

Research Article

A Real-Time Semiautonomous Audio Panning System for Music Mixing

Enrique Perez_Gonzalez and Joshua D. Reiss

*Centre for Digital Music Centre, School of Electronic Engineering and Computer Science, Queen Mary University of London,
Mile End Road, London E1 4NS, UK*

Correspondence should be addressed to Enrique Perez_Gonzalez, promix_mac@mac.com

Received 26 January 2010; Accepted 23 April 2010

Academic Editor: Udo Zoelzer

Copyright © 2010 E. Perez_Gonzalez and J. D. Reiss. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A real-time semiautonomous stereo panning system for music mixing has been implemented. The system uses spectral decomposition, constraint rules, and cross-adaptive algorithms to perform real-time placement of sources in a stereo mix. A subjective evaluation test was devised to evaluate its quality against human panning. It was shown that the automatic panning technique performed better than a nonexpert and showed no significant statistical difference to the performance of a professional mixing engineer.

1. Introduction

Stereo panning aims to transform a set of monaural signals into a two-channel signal in a pseudostereo field [1]. Many methods and panning ratios have been proposed, the most common one being the sine-cosine panning law [2, 3]. In stereo panning the ratio at which its power is spread between the left and the right channels determines the position of the source. Over the years the use of panning on music sources has evolved and some common practices can now be identified.

Now that recallable digital systems have become common, it is possible to develop intelligent expert systems capable of aiding the work of the sound engineer. An expert panning system should be capable of creating a satisfactory stereo mix of multichannel audio by using blind signal analysis, without relying on knowledge of original source locations or other visual or contextual aids.

This paper extends the work first presented in [4]. It presents an expert system capable of blindly characterizing multitrack inputs and semiautonomously panning sources with panning results comparable to a human mixing engineer. This was achieved by taking into account technical constraints and common practices for panning, while minimizing human input. Two different approaches are

described and subjective evaluation demonstrates that the semi-autonomous panner has equivalent performance to that of a professional mixing engineer.

2. Panning Constraints and Common Practices

In practice, the placement of sound sources is achieved using a combination of creative choices and technical constraints based on human perception of source localization. It is not the purpose of this paper to emulate the more artistic and subjective decisions in source placement. Rather, we seek to embed the common practices and technical constraints into an algorithm which automatically places sound sources. The idea behind developing an expert semi-autonomous panning machine is to use well-established common rules to devise the spatial positioning of a signal.

- (1) When the human expert begins to mix, he or she tends to do it from a monaural, all centered position, and gradually moves the pan pots [5]. During this process, all audio signals are running through the mixer at all times. In other words, source placement is performed in realtime based on accumulated knowledge of the sound sources and the resultant

mix, and there is no interruption to the signal path during the panning process.

- (2) Panning is not the result of individual channel decisions; it is the result of an interaction between channels. The audio engineer takes into account the content of all channels, and the interaction between them, in order to devise the correct panning position of every individual channel [6].
- (3) The sound engineer attempts to maintain balance across the stereo field [7, 8]. This helps maintain the overall energy of the mix evenly split over the stereo speakers and maximizes the dynamic use of the stereo channels.
- (4) In order to minimize spectral masking, channels with similar spectral content are placed apart from each other [6, 9]. This results in a mix where individual sources can be clearly distinguished, and this also helps when the listener uses the movement of his or her head to interpret spatial cues.
- (5) Hard panning is uncommon [10]. It has been established that panning a ratio of 8 to 12 dBs is more than enough to achieve a full left or full right image [11]. For this reason, the width of the panning positions is restricted.
- (6) Low-frequency content should not be panned. There are two main reasons for doing this. First, it ensures that the low-frequency content remains evenly distributed across speakers [12]. This minimizes audible distortions that may occur in the high-power reproduction of low-frequencies. Second, the position of a low frequency source is often psychoacoustically imperceptible. In general, we cannot correctly localize frequencies lower than 200 Hz [13]. It is thought that this is due to the fact that the use of InterTime Difference as a perceptual clue for localization of low frequency sources is highly dependent on room acoustics and loudspeaker placement, and Inter-Level Differences are not a useful perceptual cue at low frequencies since the head only provides significant attenuation of high-frequency sources [14].
- (7) High-priority sources tend to be kept towards the centre, while lower priority sources are more likely to be panned [7, 8]. For instance, the vocalist in a modern pop or rock group (often the lead performer) would often not be panned. This relates to the idea of matching a physical stage setup to the relative positions of the sources.

3. Implementation

3.1. Cross-Adaptive Implementation. The automatic panner is implemented as a cross-adaptive effect, where the output of each channel is determined from analysis of all input channels [15]. For applications that require a realtime signal processing, the signal analysis and feature extraction has been implemented using side chain processing, as depicted in Figure 1. The audio signal flow remains real-time while

