AUTOMATIC MIXING: LIVE DOWNMIXING STEREO PANNER

Enrique Perez Gonzalez and Joshua Reiss

Centre for Digital Music,
Queen Mary University of London, Electronic Engineering,
Mile End Road, E1 4NS
London, United Kingdom
enrique.perez@elec.qmul.ac.uk
josh.reiss@elec.qmul.ac.uk

ABSTRACT

An automatic stereo panning algorithm intended for live multi-track downmixing has been researched. The algorithm uses spectral analysis to determine the panning position of sources. The method uses filter bank quantitative channel dependence, priority channel architecture and constrained rules to assign panning criteria. The algorithm attempts to minimize spectral masking by allocating similar spectra to different panning spaces. The algorithm has been implemented; results on its convergence, automatic panning space allocation, and left-right inter-channel phase relationship are presented.

1. INTRODUCTION

An audio engineer carefully handcrafts the characteristics of multiple inputs to downmix it into a constrained number of channels. Creating a mix involves numerous spectral and gain processing as well as the use of several audio effects. This research explores the automatisation of one of these processes. The spatial effect, which has been investigated, is a stereo panner algorithm. The panner under study downmixes K inputs and converts them into a stereo mix. This automatic stereo panner makes panning decisions based on constrained spectral rules as well as priority criteria. The algorithm has also been optimized for live downmixing situations where a second take is not an option.

The first automatic processing for live sound environments for mixing applications can be traced to Dugan's automatic microphone mixer [1, 2]. Dugan set the basic principles of automatic gain adjustment for automatic mixing. This type of mixer was able to maintain constant gain structure regardless of the number of active microphone inputs. This mixer was completely analog and based its decisions on time domain gain compensation. Several years later a mechanical approach based on directive sensitive gating for automatic mixing was developed by Julstrom [3]. Currently, in the authors' knowledge, no frequency domain approach to automatic mixing for live applications has been proposed. With current DSP processing power and the expanding availability of fully automated digital consoles, it should be possible to develop automatic processes for automatic mixing in an easy and cost-effective manner. Automatic mixing can prove useful in live mixing for video games, live concerts and post-production.

For the purpose of this research it is important to make a distinction between *automatic* mixing processes and *automated* mixing processes. An automatic process involves an autonomous process. This autonomous process can be treated as a constrained rule problem in which the design of the control rules determines the process to be applied to the input signals. The automated process, on the other hand, is the result of playing back in sequence a series of user recorded actions. This involves playing back previously recorded and stored actions, regardless of whether automatically or manually generated.

A common task in live mixing is downmixing a series of mono inputs into a two channel stereo mix. For doing this the input channels get summed into a Left (L) and a Right (R) channel. The proportion at which these multiple mono inputs are added to each L and R channels are responsible for the perceived stereo image. Previous related work on downmixing for spatial audio coding, from 5.1 surround to 2.0 stereo, has been attempted by [4]. Processing of multiple channels for real time applications using priority has been attempted by [5], but this method requires an off-line processing stage which requires pre-processing of the audio channel in order to enhance them with descriptors. This method is suitable for game and simulations but are not optimal for live environments where the signal nature is unknown. Work on upmixing has been researched by [6, 7]. In their work, they describe methods to turn a stereo downmix into a multi-channel upmix. Although these methods can prove useful if backtracked, they are more suitable for multi-channel surround format conversion rather than for multiple input mixing. By multiple input channels we refer to the individual instruments of a live group of musicians or multiple speech inputs as opposed to multi-channel format submixes, as contained in 5.1 surround formats. In the knowledge of the authors, no current approach to stereo downmixing multiple inputs channels in a live environment has been attempted.

