Audio Engineering Society
Convention Paper 8736

Presented at the 133rd Convention
2012 October 26–29 San Francisco, CA, USA

An Autonomous System for Multi-track Stereo Pan Positioning

Stuart Mansbridge[1], Saoirse Finn[2], and Joshua D. Reiss[1]

[1] Centre for Digital Music, Queen Mary University of London, Mile End Road, London, E1 4NS, UK
stuart.mansbridge@eecs.qmul.ac.uk
josh.reiss@eecs.qmul.ac.uk

[2] Centre for Digital Music, Queen Mary University of London, Mile End Road, London, E1 4NS, UK
saoirse.finn@eecs.qmul.ac.uk
and Birmingham City University, Birmingham, B4 7XG, UK
saoirse.finn@mail.bcu.ac.uk

ABSTRACT

A real-time system for automating stereo panning positions for a multi-track mix is presented. Real-time feature extraction of loudness and frequency content, constrained rules and cross-adaptive processing are used to emulate the decisions of a sound engineer, and pan positions are updated continuously to provide spectral and spatial balance with changes in the active tracks. As such, the system is designed to be highly versatile and suitable for a wide number of applications, including both live sound and post-production. A real-time, multi-track C++ VST plug-in version has been developed. A detailed evaluation of the system is given, where formal listening tests compare the system against professional mixes from a variety of genres.

1. INTRODUCTION

Stereo positioning is the process of changing the apparent location of a sound source in a binaural audio mix. Most commonly, this is achieved by feeding left and right channels with the same sound source and adjusting the relative amplitude of the channels. This is referred to as the interaural level difference (ILD) [1], and is traditionally adjusted by the pan pots on a mixing desk.

Localisation of sound sources is also aided by temporal differences between the ear channels, known as the interaural time difference (ITD), for frequencies lower than 1.5kHz [1]. This accounts for the extra time required for longer-wavelength sounds to reach the ear, and in sound production is achieved by introducing an appropriate delay between the channels, and additionally equalization to approximate a high-frequency roll-off due to the acoustic effects of the head. ITD delay is in the region of 1-2ms, while Haas panning (without the use of

ILD) is around 20ms [2]. Delay based stereo positioning techniques can be very effective, but equally can introduce issues when listening using loudspeakers e.g. comb-filtering and the requirement for the listener to be in a central location between the two speakers. Typically, the technique is used sparingly and only in post-production. The focus of this paper is therefore on the stereo placement of sources using ILD, although the use of ITD is considered and has been built into the plug-in as an additional option.

We propose a new innovative solution to automating the task of panning a mix. The motivation behind the adopted approach, as with the majority of the work by the authors in the field of Automatic Music Production, is to determine general rules and constraints which can be adopted to emulate the performance of a sound engineer in a real-time environment. This requires the extraction of features and cross-adaptive processing to analyse all incoming tracks and reach appropriate decisions for adjusting the mixing controls. To create appropriate rules, the techniques employed by sound engineers have been studied in depth.

This paper provides a wide array of improvements over the original proof-of-concept papers [3][4], such as being intended for both live and post-production use, an arbitrary number of tracks, fully autonomous, use of spectral centroid from an Fast Fourier Transform (FFT) instead of a filter bank, better use of panning space and spectral/spatial balancing. It also draws upon the real-time processing challenges introduced in the authors' previous paper [5]. With the addition of techniques such as vocal detection, currently in development, the proposal is for a fully autonomous panning tool with minimal or no human interaction required.

The frequency content of each track and the relationship between them is used to determine panning position. However, the main objective when panning is to retain balance in the stereo domain. A key part of the algorithm is therefore the steps taken to monitor different measures of balance, and to perform adjustments where necessary.

2. CONSTRAINTS

2.1. Panning position determined by frequency content

An analysis of mixing practice shows sources with higher frequency content are progressively panned further towards the extremes. A typical drum kit, for example, places the kick drum in the centre, the toms and snare close on either side, with the high frequency cymbals furthest left and right. Furthermore, high frequency sounds diffract less as they bend around the head, and so the panning effect needs to be greater to represent this [6,7]. For these reasons an expanding panning width is needed to push higher frequency sources wider in the stereo field.

2.2. Low frequency sources centred

In addition to expanding the panning width with frequency, sources with frequency content below a certain threshold should be fixed in the centre of the mix. Having low frequency sources off centre can provide an uneven power distribution, and furthermore, due to the longer wavelength there is little or no directional information below 200Hz [6,7][8].

2.3. Minimise spectral masking

Panning techniques dictate that sources with similar spectral content be placed apart in the stereo field to minimise spectral masking [9,10]. As a result the tracks can be more easily distinguished in the mix.

2.4. Spatial Balance

Spatial balance is the comparison of signal level between left and right channels and is the most important consideration when mixing, where the aim is for both to be approximately equal [6,7]. As the activity and intensity of sources can change during a song the source placement must be able to adapt to provide a balanced mix.

2.5. Spectral Balance

Spectral balance is the ratio of intensity of frequency content between left and right, so that there is an equal spread of frequencies across the mix [6,7].

Audio signals are typically divided into well-defined frequency bands: lows, low-mids, high-mids and highs, and each band should have approximately equal content in left and right channels. Where there is a single source dominating a band, typically the source will be moved towards the middle of the mix, or the source may be duplicated in the opposite channel and a stereo effect (phase, delay, reverb etc.) applied [2].