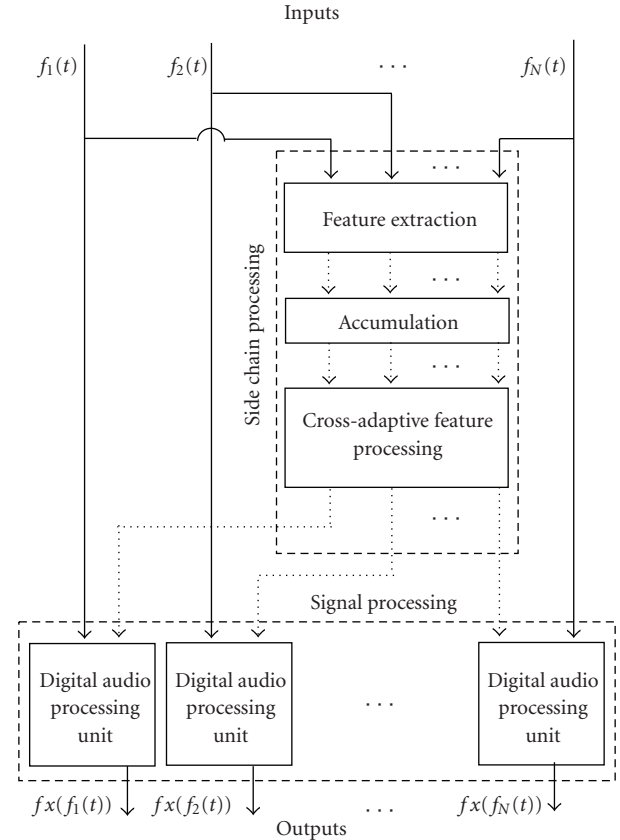


FIGURE 1: General diagram of a cross-adaptive device using side chain processing with feature accumulation.

the required analysis of the input signals is performed in separate instances. The signal analysis involves accumulating a weighted time average of extracted features. Accumulation allows us to quickly converge on an appropriate panning position in the case of a stationary signal or smoothly adjust the panning position as necessary in the case of changing signals. Once the feature extraction within the analysis side chain is completed, then the features from each channel are analyzed in order to determine new panning positions for each channel. Control signals are sent to the signal processing side in order to trigger the desired panning commands.

3.2. Adaptive Gating. Because noise on an input microphone channel may trigger undesired readings, the input signals are gated. The threshold of the gate is determined in an adaptive manner. By noise we refer not only to random ambient noise but also to interference due to nearby sources, such as the sound from adjacent instruments that are not meant to be input to a given channel.

Adaptive gating is used to ensure that features are extracted from a channel only when the intended signal is present and significantly stronger than the noise sources. The gating method is based on a method implemented in [16, 17] by Dugan. A reference microphone may be placed outside of the usable source microphone area to capture a signal representative of the undesired ambient and interference

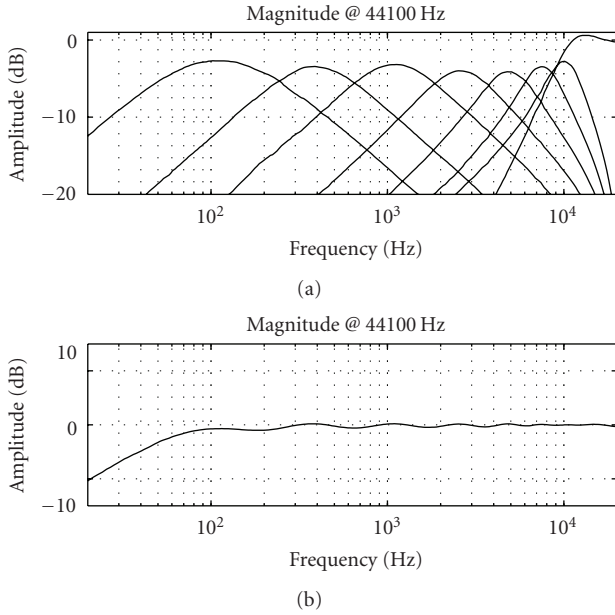


FIGURE 2: Quasiflat frequency response bandpass filter bank (Type A filter bank for $K = 8$). (a) Filter bank consisting of a set of eight second order band-pass IIR Biquadratic filters with center frequencies as follow: 100 Hz, 400 Hz, 1 kHz, 2.5 kHz, 5 kHz, 7.5 kHz, 10 kHz, and 15000 kHz. (b) Combined response of the filter bank.

noise. The reference microphone signal is used to derive an adaptive threshold by opening the gate only if the input signal magnitude is greater than the reference microphone magnitude signal. Therefore the input signal is only passed to the side processing chain when its level exceeds that of the reference microphone signal.

3.3. Filter Bank Implementation. The implementation uses a filter bank to perform spectral decomposition of each individual channel. The filter bank does not affect the audio path since it is only used in the analysis section of the algorithm. It was chosen as opposed to other methods of classifying the dominant frequency or frequency range of a signal [18] because it does not require Fourier analysis, and hence is more amenable to a real-time implementation.

For the purpose of finding the optimal spectral decomposition for performing automatic panning, two different eight-band filter banks were designed and tested. The first consisted of a quasiflat frequency response bandpass filter bank, which for the purposes of this paper we will call filter bank type A in Figure 2, and the second contained a lowpass filter decomposition filter bank, which we will call filter bank type B in Figure 3. In order to provide an adaptive frequency resolution for each filter bank, the total number of filters, K , is equal to the number of input channels that are meant to be panned. The individual gains of each filter were optimized to achieve a quasiflat frequency response.

3.4. Determination of Dominant Frequency Range. Once the filter bank has been designed, the algorithm uses the

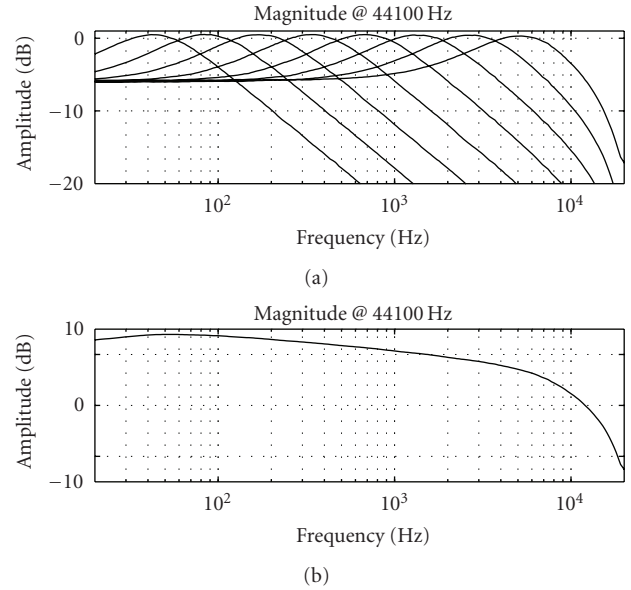


FIGURE 3: Lowpass filter decomposition filter bank (Type B filter bank for $K = 8$). (a) Filter bank comprised of a set of second order low-pass IIR Biquadratic filters with cutoff frequencies as follows: 35 Hz, 80 Hz, 187.5 Hz, 375 Hz, 750 Hz, 1.5 kHz, 3 kHz, and 6 kHz. (b) Combined response of the filter bank. All gains have been set to have a maximum peak value of 0 dBs.

