A NEW METHOD FOR ELEVATION PANNING REDUCING THE SIZE OF THE RESULTING AUDITORY EVENTS

PACS: 43.38.Md

Reinhard Gretzki (1), Andreas Silzle (2)
Institute for Communication Acoustics, Ruhr-University of Bochum, Germany
Ruhr-University of Bochum, Institute for Communication Acoustics
Universitätsstr. 150 / IC 1
44780 Bochum
Germany
Tel.: (2): +49 234 322 2490
Fax: +49 234 321 4165
E-Mail: (1): anReinhard@web.de, (2): silzle@ika.rub.de

ABSTRACT

Panning between two vertically arranged loudspeakers results in an auditory event that can cover almost the whole spanned angle. Additionally, these vertically arranged loudspeakers cause audible comb filtering if listeners are not equally spaced from each loudspeaker (some centimetres displacement is sufficient). Usually, these effects are reduced by increasing the number of loudspeakers and optimizing signal delays for one listening point. The present method diminishes the spatial extent of the auditory event to a great degree. Also, the comb filter effects occurring with misplaced/moving listeners are equally reduced.

RESUMEN

Paneo estéreo entre dos altavoces colocados verticalmente da lugar a una fuente fantasma que puede cubrir casi todo el ángulo entre los altavoces. Además, estos altavoces colocados verticalmente causan un filtrado tipo “peine” percibible si los oyentes no se encuentran situados a la misma distancia de cada altavoz (pocos centímetros de diferencia es suficiente). Normalmente, estos efectos son reducidos incrementando el número de altavoces y optimizando los retardos en una posición definida del escucha. El presente método disminuye en gran medida el espacio ocupado por la fuente fantasma. Así mismo, los efectos del filtrado tipo “peine” son igualmente reducidos.

INTRODUCTION

Different reproduction methods are used for rendering audible 3D sound fields over loudspeakers, e.g. Ambisonics, Wave Field Synthesis or Amplitude Panning. Here, we investigate Amplitude Panning for elevation reproduction. Elevation localisation of an auditory event is known to be highly listener specific, depending on the listeners’ pinnae and presented sound signals, as has been shown by various studies, see e.g. [1, pp. 44,104,310,312; 2]. It was also reported several times that panning between a pair of vertically arranged loudspeakers results in an auditory event that can cover almost the whole spanned angle, see [2,3,4]. Confronted by these problems, installations with loudspeakers for height reproduction incorporated until now either a high number of loudspeakers to keep the spatial extent of the resulting auditory event small, or they did not address the problem at all. The method presented in this paper reduces the spatial extension of the auditory event to a great degree and thus reduces the number of loudspeakers necessary to evoke a defined spatial extent.

Additionally, two vertically arranged loudspeakers give rise to audible comb filtering if the listener is not equally spaced from each loudspeaker. The effect will already arise when the listener displaces himself vertically by some centimetres, as preliminary listening tests with two expert listeners at our institute showed. For the application in view, the elevation panning algorithm should be applicable to an extended listener area and therefore this problem is also addressed.