2.6. Stereo Spread

Stereo spread is a measure of how the whole panning space has been filled. For a full stereo image there should be an even distribution of sources to avoid gaps in the stereo field [6,7].

2.7. Panning Width

An additional consideration is the overall weighting on the panning width. Generally speaking hard panning of sources is unnecessary [11], but an overall weighting on the panning width allows the extent of the utilised panning space to be controlled.

2.8. Choice of lead vocal track(s) to be centred

In popular music, a lead vocal track is likely to be the focal point of a song. To provide balance and a natural listening experience it is most common for the track to be placed in the centre of the panning space [6,7]. In this situation, user interaction to designate lead vocal tracks is desirable. For a fully autonomous system, however, vocal detection techniques can be used to automate the process. With the understanding that vocal detection is unlikely to be 100% accurate, the weighting on the decision needs to be in favour of false positives which may place more tracks than desirable in the centre and produce a sub-optimal mix, rather than false negatives that may place a lead vocal in a non-central position and be most likely to produce a poor mix.

2.9. Time-varying panning positions

Fixed pan positions are unlikely to remain optimal for the entirety of a track. Sound engineers will typically adjust pan positions over time or record automation curves in Digital Audio Workstation (DAW) software to make alterations to the mix. The algorithm therefore needs to continually tweak the pan positions to optimise the mix.

2.10. Consideration of delay-based panning and other stereo effects

As previously mentioned, techniques exist other than amplitude based panning for changing the stereo image, particularly in post-production, and their use in the algorithm should be considered.

3. ALGORITHM

A simplified block diagram of the entire system is shown in Figure 1:

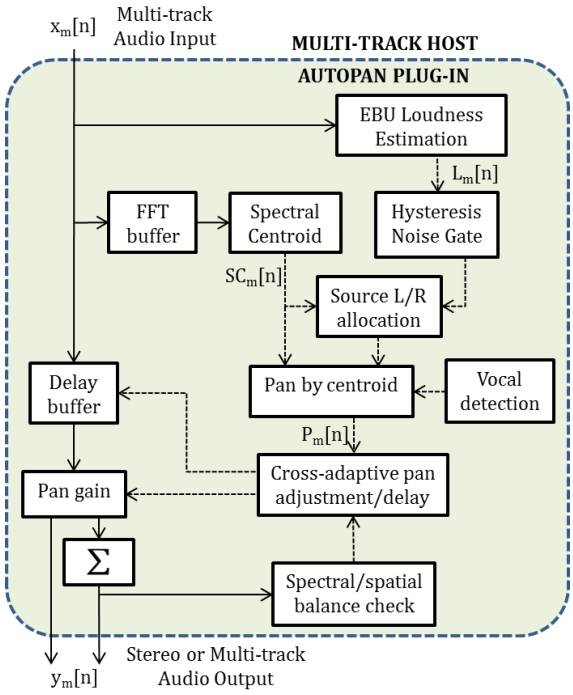

Figure 1: Full system block diagram.

3.1. Exponential Moving Average

To deal with the nature of short-frame real-time processing, the algorithm uses exponential moving average (EMA) filters extensively to provide smoothly varying data variables, as described in [12].

The EMA filter is a 1st order IIR filter, with the following difference equation, where α is a value between 0 and 1:

$$y[n] = (1 - \alpha) \cdot x[n] + \alpha \cdot y[n-1] \quad (1)$$

The value of α is adjusted according to the sample rate and frame size for a fixed filter response.

3.2. EBU Loudness R-128 and Hysteresis Noise gate

The algorithm makes use of the technique developed by the authors in the Autonomous Faders implementation [5] for determining the current active/inactive status of each track for every frame. Whilst a simpler signal level gate could be employed in this case, the intention is for the program to make use of the original calculation when the programs are combined.

A loudness value per frame is calculated using the EBU R-128 standard [13], which is an energy measurement on a signal processed by two biquadratic IIR filters. Filter coefficients adapt depending on sample rate to ensure a constant frequency response. A loudness measurement is calculated per frame and processed using an exponential moving average filter, to provide a smoothly varying loudness measurement for each incoming track.

A noise gate, with two loudness thresholds at -25 and -30LUFS (Loudness Units to digital Full Scale) and a hysteresis loop, provides a binary indication for each frame of whether a track is active or not. The hysteresis loop prevents excessive switching of state when the loudness level fluctuates above and below one threshold. Feature extraction and the control of the exponential moving average smoothing filters are determined by the noise gate, including starting smoothing when a track first becomes active.

3.3. Spectral Centroid

The spectral centroid is used to determine the 'centre of mass' of a spectrum, and provides a time-varying frequency value in Hertz (Hz) for each source every frame. The spectrum is calculated with a FFT. Because real signals are being analysed, only the first half of the spectrum need be calculated, due to the duplication above the Nyquist frequency. Spectral centroid is calculated as:

$$SC_m = \frac{\sum_{n=0}^{N/2-1} |X_m[n]| \cdot f[n]}{\sum_{n=0}^{N/2-1} |X_m[n]|} \quad (2)$$

where X_m represents the discrete Fourier transform of the m^{th} signal in the multi-track set, and $f[n]$ is the frequency represented in bin n.