Other relevant related work includes the idea of on-the-fly-mixing. On the fly multi-track mixing has as a central idea to maintain the intentions of the composer and sound engineer while providing the final user with some degree of control [8]. This system has the intention of enhancing the end user experience by providing him with controllable parameters, which have been constrained in order to keep even intention for non-expert user. The system proposed in this research differs in the idea that it searches an automatic approach to downmixing by reducing user interaction. The proposed system is to be seen as a helper to the sound engineer and composer rather than giving the engineer a new set of extra constraints parameters to be manipulated. The proposed approach seeks to enhance the user experience by automatically downmixing the input sources while reducing or eliminating the mixing tasks to the user. In particular, this paper explores an algorithm for downmixing using an automatic stereo panner.

2. AUTOMATIC PANNING

The panner is based on channel priority, the algorithm was implemented so that the user connects the most important input channels to the first channels of the downmixer and the least important channel inputs to the last channels of the downmixer. However, instrument recognition techniques can be used to implement an automatic version of channel priority. In either case channel priority gives the algorithm data, which can be used to help determine, with better accuracy, the panning positions. For example if the lead vocal is the most important channel in the mix, in the current implementation of the algorithm, it should be located on channel one. The algorithm will treat the highest priority channel as a source, which should be as centered as possible. The algorithm also is based on the assumption that panning channels which have similar spectral content to opposite sides improves intelligibility by reducing spectral masking. For the proposed algorithm the panning space is determined by the number of channel to be mixed and the available panning steps. Finally the algorithm has been set so that the panning step is dependent on the spectral resolution of the analysis. The algorithm has been optimized for MIDI control.

2.1. Design architecture

One of the most important design considerations for any automatic live sound process is the fact that input sources should be reproduced at all times. Missing just a few samples can mean losing important content of the live audio program to be heard. Figure 1 shows the general diagram of the proposed architecture that solves this problem.

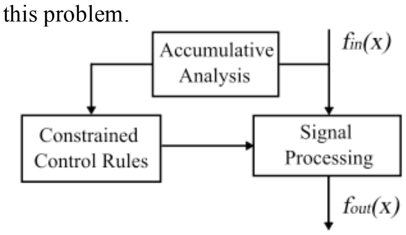

Figure 1: Diagrams of side processing algorithms for automatic live panning application.

Other important consideration is that, because of the live nature of the input signal an instantaneous decision based on the current sample can prove wrong. For this reason an accumulative knowledge based approach can prove more reliable. Because of this a side-chain accumulative analysis together with a constrained control rule structure is used to calculate the control parameters send to the desired signal processing procedure. This means that the input signal is free to reach the output even though the analysis and processing decisions have not been determined. Once the analysis and rules calculations have been finalized the processing can take place. This design gives the opportunity to the algorithm to constantly get updated by using the input data against past-accumulated data. This can influence the processing stage by using previous knowledge instead of performing decisions based on instantaneous transients, which could result in generating unwanted signal processing artifacts.

2.2. The filter bank analysis

The proposed system uses a filter bank as means of doing spectral analysis. The filter bank does not affect the output signal as it is only used for analysis. The filter bank uses first order band pass filters. The filters inside the bank are determined by the bandwidth contained between the -3dB low cut off and -3dB high cut off points of the filter. Each consecutive filter in the filter bank has double the bandwidth of the previous filter. The total number of filters inside the bank (K) is equal to the number of input channels to be mixed. This will result in an adaptive frequency analysis resolution, which is dependant on the number of input channels.

To design the channel dependent filter bank the following equations are provided.

$$Bw(k) = \frac{F_{\max}}{40K} 2^{k-1} \tag{1}$$

Where $Bw(k)$ correspond to the bandwidth of the k^{th} filter contained on the filter bank where k goes from 1 to K and K corresponds to the total number of input channels to be mixed, which is equal to the total number of filters in the filter bank. $Fmax$ is the maximum frequency to be reproduced and the constant 40 was chosen simply to give filter bandwidths at simple rational intervals.

$$LCF(k) = F_{\min}.(2)^{k-1} \tag{2}$$

Equation 2 calculates the Low Cut Frequency (LCF) for the k^{th} filter. Where k goes from 1 to K. $Fmin$ corresponds to the left -3dB cut off point of the first filter, in the current implementation $Fmin$ has a value of 20Hz. 20Hz has been chosen because it is widely accepted as the lower frequency limit of human hearing. To calculate the High Cut Frequency (HCF) for the k^{th} filter the corresponding $LCF(k)$ and $Bw(k)$ must be added together. Where k goes from 1 to K. Because K is equal to the total number of input channels to be mixed, we can say that k corresponds to the individual filter identifier number. The HCF corresponds to the right -3dB cut off point of the filter.