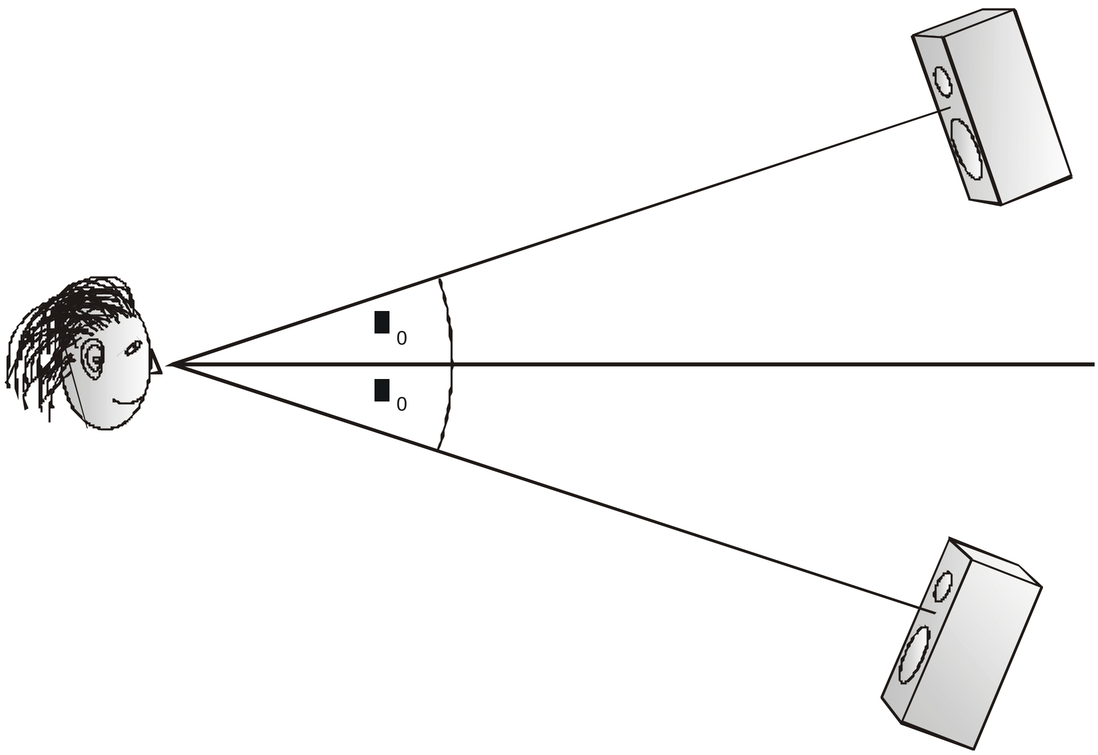

Figure 1: Listening position in preliminary tests, note the symmetric position of the speakers with regard to the listener

ELEVATION LOCALISATION

The cues used by humans to form auditory events in the horizontal plane are the interaural level and time differences as well as the frequency dependent filtering by the pinna. To form auditory events in the median plane, humans mainly use the pinna filtering. Blauert reports in [1, p.106] data from a localisation experiment with sine pulses made by Roffler/Buttler in 1968. They found out that the angle of elevation to which a sine pulse is localized to does not depend on the angle of elevation it is presented from. Rather, the frequency of the sine pulse specifies the elevation angle that listeners will perceive the pulse to be coming from. Lower frequencies are perceived below and higher frequencies above the 0° elevation angle. This effect has also been recognized in preliminary listening tests at our institute with different types of signals such as frequency filtered pink noise or music. Recently, the effect has been reported in [4].

In [1, p.310], Blauert also reports data from Wettschurek's experiments (1971) concerning the localisation blur of auditory events along the median plane. The localisation blur is defined as 50% of the localisation answers around the mean of all answers in a localisation experiment. So it does not describe the spatial extent of an auditory event, however it constitutes a measure of the interindividual differences in localisation which the tested subjects perceived. According to Wettschurek, the blur increases to about the double when 4 kHz lowpass filtered noise instead of broadband noise is presented. Keeping in mind that the outer ear is responsible for elevation localisation, see e.g. [1, p.307ff and 5, 6], this is comprehensible. The lower frequency limit, in which pinna cues are used, is known to be at about 5 kHz, see [5, 6]. Therefore it is clear that the blur increases when 4 kHz lowpass filtered signals are presented, as most of the cues one uses for elevation perception are then filtered out. The listeners have to rely more often on guesses and normally lesser relevant cues such as shoulder and torso reflections.

When two vertically arranged loudspeakers, as in figure 1, are used for sound reproduction, the pinna cues above 5 kHz should allow the listener to perceive the exact elevation position of each loudspeaker. When the two loudspeakers radiate the same signal (as in panning), this leads to an auditory event that covers almost the whole spanned angle (see [2,3,4]) and not to a clearly localizable auditory event (also called “phantom source”) as in horizontal stereo panning. One might expect the auditory event to be smaller with low pass filtered signals because of the absence of the strong elevation localisation cues. The authors could recreate this effect in preliminary tests. A pair of loudspeakers was installed twice in two rooms (anechoic room and reproduction room). The opening angle 2?f $_0$ between the loudspeakers was 50°. The listener was seated 2m in front of the loudspeakers (see figure 1 for the placement of the listener with respect to the loudspeakers). The tests showed that the more energy above 5 kHz is present in the test signals the bigger the spatial extent of the auditory event will be. Further testing with 4 test subjects who were presented with signals that were filtered to have energy in just one band between 0Hz and 5kHz showed that it is possible to reduce the spatial extent of the auditory event by panning different bands in-between 0Hz and 5khz independently, according to the elevation perception that test listeners reported when listening to each band. The tests also showed, that the elevation perception for the same panning is dependent on the room to a certain degree (anechoic room or reproduction room), but the basic effects will remain. When 5 kHz lowpass filtered signals are used, the correct bands and the correct panning gains for a certain room lead to a great reduction of the spatial extent of the auditory event. Audio reproduction not only uses 5 kHz lowpass filtered signals. Therefore, another way of reducing the spatial extent of the auditory event that works mainly for frequencies greater than 5 kHz had to be found.