The exponential moving average of the spectral centroid $SCema_m[n]$ is updated only when the noise gate determines the track to be active, preventing erroneous spectral centroid values being used.

The size of the FFT should be considered to obtain a sufficient frequency resolution to detect low or close frequency content. For a real-time plug-in implementation where the incoming frame size is controlled externally, a buffer accumulates a sufficient number of samples before calculating the FFT. The buffer size is chosen by considering bin width:

$$binWidth = \frac{f_S}{N} \quad (3)$$

where f_s is the sampling frequency in Hz and N is the FFT size. Assuming a maximum f_s of 192kHz, a frame size of 2048 or above (providing a maximum spacing of 62.5Hz) is considered sufficient.

3.4. Source left/right allocation

As a track enters the mix for the first time a decision is made as to whether the source should be dominant in the left or the right channel. This is then fixed to ensure a source cannot cross over from one channel to the other:

$$P_m[n] = \begin{cases} 0 & for\ left \\ 1 & for\ right \end{cases} \quad (4)$$

The decision made is to use the opposite polarity to the channel with the closest spectral centroid, provided it hasn't been selected as a lead channel. This way, the distance between sources with similar spectral content is maximised and spectral masking is reduced.

The mean left/right weighting of all panning positions is checked after the addition of each new

track. As a safety measure, in the event of a poor distribution (determined as when the mean strays beyond a tolerance of 0.2 from the centre), the system will automatically reset.

3.5. Frequency scaled pan locations

With the dominant channel decided the degree of panning applied to each source is scaled according to its spectral centroid, the maximum spectral centroid value of all sources and the overall panning width, to produce a panning factor for each track. A custom exponential curve, shown in Figure 2, determines the source distribution.

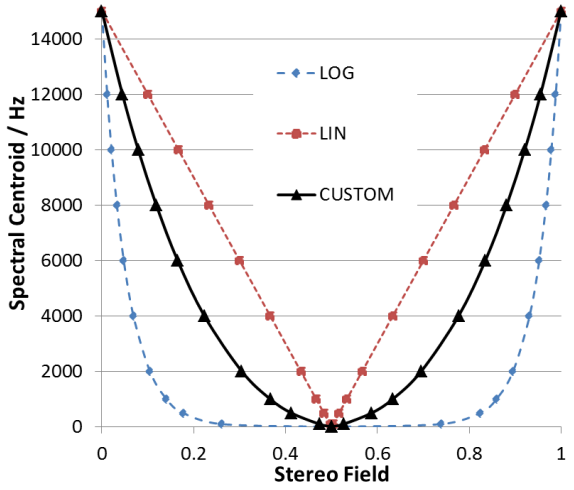

Figure 2: Graph showing pan positions in stereo field depending on frequency. Simple log and linear lines were trialed before a customised curve was chosen.

3.5.1. Maximum spectral centroid value

As a method to prevent stereo spread imbalance, the maximum spectral centroid value of all tracks is stored and updated over time. This allows the panning ratio to adapt according to the frequency range, and allows the full panning space to be utilized when the full frequency spectrum is not, as shown in Figure 3.

3.5.2. Panning Width

The panning width is a user-controlled value between 0 and 10 to extend or restrict the width of all sources, set to 5 by default for fully autonomous use. It works by moving the maximum spectral centroid value using a weighting of one third of the *SCmax* value, to adjust the angle of each track appropriately, also shown in Figure 3.

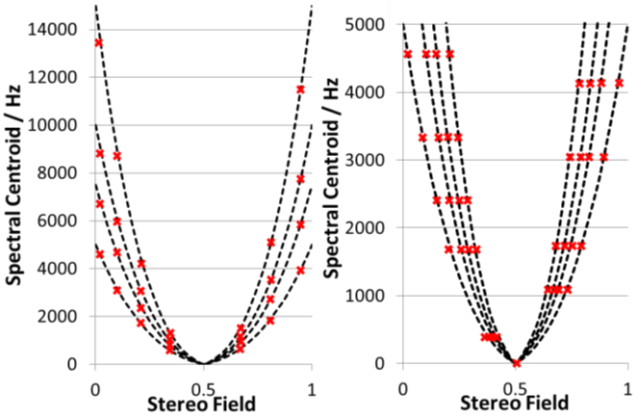

Figure 3: Panning positions on the left show consistent spread for different values of maximum spectral centroid, and on the right show the effect of the panning width control.

3.5.3. Panning Factor

The final panning factor is defined as:

$$Pf_m[n] = \left(\frac{\log(SC_m[n])}{\log\left(SCmax + \left((10-PW)\cdot\frac{SCmax}{3}\right)\right)} \right)^4 \quad (5)$$

where $SC_m[n]$ is the spectral centroid of each track, $SCmax$ is the maximum spectral centroid calculated from all tracks, and PW is the panning width factor between 0 and 10.

This is applied to the $P_m[n]$ starting value of 0 or 1 to give a panning position centered on 0.5:

$$P_m[n] = (Pf_m \cdot (2 \cdot P_m[n] - 1) + 1)/2 \quad (6)$$

3.5.4. Exceptions

Exempt from these general panning rules are designated lead tracks and tracks with a spectral centroid below the low frequency cut-off point, set nominally to 200Hz. In these cases the pan position is fixed at 0.5 in the centre of the panning space.