band limited signal in each filter’s output to obtain the absolute peak amplitude for each filter. The peak amplitude is measured within a 100 ms window. The algorithm uses the spectral output of each filter contained within the filter bank to calculate the peak amplitude of each band. By comparing these peak amplitudes, the filter with the highest peak is found. An accumulated score is maintained for the number of occurrences of the highest peak in each filter contained within the filter bank. This results in a classifier that determines the dominant filter band for an input channel from the highest accumulated score.

The block diagram of the filter bank analysis algorithm is provided in Figure 4. It should be noted that this approach uses digital logic operations of comparison and addition only, which makes it highly attractive for an efficient digital implementation.

3.5. Cross-Adaptive Mapping. Now that each input channel has been analyzed and associated with a filter, it remains to define a mapping which results in the panning position of each output channel. The rules which drive this cross-adaptive mapping are as follows.

The first rule implements the constraint that low-frequency sources should not be panned. Thus, all sources with accumulated energy contained in a filter with a high cutoff frequency below 200 Hz are not panned and remain centered at all times.

The second rule decides the available panning step of each source. This is a positioning rule which uses equidistant spacing of all sources with the same dominant frequency

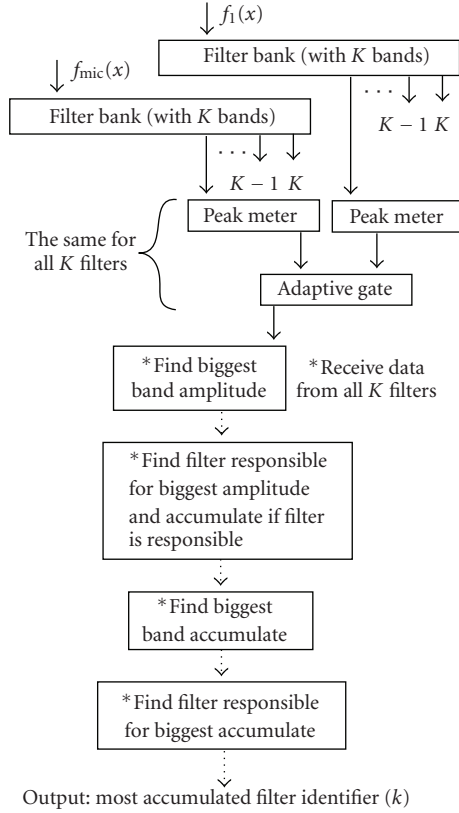


FIGURE 4: Analysis block diagram for one input channel.

range. Initially, all sources are placed in the centre. The available panning steps are calculated for every different accumulated filter, k , where k ranges from 1 to K , based on the number of sources, N_k , whose dominant frequency range resides in that filter. Due to this channel dependency the algorithm will update itself every time a new input is detected in a new channel, or if the spectral content of an input channel suffers from a drastic change over time.

For filters which have not reached maximum accumulation there is no need to calculate the panning step, which makes the algorithm less computationally expensive. If only one repetition exists for a given k th filter ($N_k = 1$) the system places the input at the center.

The following equation gives the panning space location of the i th source residing in the k th filter:

$$P(i, k) = \begin{cases} 1/2, & N_k = 1, \\ \frac{N_k - i - 1}{2(N_k - 1)}, & i + N_k \text{ is odd,} \\ 1 - \frac{N_k - i}{2(N_k - 1)}, & i + N_k \text{ is even, } N_k \neq 1, \end{cases} \quad (1)$$

where $P(i, k)$ is the i th available panning step in the k th filter, i ranges from 1 to N_k and $P(i, k) = 0.5$ corresponds to a center panning position.

Using this equation, if N_k is odd, the first source, $i = 1$, is positioned at the center. When N_k is even, the first two sources, $i = 2, 3$, are positioned either side of the center. In both cases, the next two sources is positioned either side of the previous sources and so on such that sources are positioned further away from the centre as i increases. The extreme panning positions are 0 and 1. However, as mentioned, hard panning is generally not preferred. So our current implementation provides a panning width control, P_W , used to narrow or widen the panning range. The panning with has a valid range from 0 to 0.5 where 0 equates to the wide panning possible and 0.5 equates to no panning. In our current implementation, it defaults to $P_W = 0.059$. The P_W value is subtracted for all panning positions bigger than 0.5 and added to all panning positions smaller than 0.5. In order to avoid sources originally panned left to cross to the right or sources originally panned right to cross to the left, the panning width algorithm ensures that sources in such cases default to the centre position.

3.6. Priority. Equation (1) provides the panning position for each of the sources with dominant spectral content residing in the k th filter, but it does not say how each of those sources are ordered from source $i = 1$ to $i = N_k$. The common practices mentioned earlier would suggest that certain sources, such as lead vocals, would be less likely to be panned to extremes than others, such as incidental percussions. However, the current implementation of our automatic panner does not have access to such information. Thus, the authors have proposed to use a priority driven system in which the user can label the channels according to importance. In this sense, it is a semiblind automatic system. Thus, all sources are ordered from highest to the lowest priority. For the N_k sources residing in the k th filter, the first panning step is taken by the highest priority source, the second panning step by the next highest priority source, and so on. The block diagram containing the constrained decision control rule stage of the algorithm is presented in Figure 5.