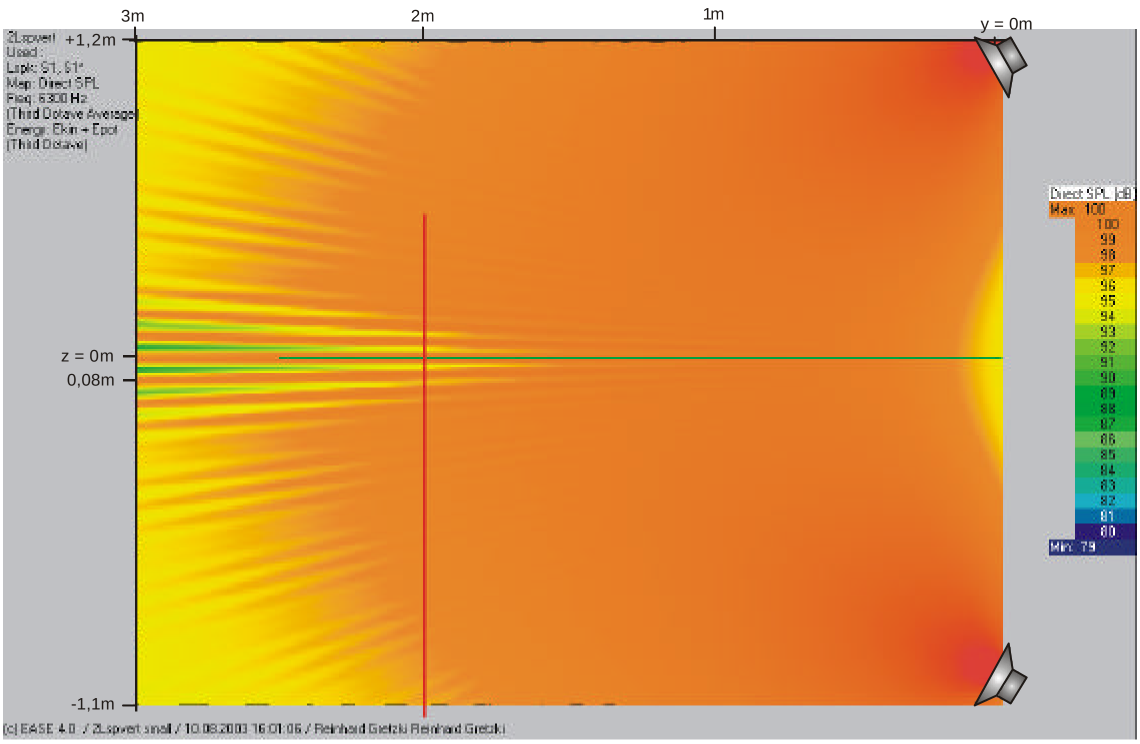

Figure 2: Comb filter effect caused by two elevated loudspeakers (at the right side of the figure), 6300Hz third octave band shown, red/green cursor points to listening position, differences between minima/maxima between 7 and 10 dB, distance between two maxima smaller than 8cm, see z axis on the left side of the figure.

COMB FILTER EFFECTS WITH TWO VERTICALLY ARRANGED LOUDSPEAKERS

Preliminary tests by the authors showed that two *vertically* arranged loudspeakers cause audible comb filtering. The tests used the same arrangement as the localisation tests (see above). Already a vertical displacement by 2 or 3 centimetres out of the symmetry point (point at which the loudspeaker above and below has the same absolute elevation) will change the comb filtering and thus make its effect on the listener more pronounced. Note that there is nearly no such effect with horizontally placed loudspeakers. See figure 2 for a plot of a simulation with the electroacoustic simulation program "EASE". The dimensions are those from the listening test setup. The simulation was only made for the direct sound, so the differences between maxima and minima are overestimated for a real room, but are printed here as an overview. Note that EASE has a minimal resolution of 1cm for mappings as the one shown here. This leads to the impression that the comb filter effects are smaller near the loudspeakers. As measured in our reproduction room, these differences amount to about 5dB for third octave bands higher than 5kHz, with no strong differences of the effect in the area 1m to 3m in front of the speakers. The smaller difference between the minima and maxima is the effect of the room reflections smoothing out the sharp interference effects.

For signals with energy up to 5 kHz, the perceptually based panning method, which is proposed here, involves independent panning of frequency bands, according to the elevation perception that test listeners reported when listening to each band. As a result, the energy of one band is mostly located at, for example, the lower loudspeaker whereas the energy of another band might be more concentrated at the upper one. In that way the energy of different bands is spatially separated and can no longer interfere. The comb filter effects in the frequency bands up to 5 kHz are thus reduced.