Once the filter bank has been designed the algorithm uses the spectral, band limited, information within each filter to obtain the absolute peak amplitude for each filter. The peak amplitude is measured within a 100ms window. Because noise contained in the signal of interest may trigger undesired readings a second threshold peak meter is used to gate the peak readings of each filter. During development it was found that the optimal threshold value for this application is -60dB, with a 10ms window. For every occurrence of the peak amplitude measurement that has is successful in passing through the gate, the following algorithm is performed.

First all K filters are scanned in search of the highest amplitude. Second all filters are searched with the aim to identify the k filter responsible for having the highest amplitude. Once this is done the result is accumulated in a register corresponding to the k^{th} filter responsible for the highest peak amplitude. Third the K accumulated registers are scanned in search of the k accumulator responsible for the highest number of occurrences. Finally the system outputs the k filter identifier number corresponding to the biggest accumulated spectral band of the input signal. This process is repeated in a continuous manner for every peak amplitude value received after the gating stage. This approach uses digital logic operations of comparison, equity and accumulation only, which makes it highly attractive for an efficient digital implementation. The block diagram of the bank filter analysis algorithm has been sketched in Figure 2.

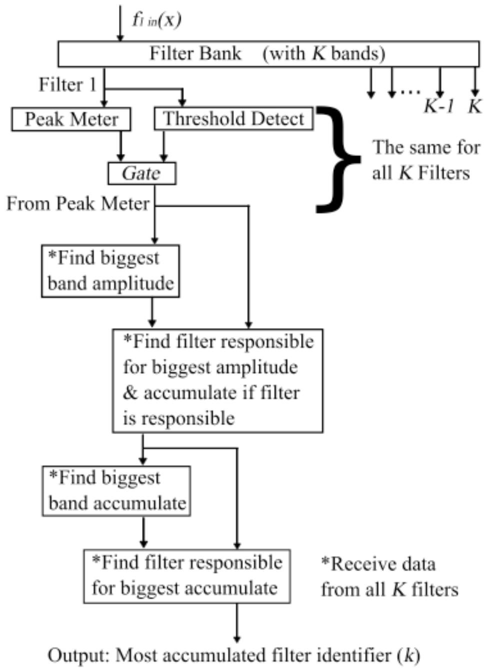

Figure 2: *Analysis block diagram for one input channel for a Filter bank of K length.*

2.3. The constrained control rules

The first rule consists of maintaining all sources whose main energy is contained in the lowest filter banks un-panned. There are two main reasons for doing this. First this ensures that the Low frequency content remains evenly distributed within speakers[9]. This ensures that reproduction of low frequencies, which are more likely to produce audible distortion at high levels, remains spread over two sources to reduce distortion. Second there is no point on panning low frequency sources below 200Hz because we fail to localize them properly [10]. It is thought that we are unable to localize properly low frequencies because its wavelength is so long that our left and right ears perceive them as coming simultaneously from the same place. It is also thought that because the head is unable to absorb the low frequencies, it is hard to localize them, in contrast, high frequencies do get some attenuation between ears which we associate with position.[11] For this reason all input signals, with accumulated energy contained in a filter with a HCF below 200Hz are not panned, and should remain centered at all times.

The second rule decides the available panning step *(PS)* a positioning rule based on equidistant spacing. The rule uses 2 parameters to determine its result. First the Integer location of the filter that contains the maximum number of accumulation *(k)*. Second the total number of occurrences of signals containing accumulated energy in the same k^{th} filter (R_k).

