

AUTOMATIC NOISE GATE SETTINGS FOR MULTITRACK DRUM RECORDINGS

Dr. M. J. Terrell and Dr. J. Reiss

Center for Digital Music
Dept. Electrical Engineering
Queen Mary University London

Abstract

A method has been developed for automating the settings of a noise gate. The method has been applied to a kick drum track containing bleed from secondary drum sources and white noise. The optimal settings are found by maximising the signal to distortion ratio (SDR). The SDR has contributions from the distortion caused to the kick drum signal, and the residual bleed and noise. These two components are weighted, enabling the gate to be controlled by a single parameter. It is shown that the improvement in the SDR can be obtained when the two components of the SDR are approximated, enabling the optimal settings to be calculated from the noisy signal and a single kick drum hit. It is found that the optimal threshold is slightly above the peak level of the noise component of the signal.

1. INTRODUCTION

There are many situations in audio production where the target signal is masked to some extent by noise. The masking effect in itself is undesirable, but if processing is applied to the noisy signal, additional problems can arise. In multiple microphone conferencing settings it is undesirable to amplify background noise, and it is possible for feedback loops to be formed with even low levels of sound. Live musical recordings are made in an environment containing multiple sources and microphones. Rather than random noise, the target source will be masked by interference from a secondary source, known as bleed. In all cases it is preferable to reduce the level of the noise or the bleed signal, whilst minimising any distortion of the target signal.

A non-linear, dynamic audio effect called a noise gate is generally used to reduce the amplitude of the noise. The transfer function of a dynamic audio effect is dependent on the input amplitude. Compressors are the most commonly used dynamic audio effect. Compression is present to some extent in all modern day recordings. The compressor reduces the dynamic range of the signal by applying an attenuation to those parts of the signal with an amplitude greater than some threshold. An expander is the opposite of a compressor, in that it increases the dynamic range. This is done by applying an attenuation the parts of the signal with an amplitude below the threshold.

A noise gate is an extreme example of an expander. A signal entering the gate which is below the threshold level is treated as noise. The gate will not open fully and will apply an attenuation to the signal (up to $-\infty$ dB). A signal entering the gate which is above the threshold level will cause the gate to open, allowing the signal to pass through unattenuated. Noise gates are used in a number of applications. They are used to gate ambient noise of microphones in conference environments, to cut the noise or hum from a heavily distorted guitar amplifier during a live performance, or in audio post production to remove breathing from a vocal track. The speed at which the gate opens and closes is determined by the dynamics of the gate. These are key parameters in determining any distortion to the target signal. If the gate opens too slowly it will cut off the start of the target signal (which at conference settings could make speech unintelligible). If the gate closes too slowly noise will be allowed to pass through.

The use of noise gates in conference settings has been investigated in the past, by Dugan [1]. Dugan identified the difficulties in setting a suitable threshold level, particularly when there are relatively high levels of

ambient noise, and there are speakers positioned at distance from the microphone. The effect of poor gate settings is the amplification of ambient noise, or distortion of the target signal. Methods to overcome these problems have been presented. In [2] it is proposed that an adaptive threshold be used, which is a function of the ambient noise. In [3] it is suggested that the gating mechanism is designed to be sensitive to the input frequency of the signal. Only an input signal in the vocal frequency range will open the gate. Julstrom et al. [4] suggested using two cardioid microphones to identify the direction of the sound source. Only sound sources which were from the correct direction would open the gate.

An area in musical recording where bleed is especially prevalent is in the recording of a drum kit. Although modern microphones are directional, and steps can be taken to isolate sources, there will inevitably be some bleed as both sources and microphones are in close proximity. Equalization can potentially be used to reduce the effect (for example a high cut filter will reduce the level of cymbal bleed on a kick drum microphone), but if the action taken is too aggressive it can have a detrimental effect on the target sound. Noise gates are commonly applied in post production to remove this type of bleed.