3.6. Balancing the mix

The processes detailed above give appropriate positions for the different sources within the mix, which will remain approximately static throughout assuming reasonably constant spectral centroid readings. As such, the mix should be reasonably balanced already, and require only optional minor tweaks to pan positions. However, balancing takes on particular importance as the dynamics of the mix change over time; when tracks drop out or come in, for example.

At all times the mix is tending to the maximum possible use of the panning space, up to the limits set by the panning factor. For this reason, balancing will only involve pulling sources inwards towards the centre and not pushing them further towards the extremes. Once balance has been achieved the mix will attempt to move the sources back to their original static positions.

3.6.1. Spectral Balance

The aim of the spectral balancing is to maintain the left/right balance across the entire frequency spectrum.

A 5-band approach is used, where the FFT of the left and right panned master channels are calculated, and the complex magnitude taken. Centre frequencies of 750, 1650, 3650, 7750 and 16000 Hz are used to cover the audible frequency spectrum. The spectral balance angle per band is calculated as the inverse tangent of the sum of magnitudes:

$$SpecB_b = \tan^{-1}\left(\sum_{k=K_b}^{K_{b+1}-1}|L[k]|/\sum_{k=K_b}^{K_{b+1}-1}|R[k]|\right) \quad (7)$$

where L and R are the FFT data from the left and right channels, b is the band between 1 and 5, k is the FFT bin and K_b is the starting bin number for each band.

The aim is to converge each of the spectral balance values to 0.5, i.e. the centre of the mix. A tolerance of 0.05 is allowed, meaning balancing will only occur on a band where $0.45 < SpecB_b < 0.55$.

For each band requiring balancing, sources are ordered by their distance from the centre frequency. Only sources on the higher-weighted channel within a certain bandwidth from the band centre frequency (using a Q-factor of 0.5) are moved. The ordering of the sources affects to what extent they are moved, with the closest sources moved the most using the following factor:

$$P_m[n] = P_m[n] + \left(dir \cdot G_{SB} \cdot \left(M_A - \frac{index_m}{M_A}\right)\right) \quad (8)$$

where M_A is the number of tracks which have become active, *index* is the source's place in the distance array and ranges from 0 (closest source) to M_A, *dir* is set to either 1 or -1 to ensure the sources move in the desired inward direction, and G_{SB} is the movement factor and is fixed by default to 0.3.

3.6.2. Spatial Balance

Spatial balance is defined as the inverse tangent of the peak signal level ratio from the left and the right channels.

$$SpatB_k = \tan^{-1}(|y_r|/|y_l|) \quad (9)$$

Similarly to the spectral balance, the aim is to converge at the 0.5 centre position, when $0.45 < SpatB_b < 0.55$. In that case, all active sources (with the exception of designated lead tracks or sources with a spectral centroid below the low frequency threshold) on the channel with the higher weighting are moved inwards by the same small factor.

It was shown experimentally that a byproduct of spectral balancing is the balancing of the mix spatially as well, making the latter process frequently unnecessary.

3.7. Stereo balancing

The proposal so far is for the placement of monaural (mono) sources. To this end, sources with existing stereo information can be mixed down to mono for replacement in the stereo field. However, this may not always be desirable. Stereo sources, recorded from coincident microphones or a stereo instrument like a piano, can contain useful stereo information that should be maintained. In this situation, the pan pots become a tool for weighting the mix towards left or right, known as 'balancing'.

In this implementation stereo information is maintained by default, with the width adjusted by the pan pot weighting, applied according to the same spatial and spectral balancing rules as for mono sources.

3.8. Delay-based Panning

Delay-based stereo placement can be used instead of, or in conjunction with, amplitude panning. As the algorithm is based on a traditional pan-pot approach, only the slight ITD delay is used to emphasise the existing source placement, typically between 1 and 2ms [2]. Delay is chosen depending on the pan pot position, with a linear relationship up to 2ms from mono to hard-panning, as described by the following equation:

$$\tau_m[n] = Pf_m[n] \cdot 2 \times 10^{-3} \quad (10)$$

Where $\tau_m[n]$ is the time delay in seconds applied to the m[th] track.

Automating the use of time delays in the algorithm is still work in progress, and whilst it is built into the software, by default the option is not currently used.

3.9. Other Stereo Effects

In addition to the amplitude and delay-based approaches there are numerous stereo effects which can be applied to both mono and stereo inputs. These include applications of reverb and chorus to provide depth, and width adjustment techniques, for example hard-panned double-tracked delayed sources. While these can provide interesting and effective additions to the stereo mix, they are largely used for artistic decisions and their use presents additional complexities to an autonomous algorithm.

Further research is required to establish rules for the use of stereo effects. A basic implementation of the double-tracked delay technique has been built into the software, which from preliminary testing has provided interesting additions to the final mix. As above, however, the option will not be in use until automation rules have been determined.

3.10. Pan Processing

There are numerous panning laws determining the ratio of spreading signal power between left and right channels. The most common is the sine/cosine -3dB pan law, which has the property of equal power from left to right, shown in Figure 4.