In order to reduce the comb filter effects also in frequency bands higher than 5 kHz, a special filtering method was developed. In the listening tests we found a very inconvenient effect of the comb filtering at higher frequencies: when the listener positions himself in a interference notch, he mostly perceives the upper and lower speaker and a "hole" in-between. To compensate for this, a special ladder filter structure was chosen, which precombs the signals and inverts them. As preliminary listening tests by the two authors showed, this filtering method leads to a perception of not only reduction of the comb filter effects but also reduction of the spatial extent of the auditory event.

The special filtering method together with the aforementioned panning method for signals with frequency up to 5 kHz results in a remarkable reduction of both the spatial extent of the auditory event and the comb filter effects as listening tests with 3 expert listeners showed. It remains to be investigated if the same result can be achieved with decorrelated signals. Early testing seems to show that the decorrelated signals reduce the comb filter effect to an equal amount, but by means of a greater computational burden.

ROOM INTERACTION EFFECTS

When conventional cone loudspeakers are placed near to a boundary such as a wall or the floor, the lower frequency response is amplified. To avoid this, one normally makes use of compensation filters ('boundary compensation') which are incorporated for example in many active studio monitor speakers. The panning method has to take this into account. The strength of the comb filter effects that were previously mentioned is dependent on the loudspeaker directivity and the absorption factors of the near boundaries. These effects partly explain the differences between the test setup in the anechoic chamber and the reproduction room.

SIGNAL PROCESSING

The method consists in panning three or more frequency bands individually, based on the elevation perception that expert test listeners report, and additionally performing a special filtering of higher frequencies, as explained above. Figure 3 gives an overview of the signal processing. The audio signal is first filtered into bands by a linear phase filterbank. The phase linearity is necessary, because the bands are later to be summed after the processing. Non-linear-phase would lead to cancellations of frequency bands. Actually, filters with equal phase response for the filterbank would suffice, but linear phase is easy to achieve by means of FIR filters. The different bands are then filtered as necessary. The signals of each filtered band are multiplied by their respective gain.

$$\frac{\tan j}{\tan j_0} = \frac{g_{high} - g_{low}}{g_{high} + g_{low}}; \qquad g_{norm,i} = \frac{g_i}{\sqrt{g_{high}^{\,2} + g_{low}^{\,2}}} \qquad \textit{(equation 1)}$$

The gains for higher bands are directly derived from the tangent panning law (see left side of eq.1). This law was formulated for horizontal stereo panning, see [2a]. The gains (here “high” and “low”) are submitted to a normalization as in the right side of equation 1. The panning gains for lower and middle frequencies are also derived from this law, but additionally trimmed according to the elevation perception of expert test listeners. This is achieved by linear interpolation between the adjustments test listeners make when asked to adjust the panning in order to move the elevation of their auditory event to a specific target position. They adjust the degree value for a tangent panning law when asked to move their percept to a low, middle or high position while listening to one specific frequency band. This yields panning values for each band and the 3 positions low, mid, high. Finally, the frequency bands are summed to signal pairs. When more then two loudspeakers are used for elevation reproduction, it is necessary to firstly select the pair which is used for a certain elevation and secondly to select which signal path will feed the upper of the two loudspeakers and which the lower.