2. NOISE GATES

2.1. Noise Gate Parameters

A simple noise gate has four main parameters; threshold and gain which are measured in decibels and attack and release - which represent the dynamics of the gate - are measured in seconds. The threshold is the level above which the gate is opened and below which the gate is closed. The gain is the reduction in the signal level caused by the closed gate. A gate which stops signals below the threshold from passing through completely has a gain of $-\infty$. The attack is the time it takes for the closed gate to fully open once the threshold is reached. The release is the time it takes for the open gate to fully close once the signal level drops below the threshold. Some noise gates also have a hold parameter. This dictates a minimum time in which the gate

must remain in the current state, and thus prevents it from switching between states too quickly, which can cause unwanted audio artifacts.

2.2. Multitrack Drum Recordings

A simple drum kit set up will comprise kick drum, snare, hi-hats, cymbals and any number of tom toms. The general microphone setup will use a kick drum mic, a snare mic, a mic for each tom tom, and a set of stereo overheads to capture a natural mix of the entire kit. In some instances a hi-hat mic will also be used. When mixing the recording, the overheads will be used as a starting point. The signal from the other microphones is then mixed into this to provide emphasis on the main rhythmic components i.e. the kick, snare and tom toms. Processing is applied to these signals to obtain the desired sound. Gating is used to limit (or ideally remove) the level of the bleed sources before this processing is applied.

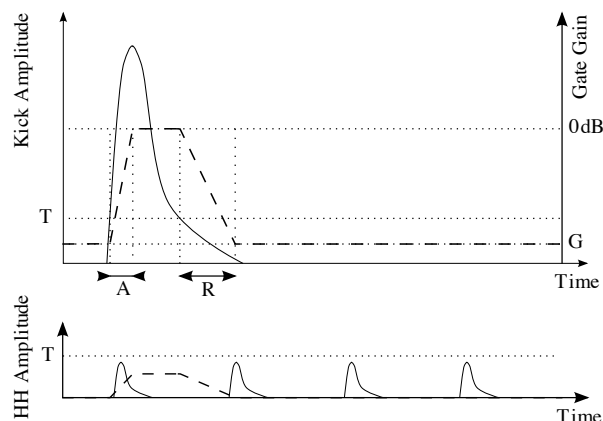


Figure 1: Kick Microphone Output with Hi-Hats Bleed

Figure 1 shows an example of the output of a kick drum mic with bleed from the hi-hats. The hi-hats signal has been plotted on a separate set of axes, and the gate gain envelope has been overlaid on both. If the gate threshold is above the peak amplitude of the hi-hats bleed, then the third and fourth hi-hat hits will not open the gate, and will be silenced (or reduced by G dB). The first hit coincides with the kick drum hit, the hi-hat signal will also be allowed to pass through

the gate unaffected. In this instance it is not possible to remove the bleed source. The second hi-hat hit coincides with the end of the release phase of the gate. If the release time is short the gate will close fully before the hi-hat hit, but the natural decay of the kick drum will be distorted. If the release time is long the gate will still be partially open, and the hi-hat will be audible to some extent, but the kick drum hit will be allowed to decay more naturally. It is necessary to strike a balance between removing the hi-hat and minimising distortion of the kick drum. If the gate threshold is below the peak amplitude of the hi-hat signals, then all hi-hat hits will open the gate.

3. METHOD

Audio files representative of a kick drum microphone containing bleed from hi-hats, snare drum, cymbal and tom toms, as well as white noise are investigated.

3.1. Audio Files

The audio files used for testing are sequenced by the author, using real drum samples, enabling the level of the bleed sources to be controlled. This results in a simulated kick drum microphone. The noise component of the signal is a combination of bleed sources: hi-hats, snare, tom toms, and cymbal; and white noise. The clean (kick) and noise signals are then combined to give a signal representative of a kick drum microphone masked by bleed and noise. A sample rate of 44,100 Hz is used. The length of each audio file is 4 bars, the tempo used is 120 bpm and the duration is 8 seconds. The drum pattern is shown in Figure 2.