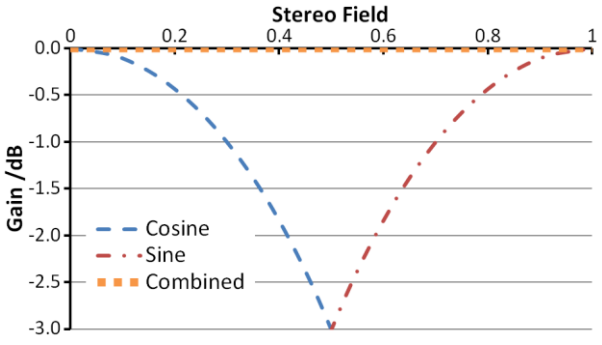

Figure 4: Equal power sine-cosine -3dB panning law.

This law is used to place the sources in the stereo mix, using the following equations:

$$y_L[n] = \cos\left(P_m[n] \cdot {}^{\pi}/_{2}\right) \quad (11)$$

$$y_R[n] = \sin\left(P_m[n] \cdot {}^{\pi}/_{2}\right)$$

4. VST PLUG-IN

The algorithm has been built into a multi-track VST plug-in. Additional routines have been added to the algorithm for real-time use, including an expandable track count by monitoring input activity, adjustment of parameters, and reset and on/off toggling capability.

The user interface, shown in Figure 5, includes switches and controls for pan width and switching on/off, and visualisations to represent spectral and spatial balance, pan positions, and a goniometer to provide real-time feedback of the stereo activity. The goniometer coordinates are determined as shown in Equations 12 and 13, where *n* is the sample number of a circular buffer of stored output samples.

$$x_{coord}[n] = y_r[n] - y_l[n] \quad (12)$$

$$y_{coord}[n] = y_r[n] + y_l[n] \quad (13)$$

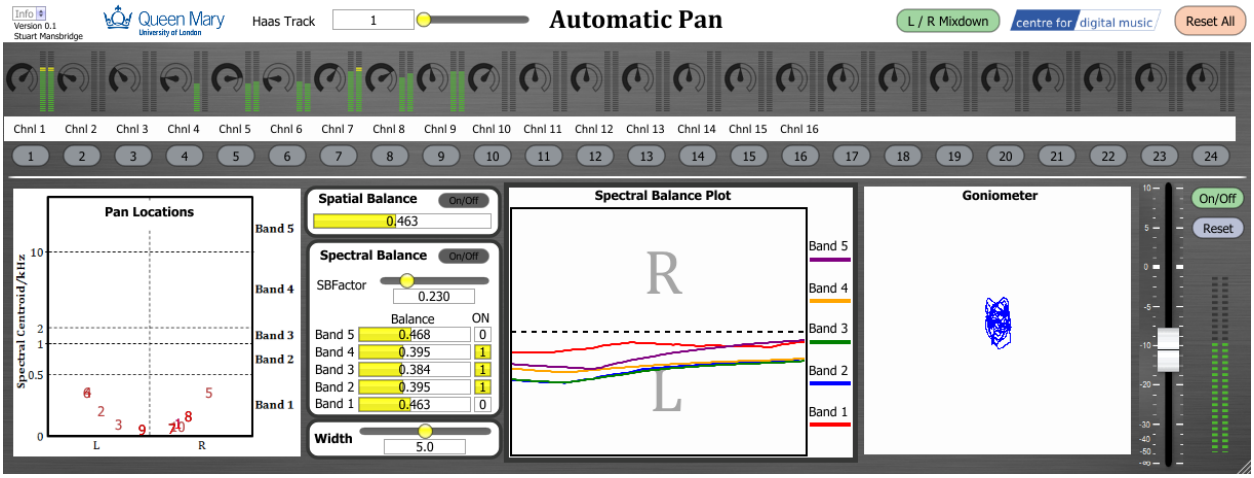

Figure 5: Screenshot of the Automatic Pan VST plug-in.

5. LISTENING TEST

A listening test was conducted to evaluate the system against professional mixes and across a variety of genres.

5.1. Method

A multiple stimulus with hidden anchor listening test was used for the subjective evaluation of the system. Similar to the MUSHRA framework [14] for perceptual audio evaluation, this allows audio content to be rated to an individual's preference against a specific criterion [15]. In this test an automatic mix was compared against three professional mixes. There was no reference included as there was no ideal mix; however a monophonic mix was used as a hidden anchor.

There were 11 participants in total for the audio evaluation. Table 1 shows the results of the preliminary test questions of participant's audio production and critical listening skills. All tests were conducted in an isolated listening room, with identical headphones, and a constant listening level.

Three sound engineers with experience in studio and live applications were asked to create mixes for the test-audio: one semi-professional with 5+ years' experience ('Eng. 1') and two professionals with 15+ years' experience ('Eng. 2' and 'Eng. 3').

In these mixes, the only parameter that was modified was panning position. The multi-tracks were raw but were loudness balanced appropriately so the engineers could focus solely on panning location. The engineers were asked to use a Digital Audio Workstation with a -3dB paw law, to correspond with the method used in the automatic pan system. However, 'Eng. 3' used a -2.5dB pan law.

Gender	Male	10
	Female	1
Audio Production experience descriptors and number of years of experience.	**Beginner**	**2**
	<5 years	*3*
	Competent	**3**
	>5 years	4
	Proficient	**2**
	>10 years	3
	Expert	**4**
	>15 years	1
Critical listening skills and details of experience.	**Beginner**	**1**
	Competent	**2**
	Proficient	**7**
	Expert	**1**
	Musician	4
	Music related training	3
	PhD related subject	4
Hearing Impairments	Yes (slight tinnitus)	1
	No	10

Table 1: Results of preliminary questions to test subjects.