3.7. The Panning Processing. Once the appropriate panning position was determined, a sine-cosine panning law [2] was used to place sources in the final sound mix:

$$\begin{aligned} f_{Rout}(x) &= \sin\left(\frac{P(i, k)\pi}{2}\right) \cdot f_{in}(x), \\ f_{Lout}(x) &= \cos\left(\frac{P(i, k)\pi}{2}\right) \cdot f_{in}(x). \end{aligned} \quad (2)$$

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

In Figure 6, the result of down-mixing 12 sinusoidal test signals through the automatic panner is shown. It can be seen that both f_1 and f_{12} are kept centered and added together because their spectral content is below 200 Hz. The three sinusoids with a frequency of 5 kHz have been evenly spread.

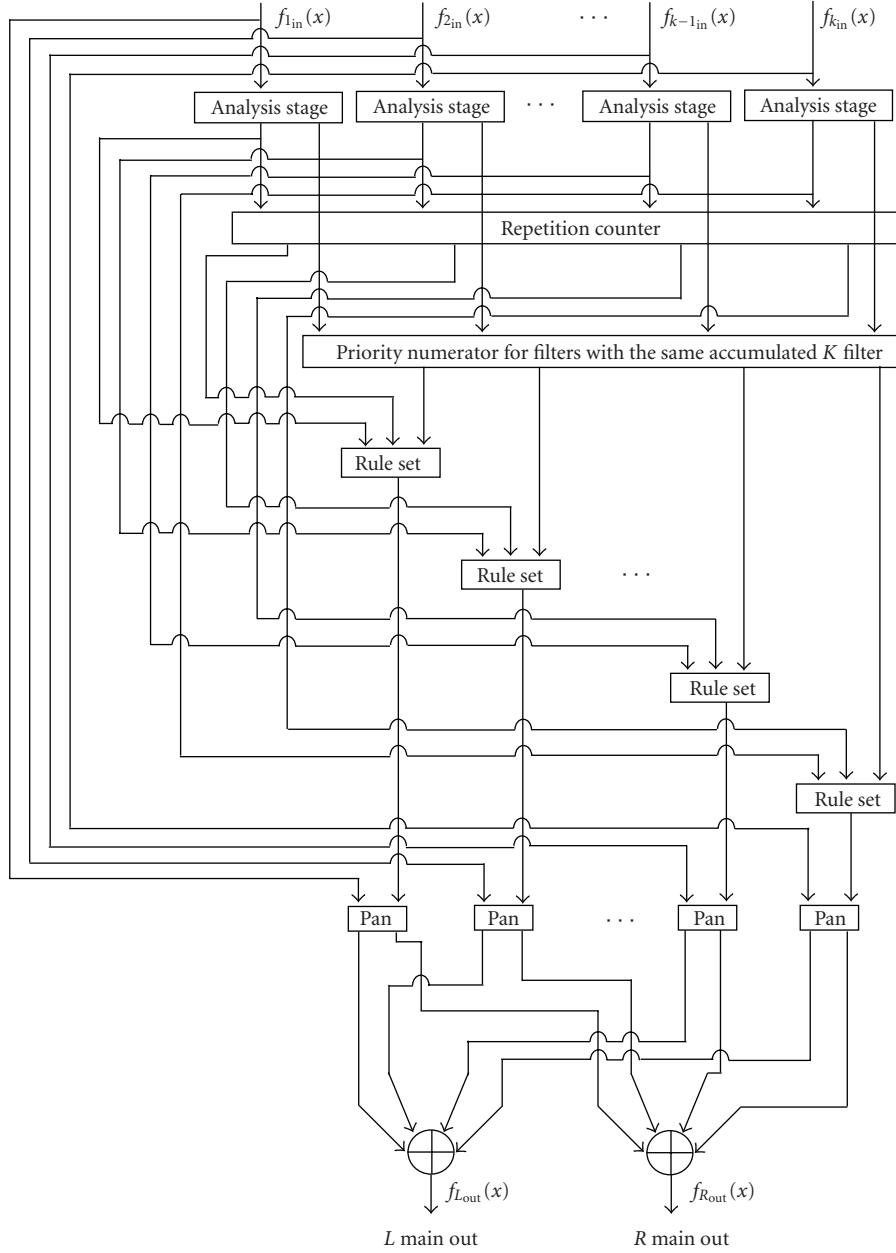


FIGURE 5: Block diagram of the automatic panner constrained control rules algorithm for K input channels.

f_2 has been allocated to the center due to priority while f_4 has been send to the left and f_6 has been send to the right, in accordance with (1). Because there is no other signal with the same spectral content than f_{11} , it has been assigned to the center. The four sinusoids with a spectral content of 15 kHz have been evenly spread. Because of priority, f_3 has been assigned a value of 0.33, f_7 has been assigned a value of 0.66, f_9 has been assigned all the way to the left, and f_{10} has been assigned all the way to the right, in accordance with (1). Finally, the two sinusoids with a spectral content of 20 KHz have been panned to opposite sides. All results prove to be in accordance with the constraint rules proposed for cross-adaptive mapping.

4. Results

4.1. Test Design. In order to evaluate the performance of the semiautonomous panner algorithm against human performance, a double blind test was designed. Both of auto-panning algorithms were tested, the bandpass filter classifier known as algorithm type A, and the low-pass classifier known as algorithm type B. Algorithms were randomly tested in a double blind fashion.