Figure 2: Musical score of the drum pattern

The clean and noise signals are identified by Y_C and Y_N respectively. The peak amplitudes of signals Y_C and Y_N are -0.5 dB and -9.2 dB respectively. The total signal is identified by the subscript S ,

$$Y_S = Y_C + Y_N. \quad (1)$$

3.2. Signal Distortion

Vincent et al. [5] proposed performance measures to be used in source separation algorithms. There are three components to the measure of distortion; interference from sources other than the target, noise, and artifacts which result from the source separation algorithm. In [5] the total distortion is defined as,

$$D_T = \frac{\|\hat{s}[n]\|^2 - \|\hat{s}[n].s[n]\|^2}{\|\hat{s}[n].s[n]\|^2}, \quad (2)$$

where $\hat{s}[n]$ is the approximation of the target signal $s[n]$. The approximation of the target signal is projected onto the target signal. If the two signals correlate exactly then the distortion is zero. If the signals are orthogonal then the distortion is infinite. When applying this distortion measure to noise gates, the distortion of the signal will be limited to the transient regions of the gate. This amounts to a very small percentage of the signal. As a result the approximate and target signals will always have a strong correlation, and differences in the total distortion may be hard to gauge. For this reason the following more classical distortion measure is used,

$$D_T = \frac{\|\hat{s}[n] - s[n]\|^2}{\|s[n]\|^2}. \quad (3)$$

The equivalent parameters in this paper are,

$$\hat{s}[n] \equiv f[n]^T Y_S, \quad (4)$$

and,

$$s[n] \equiv Y_C, \quad (5)$$

where the function $f[n]$ is the resultant gate function from the input Y_S , and the current gate parameters. The total error vector of the signal, $e_T[n]$ is given by,

$$e_T[n] = f[n]^T Y_S[n] - Y_C[n]. \quad (6)$$

Following [5], the error is split into three components, interference, noise and artifacts. The interference is analogous to the bleed, and the noise is analogous to the white noise added to the signal. The interference and noise are grouped into a single noise component here. The artifacts are the residual effects introduced by the process itself. This is analogous to the distortion of the clean signal caused by the gate. The total error is therefore a combination of the artifact and noise errors,

$$e_T[n] = e_A[n] + e_N[n]. \quad (7)$$

Using Equations 1 and 6,

$$e_A[n] = f[n]^T Y_C[n] - Y_C[n], \quad (8)$$

and,

$$e_N[n] = f[n]^T \cdot Y_N[n]. \quad (9)$$

From these error functions, the distortion due to the artifact error is given by,

$$D_A = \frac{\|f[n]^T Y_C[n] - Y_C[n]\|^2}{\|Y_C[n]\|^2}, \quad (10)$$

and the distortion due to noise is given by,

$$D_N = \frac{\|f[n]^T Y_N[n]\|^2}{\|Y_C[n]\|^2}. \quad (11)$$

The signal to distortion ratio (SDR), the signal to artifact ratio (SAR) and the signal to noise ratio (SNR) are defined as follows:

$$\text{SDR} = 10 \log_{10} D_T^{-1}, \quad (12)$$

$$\text{SAR} = 10 \log_{10} D_A^{-1}, \quad (13)$$

$$\text{SNR} = 10 \log_{10} D_N^{-1}. \quad (14)$$

3.3. Initial SDR measurements

Fast attack times will maximise SAR, and will not have a significantly detrimental effect on the SNR. The attack time of the gate is therefore set to the minimum setting of $1ms$. In order to maximise SNR, the noise level should be reduced as much as possible. The gain parameter of the noise gate is therefore set to $-\infty dB$. Figure 3 shows a plot of the SDR over a range of release and threshold settings. It can be seen that there is a maximum $\text{SDR} = 12.6dB$ (increased from $9.6dB$), achieved with $T = -9.03dB$ and $R = 0.225s$. The optimal threshold is slightly above the peak amplitude of the noise ($-9.2dB$).

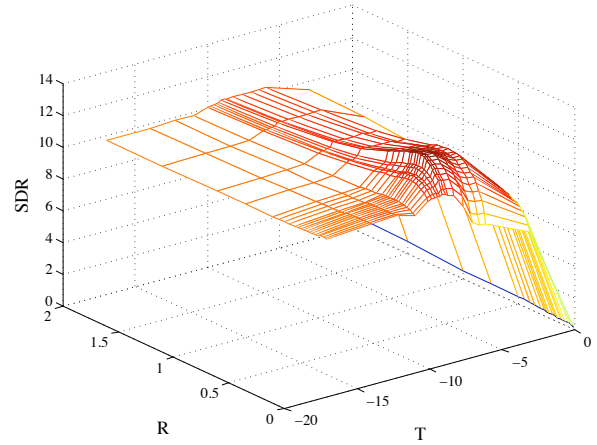


Figure 3: SDR over a range of release and threshold settings.