This PS has to be calculated for every different accumulated k^{th} filter. This means that for k^{th} filters which have not reached maximum accumulation there is no need to calculate the PS, this makes the algorithm less computationnally expensive. If only one repetition exists for a given k^{th} filter (R_k=*1*) the system is default to pan the input to the center (64 or 63 in the MIDI case). The system default initial value is to have all channels centered.

In order to obtain all the available panning space locations for R_k bigger than one equation 3 is provided.

$$PS_k(i) = RND[\frac{(i-1)\max PS}{R_k - 1}] \tag{3}$$

Where $PS_k(i)$ is the ith available panning step and i has a range from 1 to R_k. i and k have a range from 1 to K. Because the algorithm has been optimized for MIDI control the *maxPS* is 127, and the rounding function *RND* is to be used to obtain the proper discrete MIDI *PS*.

The third rule is priority-panning assignment. Priority is based on priority numbers (Pr_k) assigned to all R_k. In order to assign the panning priority first we must calculate PS with equation 3. An alternating priority algorithm has been used. This priority assignation works as follows; once K has been obtain by the algorithm, the system scans all channels to obtain the number of input channel sources which share the same accumulated filter R_k. Then the system proceeds to assign the priority channel number Pr_k from left to right. The Pr_k is done from left to right because it goes in accordance with the idea that the channel closest to the first mixer channel has more priority than the last one. This ensures that the channels closer to the first physical channels remain as centered as possible. This approach reduces spectral masking by separating inputs with similar spectral content as far as possible from each other. Since the implementation of this algorithm was intended for compatibility with MIDI controlled automated mixers the panning step has a range from 0 to 127 steps. Where centre is 64 or 63 and *maxPS*=127. In Table 1 it is presented all possible PS for an 8CH automatic panning system.

PS(i)	Pr_1	Pr_2	Pr_3	Pr_4	Pr_5	Pr_6	Pr_7	Pr_8
PS(1)	64		-	-	-	-	-	-
PS(2)	0	127	-	-	-	-	-	-
PS(3)	64	0	127	-	-	-	-	-
PS(4)	43	84	0	127	-	-	-	-
PS(5)	64	32	95	0	127	-	-	-
PS(6)	51	76	25	102	0	127	-	-
PS(7)	64	42	85	0	21	106	127	-
PS(8)	54	73	36	91	18	109	0	127

Table 1: *Discrete panning rule by implementing alternate priority*

All values have been obtained using equation 3 and by implementing alternate priority. It is important to notice that the maximum value of i is equal to the number of channels of the mixer, and the same applies for the maximum assignable priority numbers. Based on this, it can be stated that the total panning space can be calculated as the number of panning steps available times the filter bank filter which HCF are above 200Hz. It is also important to realise that thanks to this channel dependency the algorithm will update itself every time a new input is detected in a new channel. The algorithm will also update itself if the spectral content of a input channel suffers from a drastic change over time of spectral content. The block diagram containing the constrained decision control rule stage of the algorithm is presented in Figure 3.