As the system performs time-varying pan positioning, the engineers were instructed to use automation where they thought appropriate. 'Eng. 1'

and 'Eng. 3' created studio mixes where they were able to listen to and make changes any number of times to their preference, 'Eng. 2' performed a live mix, where the mixes could only be listened through to once or twice before real-time decisions had to be made. This was done to explore the different approaches. All mixes were created in an appropriate studio environment and the engineers provided detail of location, software and hardware used.

There were six multi-tracks chosen with varying genres including: 'Funk/Rock', 'Reggae', 'Jazz/Folk', 'Opera', 'Alt. Pop' and 'Gothic Electro'. The multi-tracks were taken from the Sound on Sound 'Free Multi-track Download Library' [16]. Overall, the test-audio consisted of twenty-second excerpts of each song including three professional mixes ('Eng. 1', 'Eng. 2' and 'Eng. 3'), one auto-pan mix ('Auto') and a mono mix ('Mono').

5.2. Results

5.2.1. Question 1

For the first test participants were asked to rate the sound mixes in terms of their preference. The results are therefore entirely subjective. Figure 6 shows the mean with error bars displaying the 85% confidence intervals using the T-distribution.

The professional mixes rate consistently high throughout with the 'Auto' mix scoring similarly or just below. However the 'Auto' mix out-performs the professional mixes on the 'Reggae' track. 'Eng. 2' and 'Eng. 3" in particular perform consistently well, with 'Eng. 1' performing well throughout and even outperforming all other mixes in the 'Jazz/Folk' and 'Opera' tracks but being least consistent overall. In 'Alt. Pop', 'Eng. 1' rates extremely low due to a corrupt audio file that had not been identified.

It can also be seen that the 'Mono' mixes are consistently rated lowest in all of the genres, with an exception for the 'Alt. Pop' song where it rates fairly high. This indicates a preference for a narrower stereo image for this song. The professional mixes that were rated highly had an audibly narrower stereo image compared to the 'Auto' mix, which was rated poorly. However this should be considered reflective of the individual song and not of the entire genre.

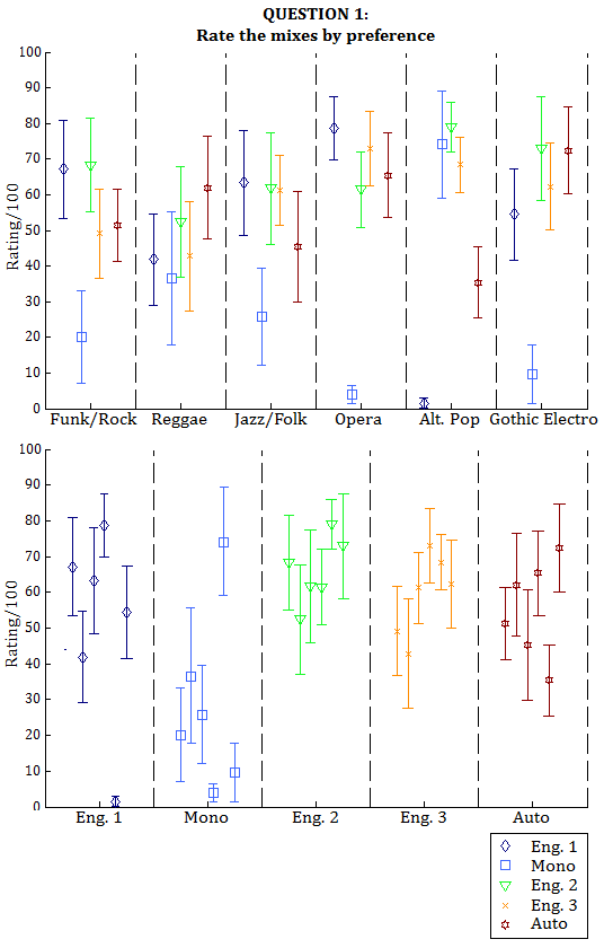

Figure 6: Mean and 85% confidence interval results for Question 1, for individual genre and mix type.

5.2.2. Question 2

In this test the participants were asked to "Rate the appropriate use of stereo mixing considering: placement and balance of sources, placement of frequency content in the mix between left and right channels, and balance of overall content in the mix between left and right channels." Results are shown in Figure 7.