3.4. Controlling the strength of the gate

The noise gate settings which yield the maximum SDR have been found. This does not necessarily mean, however, that the settings are the same as those that would be used by an engineer. There is a subjective element to the choice of settings; it may be the case that all noise must be removed resulting in a stronger gate. It may also be the case that there must be minimal distortion to the kick signal resulting in a gentler gate. To account for this, the components of the total error function given in Equation 7 are weighted,

$$e_T[n] = (1 + W)e_A[n] + (1 - W)e_N[n], \quad (15)$$

where,

$$-1 \leq W \leq 1. \tag{16}$$

When $W=-1$, noise distortion only is measured, and the peak on the SDR plot will represent a closed gate. Conversely when $W=1$, artifact distortion only is measured, and the peak on the SDR plot will represent an open gate.

Columns 2 and 3 of Table 1 contain the noise gate settings which give the maximum SDR. The improvement in SDR can be seen in Figure 4, identified by SDR_{CE} . It can be seen that when the weighting parameter is negative the gate is strong (high threshold) and minimal signal or noise is allowed to pass through. Conversely for positive values of the weighting parameter the gate is open, and all signal and noise is allowed to pass through. In the mid range of the weighting parameter the threshold has converged to a level slightly above the peak of the bleed component of the noisy signal. Within this range of $W = -0.4 : 0.4$, the release parameter slowly increases. This will cause less unwanted distortion of the target signal, but will allow more noise to pass through as the gate closes more slowly. In all cases the SDR has increased. As W increases, the noise component of the SDR reduces (for both the gated and ungated signals), and as a result the increase in SDR after gating is reduced.

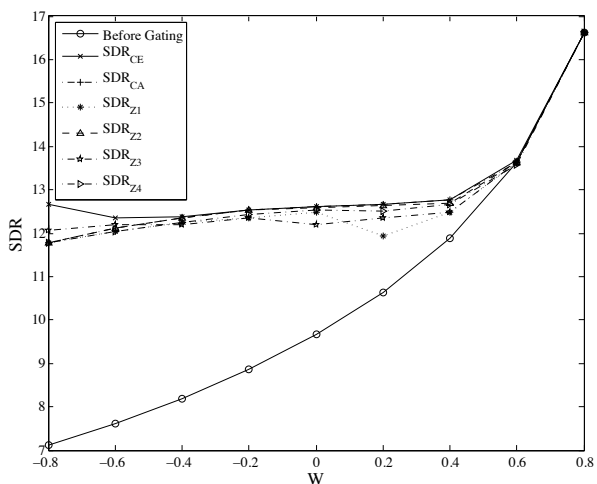


Figure 4: Comparison in the improvement of the SDR when using the exact and approximate error vectors

Table 1: Comparison of the optimal noise gate settings when using exact and approximate error vectors.

W	R_{CE}	T_{CE}	R_{CA}	T_{CA}	R_{Z1}	T_{Z1}
-0.8	0.025	-6.48	0.150	-9.03	0.150	-9.03
-0.6	0.125	-6.99	0.150	-9.03	0.175	-9.03
-0.4	0.150	-9.03	0.175	-9.03	0.200	-9.03
-0.2	0.175	-9.03	0.175	-9.03	0.250	-9.03
0	0.225	-9.03	0.225	-9.03	0.350	-9.03
0.2	0.300	-9.03	0.300	-9.03	0.350	-9.79
0.4	0.400	-9.03	0.400	-9.03	0.500	-9.79
0.6	0.250	-14.0	0.500	$-\infty$	1.500	-9.79
0.8	0.010	$-\infty$	0.500	$-\infty$	1.500	$-\infty$