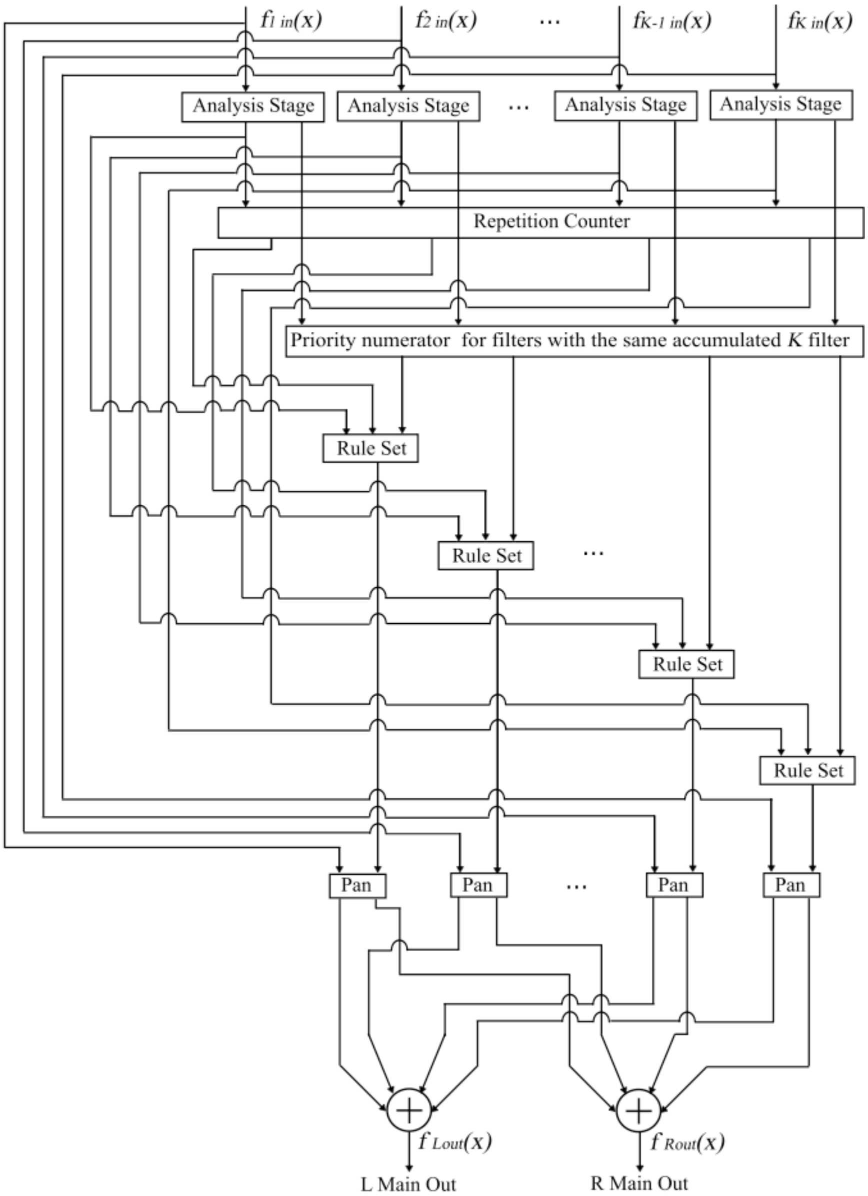

Figure 3: *Block diagram of Automatic Panner algorithm constrain control rules algorithm for K input Channels*

2.4. The panning processing

The panning architecture used consists of a -3dB panning law. This means that when panning is centred the left (L) and right (R) channel have a -3dB gain. This is so, that when L and R channels are in phase and added, a 0dB gain is achieved. When the panner goes all the way to the left the L channel has a normalized gain of 0dB and the R channel has a gain of -∞, and the inverse when the system is panned to the R channel. This approach keeps a constant 0dB gain regardless of the panning angle. The panning angle is the degree measurement of rotation of a pan-pot. The panning step is the digital analogy used for MIDI implementation of a panning angle. The panning step is related to the signal gain multiplier. The panning algorithm uses the following equations for processing the right and left channel.

$$f_{Rout}(x) = (PS_k(i) / \max PS).f_{in}(x) \quad (4)$$

$$f_{Lout}(x) = [1 - (PS_k(i) / \max PS)].f_{in}(x) \quad (5)$$

The term multiplying $f_{in}(x)$ in equations 4 and 5 have a range from 0 to 1 and it is called the Panning Factor (PF). It has a maximum range of 1 In order to maintain the resulting signal normalized. The PF of one channel is complementary to the PF of the other. For this reason the sum of the squares of boat PF is always 0dB, making the overall amplitude equal to the original amplitude of $f_{in}(x)$ regardless of the panning position.