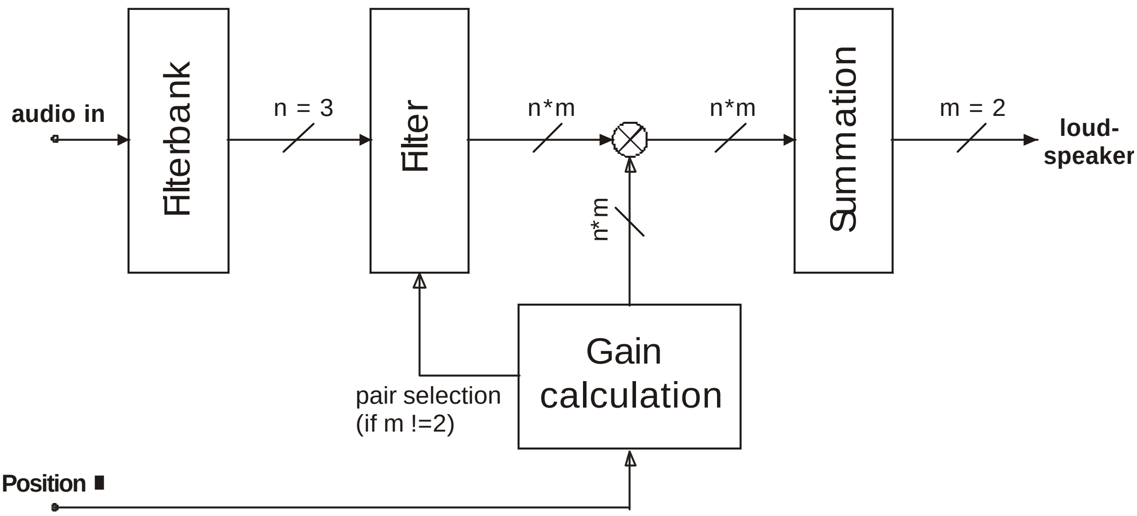

Figure 3: Signal flow graph of elevation panning method, n: number of frequency bands, m: number of elevated loudspeakers for one azimuth direction

IMPLEMENTATION OF THE METHOD

The new elevation panning method was integrated with an already known amplitude panning method developed by Ville Pulkki: VBAP/MDAP. The VBAP (“Vector Base Amplitude Panning”) algorithm is explained in [2a]. It is basically a vector formulation of the tangent law. A loudspeaker setup is divided into pairs (2D) or triangles (3D). Each pair or triangle is used for panning in the area it spans. See equation 2 for its formulation for one pair of loudspeakers in a 2D setup. g_1, g_2 are the yet to be normalized gains for one loudspeaker pair, $l_{i,x}$ and $l_{i,y}$ are the coordinates of loudspeaker i=1,2. p_x and p_y are the coordinates of the direction to which the sound is panned to. The gains are normalized as in equation 1 by the standard vector norm. Because of its vector formulation, VBAP is easy to use with more than 2 loudspeakers, i.e. in

3D setups. VBAP uses one loudspeaker alone, when the panning direction coincides with a loudspeaker direction. This behaviour generates the sharpest virtual source possible. When moving sources are generated, this causes the listener to perceive the actual loudspeaker directions, because the sound is radiated by 3, 2 or just 1 loudspeaker, depending on the panning direction. This defect is prevented by applying MDAP ("Multiple Direction Amplitude Panning"), which specifies multiple directions for the panning calculation. Each direction yields a gain vector. The resulting gains are summed up. By specifying the correct spread angle between the multiple directions, a virtual source always uses the same number of loudspeakers, whether the panning direction equals a loudspeaker direction or not. This prevents the effect of perception of the loudspeaker directions with moving sources. For more details see [2c].

$$\vec{g}_{unnorm} = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} = \begin{bmatrix} l_{1,x} & l_{2,x} \\ l_{1,y} & l_{2,y} \end{bmatrix} \begin{bmatrix} p_x \\ p_y \end{bmatrix}; \qquad \vec{g}_{norm} = \left\| \vec{g}_{unnorm} \right\| \qquad \textit{(equation 2)}$$

The current implementation of the elevation panning method uses between 250 and 550 floating point operations per source, depending on the quality of the filterbank implementation.

SUMMARY

Elevation perception of auditory events, evoked by employing amplitude panning over two vertically arranged loudspeakers, suffers from two main defects. The spatial extent of the auditory event can cover nearly the whole opening angle between the loudspeakers. This can be reduced by applying the proposed panning algorithm which consists of different panning laws on sub-bands. Additionally, comb filter effects are caused by the two loudspeakers radiating the same signal. This effect is much more audible with two vertically arranged loudspeakers than with the standard stereo setup in the horizontal plane. The comb filtering can also be reduced by the new panning method. Additional filtering can reduce the annoyance of this effect further. The presented panning method is subject of a patent application.

References

[1] "Spatial Hearing, The Psychophysics of Human Sound Localisation", Blauert, J., 1997, MIT Press

[2] "Localization of amplitude-panned virtual sources II: 2 and 3-dimensional panning", Pulkki, V., JAES vol.49(9), 2001

[2a] "Virtual Sound Source Positioning Using Vector Base Amplitude Panning", Pulkki, V., JAES vol.45(6), 1997

[2c] "Uniform spreading of amplitude panned virtual sources", Pulkki, Proceedings of the IEEE Mohonk workshop, 1999

[3] "Untersuchungen zur Summenlokalisation in der Medianebene", Behrens, T., presented at DAGA 1994, Dresden, Germany

[4] "Elevation perception: Phantom images in the vertical hemi-sphere", Barbour, J.L., 24th AES int. conf. on Multichannel Audio, 2003, Banff, Canada

[5] "A structural model for binaural sound synthesis", Brown/Duda, IEEE transactions on speech and audio processing, vol.6(5), 1998

[6] "Elevation localization and head-related transfer function analysis at low frequencies", Algazi/Avendano/Duda, JASA vol.109(3), 2001