The control group consisted of three professional human experts and one nonexpert, who had never panned music before. The test material consisted of 12 multitrack songs of different styles of music. Stereo sources were used in the form

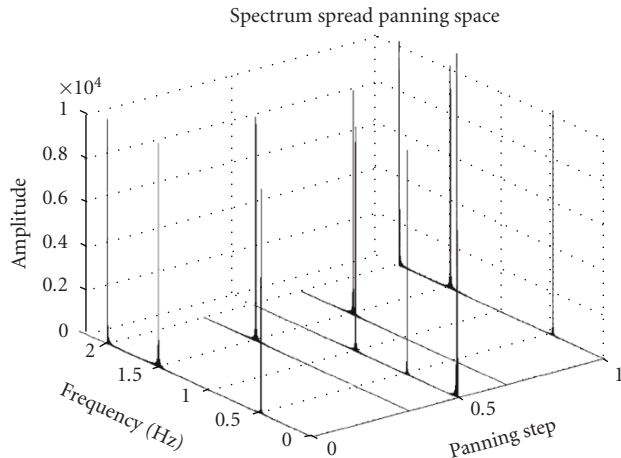


FIGURE 6: Results of automatic panning based on the proposed design. The test inputs were 12 sinusoids with amplitude equal to one and the following frequencies: $f_1 = 125$ Hz, $f_2 = 5$ kHz, $f_3 = 15$ kHz, $f_4 = 5$ kHz, $f_5 = 20$ kHz, $f_6 = 5$ kHz, $f_7 = 15$ kHz, $f_8 = 20$ kHz, $f_9 = 15$ kHz, $f_{10} = 15$ kHz, $f_{11} = 10$ kHz, and $f_{12} = 125$ Hz.

of two separate monotracks. Where acoustic drums were used, they would be recorded with multiple microphones and then premixed down into a stereo mix. Humans and algorithms used the same stereo drum and keyboard tracks as separate left and right monofiles. All 12 songs were panned by the expert human mixers and by the nonexpert human mixer. They were asked to pan the song while listening for the first time. They had the length of the song to determine their definitive panning positions. The same songs were passed through algorithms A and B only once for the entire length of the song. Although the goal was to give the human and machine mixers as close to the same information as possible, human mixers had the advantage of knowing which type of instrument it was. Therefore, they assigned priority according to this prior known knowledge. For this reason a similar priority scheme was chosen to compensate for this. Both A and B algorithms used the same priority schema. Mixes used during the test contain music freely available under creative commons copyright can be located in [19].

As shown in Figure 7, the test used two questions to measure the perceived overall quality of the panning for each audio comparison. For the first question, “how different is the panning of A compared to B?”, a continuous slider with extremities marked “exactly the same” and “completely different” was used. The answer obtained in this question was used as a weighting factor in order to decide the validity of the next question. The second question, “which file, A or B, has better panning?”, used a continuous slider with extremes marked “A quality is ideal” and “B quality is ideal”. For both of these questions, no visible scale was added in order not to influence their decision. The test subjects were also provided with a comment box that was used for them to justify their answers to the previous two questions. During the test it was observed that expert subjects tend to use the name of the instrument to influence their panning decisions. In other words they would look for the “bass” label to make

sure that they kept it center. This was an encouraging sign that panning amongst professionals follows constraint rules similar to those that were implemented in the algorithms.

The tested population consisted of 20 professional sound mixing engineers, with an average experience of 6-year work in the professional audio market. The tests were performed over headphones, and both the human mixers and the test subjects used exactly the same headphones. The test lasted an average time of 82 minutes.

Double blind A/B testing was used with all possible permutations of algorithm A, algorithm B, expert and amateur panning. Each tested user answered a total of 32 questions, two of which were control questions, in order to test the subject’s ability to identify stereo panning. The first control question consisted of asking the test subjects to rate their preference between a stereo and a monaural signal. During the initial briefing it was stressed to the test subject that stereo is not necessarily better than monaural audio. The second control question compared two stereo signals that had been panned in exactly the same manner.

4.2. Result Analysis. All resulting permutations were classified into the following categories: monaural versus stereo, same stereo versus same stereo file, method A versus method B, method A versus nonexpert mix, method B versus nonexpert mix, method A versus expert mix, and method B versus expert mix.

Results obtained on the question “How different is panning A compared to B?” were used to weight the results obtained for the second question “Which file, A or B, has better panning quality?”. This is in order to have a form of neglecting incoherent answers such as “I find no difference between files A or B but I find the quality of B to be better compared to A”.

Answers to the first question showed that, with at least 95% confidence, the test subjects strongly preferred stereo to monaural mixes. The second question also confirmed with at least 95% confidence that professional audio engineers find no significant difference when asked to compare two identical stereo tracks. The results are summarized in Table 1, and the evaluation results with 95% confidence intervals are depicted in Figure 8.

The remaining tests compared the two panning techniques against each other and against expert and nonexpert mixes. The tested audio engineers preferred the expert mixes to panning method A, but this result could only be given with 80% confidence. On average, non-expert mixes also were preferred to panning method A, but this result could not be considered significant, even with 80% confidence.

In contrast, panning method B was preferred over non-expert mixes with over 90% confidence. With at least 95% confidence, we can also state that method B was preferred over method A. Yet when method B is compared against expert mixes, there is no significant difference.