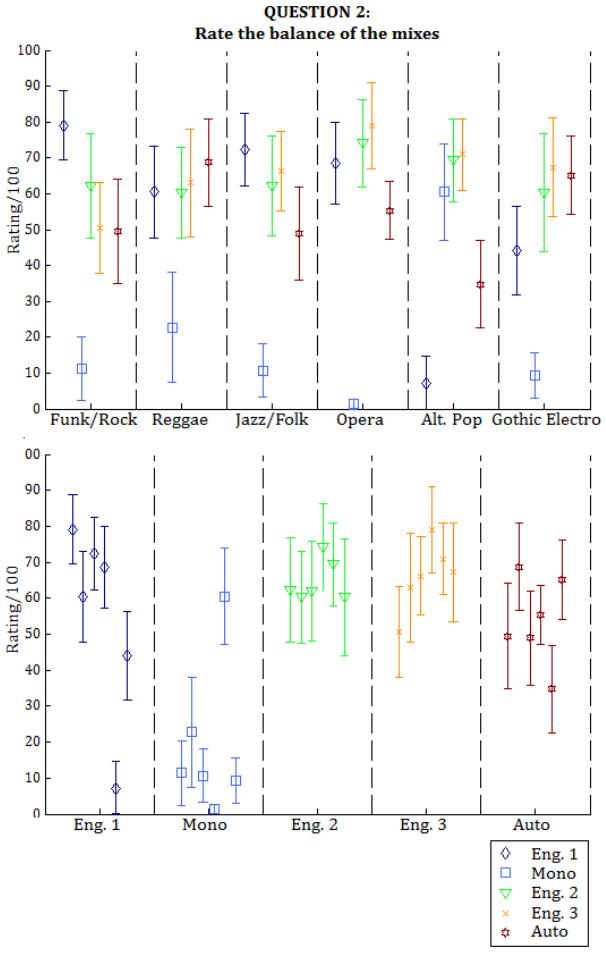

Figure 7: Mean and 85% confidence interval results for Question 2, for individual genre and mix type.

This question was designed to make the participants focus on how well the mixes met the panning constraints, particularly in terms of the balance of left/right content. The 'Mono' mix was used as a hidden reference to expel unreliable results [15].

Similar to the results in Question 1 the 'Mono' mixes were rated consistently poorly except for the 'Alt. Pop' song. Overall, the professional mixes are consistently rated highly with the 'Auto' mix just below.

Averaged mean and median results for all songs are displayed in Figure 8 for each mix type, and for both tests. These give a clearer depiction of the overall performance of each mix type.

It can be seen that 'Eng. 2' performs best in Q1 for the mean and median, and in Q2 for the median. 'Eng. 3' performs best in Q2 for the mean.

The averages differ because the mean is more affected by outliers, as shown in 'Eng. 1', as there is one very low score and generally much more varied results throughout. The median however takes the middle value and so is less affected by extremes and more by consistency, such as seen in 'Eng. 2' throughout Q1 and Q2.

With regard to the live and studio approaches to mixing, the live 'Eng. 2' mix performs most consistently overall, with the exception of the 'Eng. 3' studio mix for the mean of Q2. This was unexpected as 'Eng. 2' had less opportunity to modify decisions. However it could be due to personal experience as a professional live engineer and technical consistency.

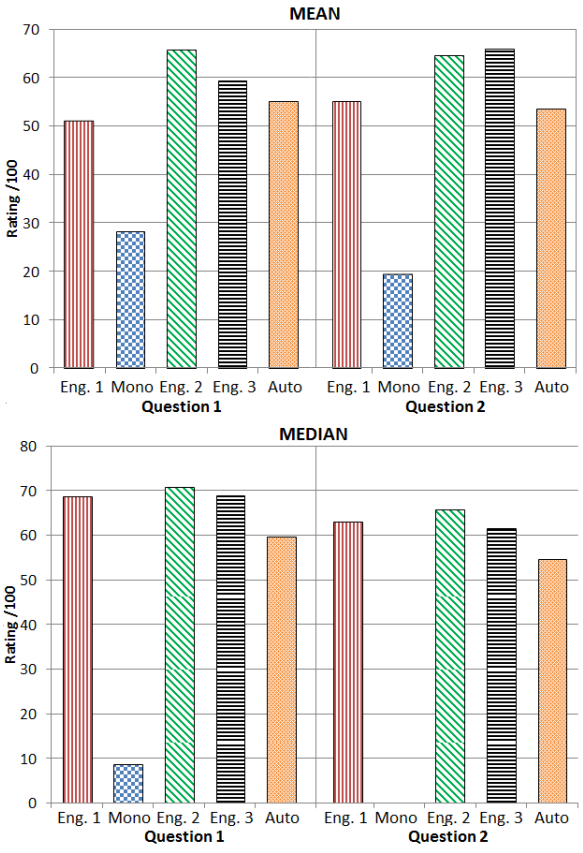

Figure 8: Overall mean and medium results for both tests.

5.3. Evaluation

Overall, the results show that the auto mix performs consistently in Q1 and Q2 across genres. Generally it rates just below the professional mixes, except in the 'Reggae' track which it out-performs, and the 'Alt. Pop' song where it performs badly. This indicates the generally successful application of the system across genres, rating closely to professional engineered mixes. With the exception of the 'Mono' mixes error bars are largely consistent throughout.

The results also indicate that the defined panning constraints are correct, due to the correlation between the results of Q1 and Q2. This shows that the system is closely following the approach that professionals take, which was a major objective at the start.

Generally the results were as expected. It was assumed that the professional mixes would out-perform the 'Auto', with the more experienced engineers 'Eng.2' and 'Eng. 3' performing above and 'Eng. 1' performing below or similarly. However, the most important result to highlight is the ability of the auto-pan to consistently work across multiple genres, falling only just below the standard of professionally engineered mixes.

A surprising result was that the 'Mono' tracks on occasion rated quite highly. It seems apparent that certain genres may benefit from a narrower stereo image, such as the 'Alt. Pop' and 'Reggae' tracks.

