiment to evaluate equalization filters. Stimuli were bursts of white noise (30 ms duration, 70 ms pause) with various upper cut-off frequencies (300 Hz

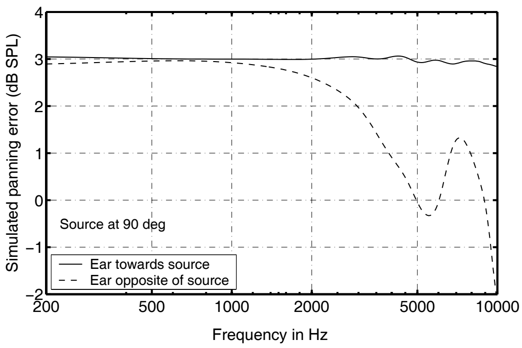

Figure 2: Panning error in dB SPL for each ear for 90 ° sound incidence. Error simulated using KEMAR-HRTFs as level difference for panning between two loudspeakers with 5 ° spacing relative to level from speaker at midpoint between the two speakers.

- 510/1080/2000/2700/3700/5300/7700/10000 Hz), Fastl noise, speech babble, and the CVCs 'shape' and 'wide'. Level was roved in 3 dB-steps within ±3 dB from a base level of 60 dB(A) (55 dB(A) for CVCs). Subjects had 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 $45° + [0, \pm 7.5]°$) 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 all 3 filters. Feedback was given on each trial and subjects were trained before each experiment.

5 Equalization evaluation – Results

Figure 3 shows results for the no-equalization baseline ($*$) and for equalization using filters derived from KEMAR-HRTFs (□). Evaluation was done using signal detection theory for a yes/no-task. In figure 3 responses were collapsed over levels, directions and subjects. 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 subject who detected panning for speech sounds and for wide-band stimuli with an upper frequency ≥ 7.7 kHz. 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. Panning equalization is

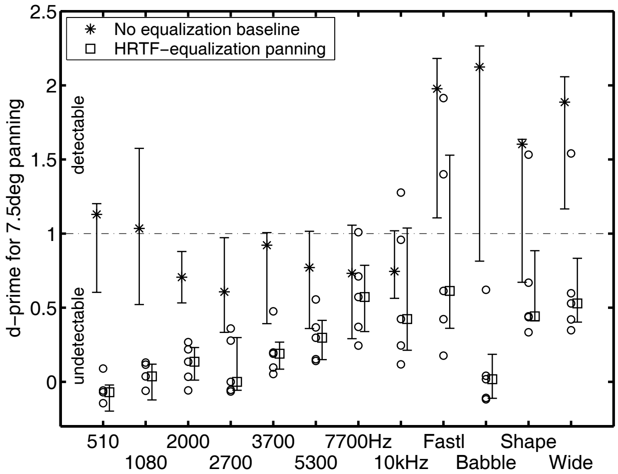

Figure 3: Detectability of differences between panned and direct sources without equalization ($*$) and with panning equalization derived from KEMAR-HRTFs (□). Medians and quartiles are given. Equalization renders panning errors undetectable.

thus valid for subjects wearing hearing devices with microphones placed outside the pinna.

An additional listening test addressed if remaining panning errors stemmed from spectrotemporal or overall level effects. 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 or temporal cues could not be used, which is important for practical applications in the music industry.

A further test assessed detection of panning errors of the primary sound in the presence of simulated room reflections. The introduction of reflections rendered panning errors inaudible.

6 Conclusions

Simple equalization is sufficient to render panned sources from nearby speakers perceptually equivalent to real sources placed at the panned location. This is particularly true for practical applications where reflections are present. Accurate reproduction of timbral cues can be obtained.

We thankfully acknowledge the support by NIH RO1 DCD 00087 and the MRC (BS).

References

[1] R.O. Duda and W.L. Martens. *J. Acoust. Soc. Am.*, 104(5):3048–3058, 1998.

[2] B. Gardner and K. Martin. Technical Report 280, MIT Media Lab, May 1994.

[3] E. Hafter and B. Seeber. *Proc. 18th ICA, Kyoto*, vol. V, pages 3751–3754, 2004.

[4] V. Pulkki. *J. Audio Eng. Soc.*, 45(6):456–466, 1997.

[5] V. Pulkki and M. Karjalainen. *J. Audio Eng. Soc.*, 49(9):739–752, 2001.