An interpolation algorithm has been coded into the panner to avoid rapid changes of signal level. The interpolator has a 22ms fade-in and fade-out, which ensures a smooth natural transition when the panning control step is changed.

3. RESULTS

Several sinusoidal test signals and music tracks simulating a live playing band were used as a mean to test the automatic panning algorithm. The multi-track data used was obtained from the BASS-dB database [12]. BASS-dB is the Blind Audio Source Separation evaluation database; it contains links to multi-track recordings which license allows modification and redistribution of the data for non-commercial purposes.

In all studied cases the algorithm was able to converge. In Figure 4 we can see the convergence for 4 different sources. The 4 sources were selected from a set of measurements obtained from an 8ch automatic panning downmixer. The plot shows the panning factor as it approaches stable state, as applied to the input signal.

The dotted line, in Figure 4, is the result of plotting the PF of one of 4 tracks that have similar spectral content; the algorithm has spread all four signals equidistantly. The dotted line corresponds to the highest priority channel out of the four tracks. The dotted line has converged into a panning factor equal to 0.33858 or a MIDI panning step of 43. The other 3 sources not shown in this plot converged in accordance to Table 1 for a *PS(4)*.

It is important to notice that the speed at which the algorithm converges is dependant on the spectral content of the overall input channels. Also a track containing similar spectral content which start later in time than others can cause a panning space re-assignation. Others convergence values where the source has been panned fully to the sides have been also plotted (PF=0 and PF =1). Finally the solid line represents a drum kit signal, which although the algorithm struggles to decide wheatear its spectral content is of mainly high frequencies (due to the hi-hat) or low frequencies (due to the kick drum) it manages to converge into central position (PF=0.5), which is technically the most convenient panning position for a signal containing very low frequencies.

In Figure 5 we can see the panning factor depicted as a solid line in Figure 4 superimposed on the MIDI panning step calculated by the algorithm. The MIDI panning step is one stage before the 2000 sample interpolation is applied to the panning process. These results show how the interpolation step makes the automatic panner more resilient to panning positioning flutter while achieving a more natural pan.

In Figure 6 the result of downmixing 12 sinusoidal test signals through the automatic panner are shown. It can be seen that both *f1* and *f12* are kept centered and added together because their spectral content is below 200Hz. The three sinusoids with a frequency of 5KHz have been evenly spread. *F2* has been allocated to the center due to priority; while *f4* has been send to the left and *f6* has been send to the right, in accordance to Table 1 for a *PS(3)*. Because there is no other signal with the same spectral content than *f11* it has been assigned to the center. The four sinusoids with a spectral content of 15Khz have been evenly spread. Because of priority *f3* has been assigned a MIDI step of 43, *f7* has been assigned a MIDI step of 84 steps, *f9* has been assigned all the way to the left, and *f10* has been assigned all the way to the right, in accordance to Table 1 for a *PS(4)*. Finally the two sinusoids with a spectral content of 20KHz have been panned to opposite sides. *f5* has been send to the left while *f8* has been send to the right, in accordance to Table 1 for a *PS(2)*. All results prove to be in accordance to the constrained rule equations proposed in section 2.3.

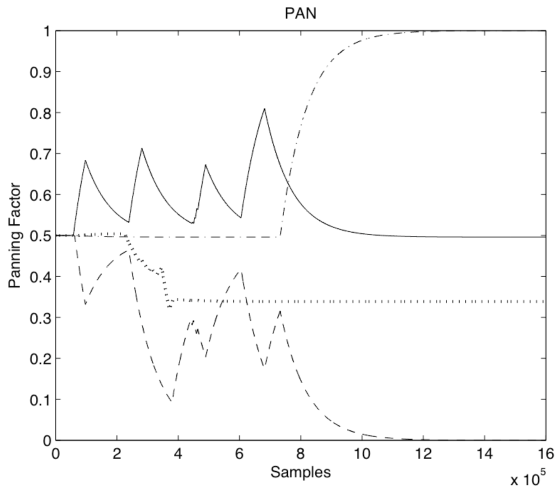

Figure 4: *Convergence of automatic panning algorithm for 4 different convergence values. (-) Panning Factor for a drum kit track,(--)panning Factor for a bass guitar, (.-) panning Factor for a vocal track, and (..) panning Factor for a channel input which spectral content is concentrated in the same filter.*