W	R_{Z2}	T_{Z2}	R_{Z3}	T_{Z3}	R_{Z4}	T_{Z4}
-0.8	0.150	-9.03	0.200	-6.99	0.150	-9.03
-0.6	0.150	-9.03	0.225	-6.99	0.175	-9.03
-0.4	0.175	-9.03	0.250	-6.99	0.200	-9.03
-0.2	0.175	-9.03	0.250	-9.03	0.225	-9.03
0	0.200	-9.03	0.350	-9.03	0.275	-9.03
0.2	0.250	-9.03	0.450	-9.03	0.350	-8.86
0.4	0.350	-9.03	1.500	-9.03	0.450	-8.86
0.6	0.010	$-\infty$	1.500	$-\infty$	1.500	-8.86
0.8	0.010	$-\infty$	1.500	$-\infty$	1.500	$-\infty$

3.5. Working Blind

The work presented so far makes the assumption that the clean signal is available when calculating the SDR. This enables $e_A[n]$ and $e_N[n]$ to be evaluated directly. In practical situations this is never the case.

Noise gate parameters are defined subjectively by the sound engineer. The human ear and brain are able to distinguish between what is signal and what is noise. As a result it is possible to judge suitable relative levels of D_A and D_N . If the gate parameters are to be set automatically it is necessary to identify regions of signal and noise when only the noisy signal is available. When considering the equivalent human operation, the sound engineer will have prior knowledge of what the clean signal sounds like, i.e. the sound engineer will know that the clean signal is a kick drum. For this reason the blind method will also have available an input, which identifies what the signal sounds like. The additional input will be a single kick drum hit.

In order to estimate the artifact and noise error it is necessary to split the noisy signal into regions which contain the kick drum, and regions which contain only noise. Work has been done previously on automatic drum transcription and source separation. Onset de-

tection functions are used to find prominent beats in a passage of music, enabling tempo to be identified. This is demonstrated by Yoshii et al. in [6]. A summary of drum transcription and source separation techniques can be found in FitzGerald's Ph.D thesis, [7]. Recent work on drum transcription has been focussed on extracting drum patterns from polyphonic audio files. In [8], Tzanetakis et al. used onset detection functions, split into low and high frequency bands to detect kick and snare drum events. Gillet and Richard [9] removed the non-rhythmic component of the signal and performed transcription on the residual signal. The situation presented here is far simpler, as there is no non-rhythmic component to mask the signal, and the target signal is of a significantly higher amplitude than the noise.

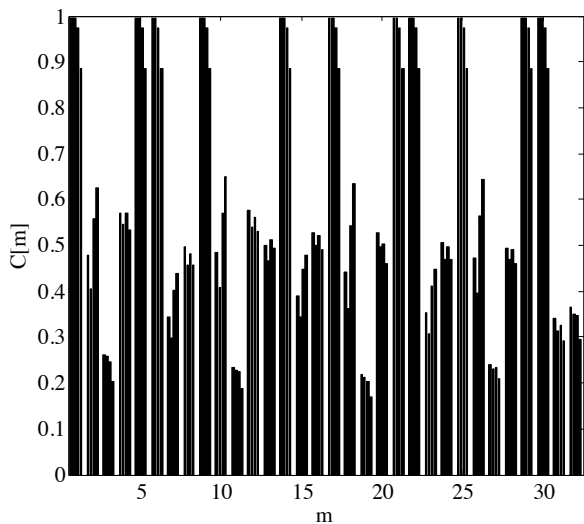


Figure 5: Correlation of a single kick from each layer of the sampled drum instrument with each section of the noisy signal

A simple method to extract the kick drum pattern is used here. The noisy signal, Y_N is split into m discrete windows. The spectral power of each window is correlated with the spectral power of the single kick drum hit. High correlation is obtained in windows that are predominantly kick drum. All other windows are attributed to noise. This is particularly applicable with drum recordings as each drum has a significantly different spectral content. The power spectrum of the single kick is denoted as P_Z , the power spectrum of window m of the noisy signal is denoted by $P_N[m]$ and

the correlation between the two vectors is denoted by $C[m]$. If $C[m]$ is greater than some threshold, window m is assigned to signal, otherwise it is assigned to noise. The noisy signal is split into 32 windows which corresponds to 8th notes (semi-quavers), and matches the quantization of the recording. Figure 5 contains a bar chart which shows the correlation of five different kick drum hits (taken from the sampled instrument used to generate the clean and noisy signals), with each window of the noisy signal. Strong correlation is found for all windows which coincide with sections of the noisy signal that contain a kick drum hit. The threshold for correlation is set at 0.8.