The preference for panning method B implies that low-pass spectral decomposition is preferred over band-pass spectral decomposition as a means of signal classification for the purpose of semi-autonomous panning. Furthermore, the

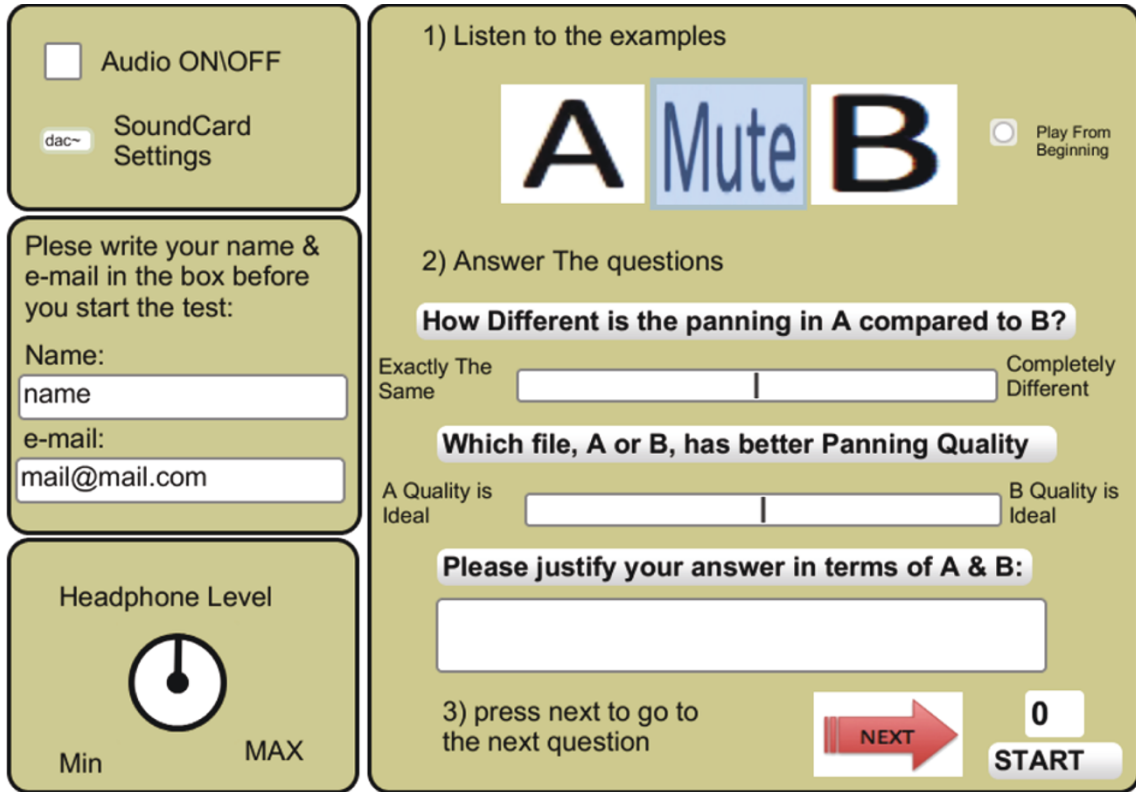


FIGURE 7: Double blind panning quality evaluation test interface.

lack of any statistical difference between panning method B and expert mixe, (in contrast to the significant preference for method B over non-expert mixes, and for expert mixes over method A) leads us to conclude that the semi-autonomous panning method B performs roughly equivalently to an expert mixing engineer.

It was found that the band-pass filter bank, method A, tended to assign input channels to less filters than the low-pass filter bank, method B. The distribution of input tracks among filters for an 8-channel song for both methods is depicted in Figure 9. In effect, panning method B is more discriminating as to whether two inputs have overlapping spectral content and hence is less likely to unnecessarily place sources far from each other. This may account for the preference of panning method B over panning method A.

The subjects justified their answers in accordance with the common practices mentioned previously. They relied heavily on manual instrument recognition to determine the appropriate position of each channel. It was also found that any violation of common practice, such as panning the lead vocals, would result in a significantly low measure of panning quality. One of the most interesting findings was that spatial balance seemed to be not only a significant cue used to determine panning quality but was also a distinguishing factor between expert and non-expert mixes. Non-expert mixes were often weighted to one side, whereas almost universally, expert mixes had the average source position in the centre. Both panning methods A and B were devised to

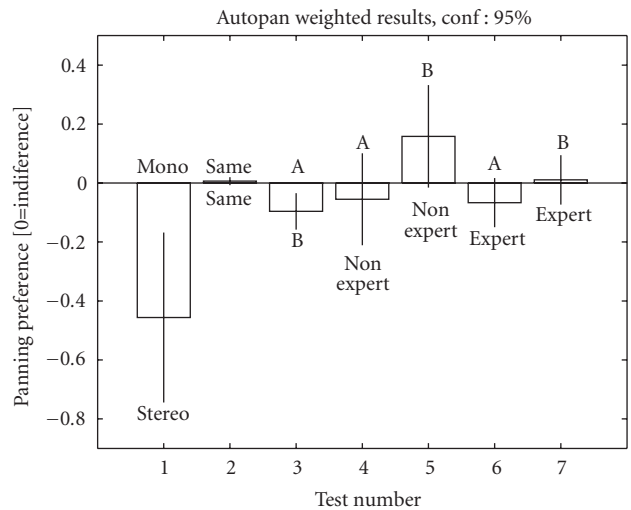


FIGURE 8: Summarized results for the subjective evaluation. The first two tests were references (comparing stereo against monaural, and comparing identical files), and the remaining questions compared the two proposed auto-panning methods against each other and against expert and non-expert mixes. 95% confidence intervals are provided.

perform optimal left to right balancing. Histograms of source positions which demonstrate these behaviors are depicted in Figure 10.