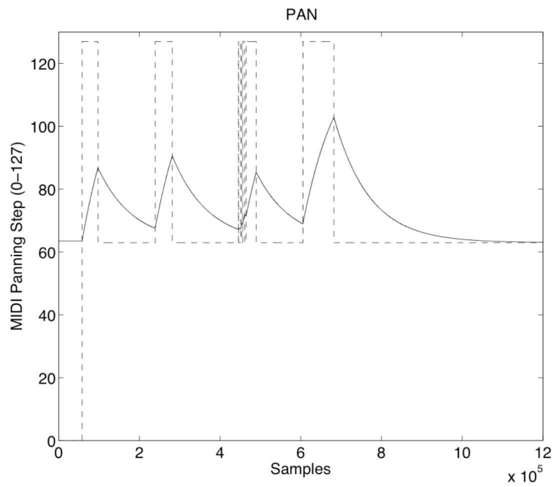

Figure 5: *Discrete panning step (- -). Super imposed interpolative panner angle (-) as applied to an input signal consisting of a drum kit recording.*

A Lissajous curve or stereogram is a two dimensional representation of a stereophonic signal and is usually perform by using a oscilloscope in XY mode or by using a vector oscilloscope. The stereogram can be obtained by plotting in time-synchronicity the left channel against the right channel. This measurement provides detailed information concerning inter-channel phase relationship [13]. The data contained in the stereogram of Figure 7 is widely spread in an oval. This means that the phase relation between the left and right channel is close to 90deg. This means we have achieved a wide spread stereo signal. In order to have a reference, a mono signal, which is represented by the diagonal separating the left and right planes of the stereogram has been plotted. The plot also shows a good data equilibrium between the right and left channel.

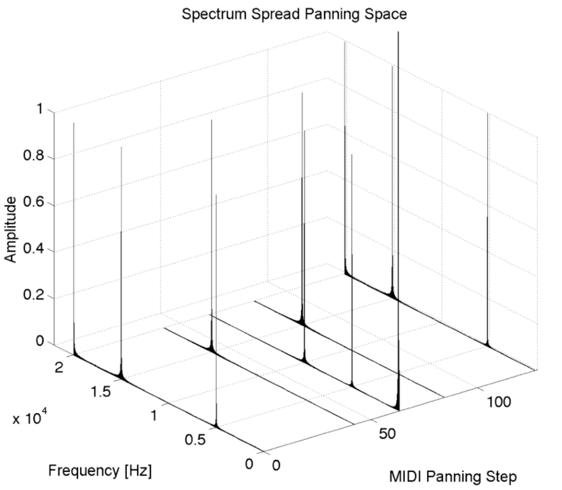

Figure 6: *Spectrum spread panning space for a 12 input CH automatic panner based on the proposed design. The test inputs were sinusoids with amplitude equal to one and the following frequencies: f1*=125Hz, *f2*=5KHz, *f3*=15KHz, *f4*=5KHz, *f5*=20KHz, *f6*=5KHz, *f7*=15KHz, *f8*=20KHz, *f9*=15KHz, *f10*=15KHz, *f11*=10KHz, and *f12*=125 Hz.