Despite the success of the listening test, it is recognised that it could be improved with more than six genres under test, and the use of more test subjects. There were originally 14 participants, of which 3 were excluded for rating the hidden anchor highly in Q2. It was found that these participants had also rated themselves as beginners in audio production.

6. CONCLUSIONS

6.1. Overview

A comprehensive new approach to the autonomous stereo positioning for music production has been presented, implemented and evaluated. It has applications in both live and off-line applications, and scored highly in a listening test in comparison with three professional mixes.

6.2. Future Work

A few limitations of the algorithm have become apparent during testing, and areas for future work are given below.

There is some reliance on a reasonable spacing between spectral centroids if there are more than two similar sources, as two sources will end up close together. In this case, often spectral balancing will move one of the sources to a more desirable location, but a more reliable solution should be investigated.

Also, the process of balancing stereo inputs requires more thought. As mentioned in [6,7], a major limitation of using one pan pot for balancing is the source will remain tied to one extreme although the width may be restricted, and so a solution using two pan pots may be appropriate.

Large variations in the range of spectral centroid values have been noted song to song, and while the self-adjusting maximum helps to deal with this, it is an indication that spectral centroid may not be the ideal measure of frequency content. Low frequency sounds seem to be worst represented, presenting problems with the use of the low frequency threshold. It is thought alterations to the spectral centroid calculation to eliminate noise would provide more suitable results, particularly in the high end of the spectrum at large sample rates. This could be achieved using a threshold to disregard FFT bins below a certain magnitude, or to calculate from a set number of harmonics.

The listening test indicated a preference of a narrower stereo image for certain genres. A genre-specific selection of the panning width control should be investigated. Vocal detection techniques to automatically determine lead tracks have been researched, but require further testing and implementation in the real-time algorithm.

Finally, whilst other forms of stereo positioning have been investigated in this paper, the panning algorithm presented has been based on the use of traditional pan pots only. Delay, phase and reverb stereo effects are often used in post-production,

typically sparingly, and a method of automating their usage needs to be determined to produce a tool that can replicate studio production.

7. ACKNOWLEDGEMENTS

This work was supported by the FP7 European Project DigiBIC. The authors would like to thank the sound engineers James Morrell, Pedro Pestana and Sebastien Fournier for their time in preparing the comparison mixes for the listening tests, and the volunteers from Queen Mary University of London and elsewhere who participated in the listening tests.

8. REFERENCES

[1] V. Pulkki, T. Lokki, D. Rocchesso, "Spatial Effects," in *DAFX*, U. Zölzer, Ed., chapter 5, pp. 139–147, John Wiley & Sons Ltd, Chichester, UK, 2nd edition, 2011.

[2] P. White, "Improving Your Stereo Mixing" *Sound On Sound Magazine*, October 2002, http://www.soundonsound.com/sos/oct00/articles/stereomix.htm

[3] E. Perez Gonzalez and J. D. Reiss, A real-time semi-autonomous audio panning system for music mixing, EURASIP Journal on Advances in Signal Processing, v2010, Article ID 436895, p. 1-10, 2010.

[4] E. Perez Gonzalez, J. D. Reiss, "Automatic Mixing: Live Downmixing Stereo Panner", in *10th International Conference on Digital Audio Effects (DAFx-07),* Bordeaux, France, September 10-15, 2007.

[5] S. Mansbridge, S. Finn, J. D. Reiss, "Implementation and Evaluation of Autonomous Multi-Track Fader Control," in *132nd Audio Engineering Society Convention*, Budapest, April 26-29, 2012

[6] 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.

[7] R. Izhaki, "Panning," in Mixing Audio: Concepts, Practices *and Tools*, chapter 13, pp. 184–203, Focal Press/Elsevier, Burlington, Vt, USA, 1st edition, 2007.

[8] E. Benjamin, "An experimental verification of localization in two-channel stereo," in *Proceedings of the 121st Convention Audio Engineering Society*, San Francisco, Calif, USA, 2006.

[9] R. Neiman, "Panning for gold: tutorials," *Electronic Musician Magazine*, 2002, http://emusician.com/tutorials/emusicpanning gold/.

[10] 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.

[11] B. Owsinski, "Element two: panorama—placing the sound in the soundfield," in *TheMixing Engineer's Handbook*, chapter 4, pp. 20–24, Mix Books, Vallejo, Calif, USA, 2nd edition, 2006.

[12] J. D. Reiss, "Intelligent Systems for Mixing Multichannel Audio", *17th International Conference on Digital Signal Processing (DSP2011)*, Corfu, Greece, p. 1-6, 6-8 July, 2011.

[13] International Telecommunication Union. Rec. ITU-R BS.1770-2, "*Algorithms to measure audio programme loudness and true-peak audio level*". Geneva, 2011.

[14] International Telecommunication Union, "*Multiple Stimuli with Hidden Reference and Anchor*", ITU-R BS. 1534-1, 2003.

[15] T. Sporer, J. Liebetrau and S. Schneider, "Statistic of MUSHRA Revisited", in *Proc. of the AES 127 Convention,* 9-12 October, 2009.

[16] M. Senior. "*The 'Mixing Secrets' Free Multi-track Download Library*", http://www.cambridge-mt.com/ms-mtk.htm.