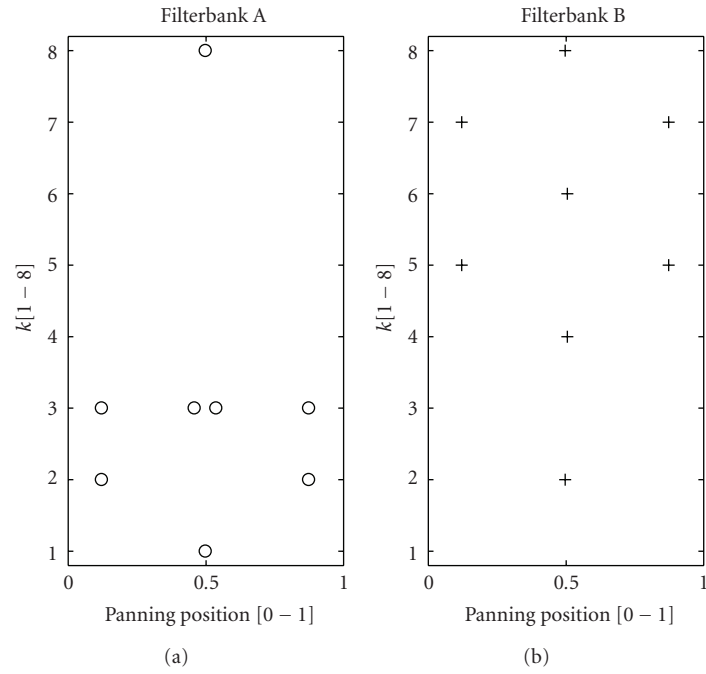


FIGURE 9: Distribution of sources among filters for methods A and B for the same song ($P_W = 0.059$).

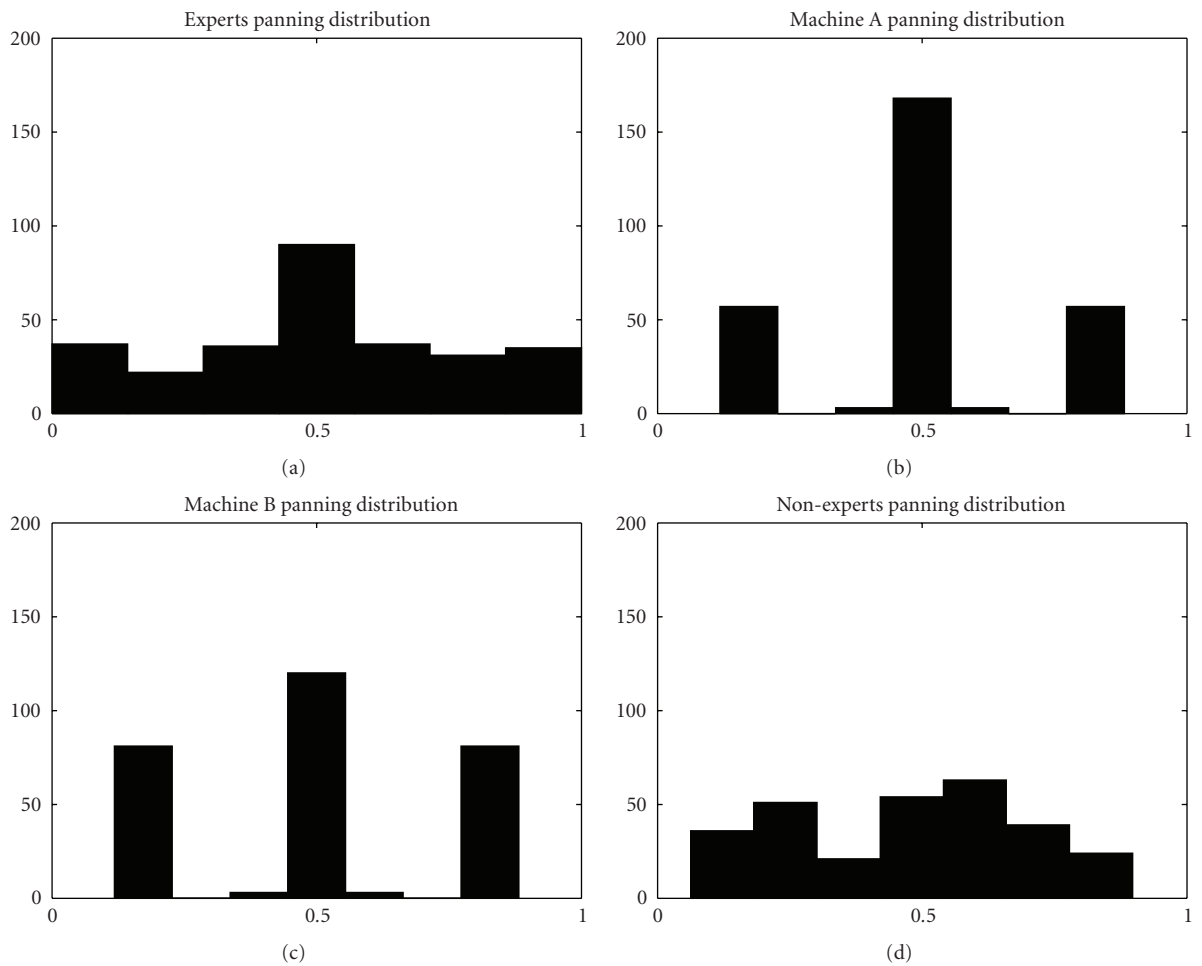


FIGURE 10: Panning space distribution histograms ($P_W = 0.059$).

TABLE 1: Double blind panning quality evaluation table.