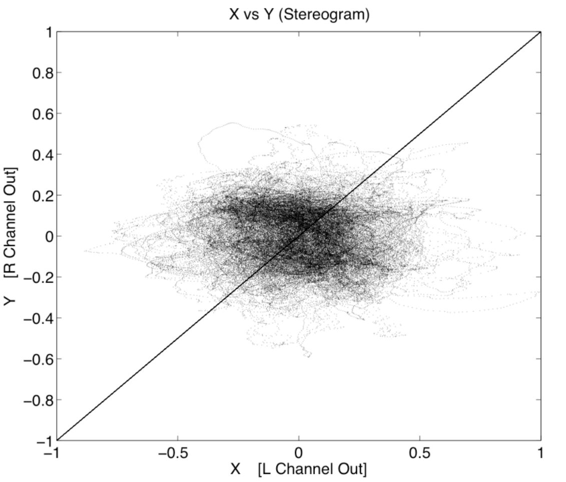

Figure 7: *Stereogram of 100,000 samples of a 5CH automatic panner. The samples correspond to a section in time where all 5-channel instruments are interacting simultaneously.*

4. CONCLUSIONS

An automatic panning algorithm for live multi-track sources has been successfully implemented. The algorithm reduces spectral masking while achieving a wide stereo signal. The system achieves panning by using constrained rules and bank filter accumulative techniques. Several test signals and multi-track signals have been tested. Results on its convergence, automatic panning space allocation, and stereogram have been presented. Subjective testing needs to be done in order to optimize the constrained rules and to verify the validity of our assumptions concerning placement of instruments and avoidance of spectral masking. Further work will also use instrument recognition techniques to automate priority rules that are independent of user channel connections. Filter bank optimization using psycho acoustical optimized band pass filters remains to be researched.

5. REFERENCES

[1] D. Dugan, "Automatic Microphone Mixing," in *51st Convention Audio Engineering Society*, San Fransico, California, 1975.

[2] D. Dugan, "Application of Automatic Mixing Techniques to Audio Consoles," in *87th Convention Audio Engineering Society* New York, 1989.

[3] S. Julstrom, and T. Tichy, "Direction-Sensitive Gating: A New Approach to Automatic Mixing," *Audio Engineering Society,* vol. 32, pp. 490-506, Julay/August 1984.

[4] B. Schick, R. Maillard, and C-C. Spenger, "First investigations on the use of manually and automatically generated stereo downmixes for spatial audio coding," in *118th Convention Audio Engineering Society* Barcelona, Spain, 2005.

[5] T. Nicolas, "Scalable Perceptual Mixing and Filtering of Audio Signals Using an Augmented Spectral Representation," in *DAFx' 05* Sophia Antipolis, France, 2005.

[6] C. Advendano, and J.-M. Jot, "A frequency-Domain Approach to Multichannel Upmix," *Journal Audio Engineering Society,* vol. 52, pp. 740-749, 2004.

[7] Y. Li, and P. F. Driessen, "An Unsupervised Adaptive Filtering Approach of 2-To-5 Channel Upmix," in *119th Convention Audio Engeneering Society* New York, 2005.

[8] F. Pachet, and O. Delerue, "On-the-Fly Multi-Track Mixing," in *109th Convention Audio Engineering Society* Los Angeles, California, USA, 2000.

[9] P. White, "Pan Position," in *The Sound on Sound book od desktop digital sound.*, M. Books, Ed., 2000, pp. 169-170.

[10] E. Benjamin, "An Experimental Verification of Localization in Two-Channel Stereo," in *121st Convention Audio Engineering Society* San Fransisco, California, USA, 2006.

[11] J. Beament, "The direction-finding system," in *How we Hear Music, The relationship Between Music and the Hearing Mechanism*, T. B. Press, Ed. Suffolk UK, 2001, pp. 127-130.

[12] E. Vincent, R. Gibonval, C. Favotte, and al., "BASS-dB: The Blind Separation Audio Source Separation Evaluation Database, http://www.irisa.fr/metiss/BASS-dB/," 2006.

[13] E. B. Brixen, "Audio Levels and Readings, http://www.dk-technologies.com/downloads/Audio%20Levels.pdf," in *Addition to User's manual of The Master Stereo Displays from DK-Audio*: DK-Audio, 2007.