Test	Number of Comparisons	Preference	Confidence	Standard Deviation	Mean
Stereo versus Mono	20	Stereo	95%	0.61545	-0.4561
Stereo versus Stereo	20	Identify them to be the same	95%	0.0287	0.0064
Human Expert versus Method A	144	Human	80%	0.5105	-0.666
Method A versus Non-expert	56	No significant difference between algorithms	95%	0.5956	-0.0552
Method B versus Non-expert	56	Method B	90%	0.6631	0.1583
Method B versus Expert	144	No significant difference between algorithms	95%	0.5131	0.0108
Method A versus Method B	200	Method B	95%	0.4474	-0.0962

5. Conclusions and Future Work

In terms of generating blind stereo panning up-mixes with minimum human interactions, we can conclude that it is possible to generate intelligent expert systems capable of performing better than a non-expert human while having no statistical difference when compared to a human expert. According to the subjective evaluation, low-pass filter-bank accumulative spectral decomposition features seem to perform significantly better than band-pass decompositions.

More sophisticated forms of performing source priority identification in an unaided manner need to be investigated. To further automate the panning technique, instrument identification and other feature extraction techniques could be employed to identify those channels with high priority. Better methods of eliminating microphone cross-talk noise need to be researched. Furthermore, in live sound situations, the sound engineer would have visual cues to aid in panning. For instance, the relative positions of the instruments on stage are often used to map sound sources in the stereo field. Video analysis techniques could be used to incorporate this into the panning constraints. Future subjective tests should include visual cues and be performed in real sound reinforcement conditions.

Finally, the work presented herein was restricted to stereo panning. In a two- or three-dimensional sound field, there are more degrees of freedom, but rules still apply. For instance, low-frequency sources are often placed towards the ground, while high-frequency sources are often placed above, corresponding to the fact that high frequency sources emitted near or below the ground would be heavily attenuated [20]. It is the intent of the authors to extend this work to automatic placement of sound sources in a multispeaker, spatial audio environment.

References

- [1] M. A. Gerzon, "Signal processing for simulating realistic stereo images," in *Proceedings of the 93rd Convention Audio Engineering Society*, San Francisco, Calif, USA, October 1992.
- [2] J. L. Anderson, "Classic stereo imaging transforms—a review," submitted to *Computer Music Journal*, http://www.dxarts.washington.edu/courses/567/08WIN/JL_Anderson_Stereo.pdf.
- [3] D. Griesinger, "Stereo and surround panning in practice," in *Proceedings of the 112th Audio Engineering Society Convention*, Munich, Germany, May 2002.
- [4] E. Perez_Gonzalez and J. Reiss, "Automatic mixing: live down-mixing stereo panner," in *Proceedings of the 7th International Conference on Digital Audio Effects (DAFx '07)*, pp. 63–68, Bordeaux, France, 2007.
- [5] D. Self, et al., "Recording consoles," in *Audio Engineering: Know It All*, D. Self, Ed., vol. 1, chapter 27, pp. 761–807, Newnes/Elsevier, Oxford, UK, 1st edition, 2009.
- [6] R. Neiman, "Panning for gold: tutorials," *Electronic Musician Magazine*, 2002, http://emusician.com/tutorials/emu-sic_panning_gold/.
- [7] R. Izhaki, "Mixing domains and objectives," in *Mixing Audio: Concepts, Practices and Tools*, chapter 6, pp. 58–71, Focal Press/Elsevier, Burlington, Vt, USA, 1st edition, 2007.
- [8] R. Izhaki, "Panning," in *Mixing Audio: Concepts, Practices and Tools*, chapter 13, pp. 184–203, Focal Press/Elsevier, Burlington, Vt, USA, 1st edition, 2007.
- [9] B. Bartlett and J. Bartlett, "Recorder-mixers and mixing consoles," in *Practical Recording Techniques*, chapter 12, pp. 259–275, Focal Press/Elsevier, Oxford, UK, 3rd edition, 2009.
- [10] B. Owsinski, "Element two: panorama—placing the sound in the soundfield," in *The Mixing Engineer's Handbook*, chapter 4, pp. 20–24, Mix Books, Vallejo, Calif, USA, 2nd edition, 2006.
- [11] F. Rumsey and T. McCormick, "Mixers," in *Sound and Recording: An Introduction*, chapter 5, pp. 96–153, Focal Press / Elsevier, Oxford, UK, 1st edition, 2006.

- [12] P. White, "The creative process: pan position," in *The Sound on Sound Book of Desktop Digital Sound*, pp. 169–170, MPG Books, UK, 1st edition, 2000.
- [13] E. Benjamin, "An experimental verification of localization in two-channel stereo," in *Proceedings of the 121st Convention Audio Engineering Society*, San Fransisco, Calif, USA, 2006.
- [14] J. Beament, "The direction-finding system," in *How we Hear Music: The relationship Between Music and the Hearing Mechanism*, pp. 127–130, The Boydell Press, Suffolk, UK, 2001.
- [15] V. Verfaillie, U. Zölzer, and D. Arfib, "Adaptive Digital Audio Effects (A-DAFx): a new class of sound transformations," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 5, pp. 1817–1831, 2006.
- [16] D. W. Dugan, "Automatic microphone mixing," *Journal of the Audio Engineering Society*, vol. 23, no. 6, pp. 442–449, 1975.
- [17] D. W. Dugan, "Application of automatic mixing techniques to audio consoles," in *Proceedings of the 87th Convention of the Audio Engineering Society*, pp. 18–21, New York, NY, USA, October 1989.
- [18] W. A. Sethares, A. J. Milne, S. Tiedje, A. Prechtel, and J. Plamondon, "Spectral tools for dynamic tonality and audio morphing," *Computer Music Journal*, vol. 33, no. 2, pp. 71–84, 2009.
- [19] E. Perez-Gonzalez and J. Reiss, "Automatic mixing tools for audio and music production," 2010, <http://www.elec.qmul.ac.uk/digitalmusic/automaticmixing/>.
- [20] D. Gibson and G. Peterson, *The Art of Mixing: A Visual Guide to Recording, Engineering and Production*, Mix Books / ArtistPro Press, USA, 1st edition, 1997.