The single kick drum hit is used to approximate the clean signal, by aligning a kick drum hit with each window of the noisy signal which had a correlation above the threshold. Each hit is scaled to have the same peak amplitude as the corresponding point in the noisy signal. The approximation of the clean signal is called the synthesized signal, and is denoted by Y_Z . The resultant noise - which contains all windows of the noisy signal with a correlation below the threshold - is denoted by Y_R .

3.5.1. Artifact Error Approximation

The synthesized signal is passed through the noise gate which results in the gate function $z[n]$. From this, an approximation to $e_A[n]$ can be made,

$$e_A[n] \simeq z[n]^T Y_Z[n] - Y_Z[n], \quad (17)$$

If the drum samples used to build the synthesized signal are significantly different to the actual signal, the approximation of the error vector Equation 17 will be less accurate. However, the trend in the changes will always be the same regardless of the sample used, as a stronger gate will always cause more distortion. In live situations a single kick sample can be obtained from the drum kit during sound check. For post production applications a single kick can be recorded before or after the main performance. Alternatively a sample could be taken from a library of drum samples.

3.5.2. Noise Error Approximation

The gate function $f[n]$ is applied to the residual noise signal, to gain an estimate of the noise error vector,

$$e_N[n] \simeq f[n]^T Y_R. \quad (18)$$

The key to getting good noise gate settings is to ensure that the attack part of all bleed drum hits are captured in the approximation of the noise component of the signal. The examples here have been applied to a quantized drum beat. Discretising a quantized signal into m windows will yield divisions exactly on beat boundaries (if m is chosen sensibly), hence the attack phase of each bleed hit is retained. Many modern recordings use a quantized grid, and the use of advanced production software, such as Pro-Tools, means that the rigid quantized grid is adhered to (i.e. it is not entirely dependent on the drummer's ability!). In situations where quantization is not used, more sophisticated drum transcription techniques, similar to those mentioned in 3.5 are needed. For example, rather than splitting the signal into equal length windows, the signal can be split using identified beat onsets. Each window can then be assigned to either signal or noise.

4. RESULTS

Table 1 contains the optimal noise gate settings. The subscripts CE and CA identify cases where the clean signal is available, and the exact and approximate error vectors equations are used respectively. The subscript Z identifies cases where the clean signal is synthesized. Synthesized signals generated from four different kick drum hits were tested. The improvement in the SDR is shown in Figure 4. For all cases other than SDR_{CE} , the improvement has been calculated using the optimal gate settings found using the approximate error vectors (Equations 17 and 18), substituted into the exact equations for the error vectors (Equations 8 and 9).

5. DISCUSSION

Figure 4 shows the improvement in the SDR before and after gating. It can be seen that the improvement

in the SDR is similar for all synthesized signals, and that the improvement is very close to that seen when the clean signal is available, and the exact equations for the error vectors are used. The trends in the optimal settings over the range of W are similar for all synthesized signals. At extreme values of W the gate is either fully open or fully closed. At intermediate values the threshold is slightly above the peak noise level, and the release time gradually increases as the W is increased, and the gate is made gentler. The release settings have been transposed over the range of W for the synthesized signals. This is because the artifact error is different in each case, and the balance between it, and the noise error has been altered.

For lower values of W the gate is made stronger. The amplitude envelope of the synthesized signal will determine whether the artifact error is effected more by a fast release, or a high threshold. A kick drum hit with a fast natural decay will suffer less distortion from a low threshold and fast release gate when compared to a kick drum hit with a slow natural decay time. The artifact error of synthesized signal Z_3 was lessened by increasing the threshold, rather than reducing the release time as was the case for all other signals.

When the approximate error vectors are used, the noise is apportioned to discrete sections of the audio file. No account is taken of the noise which overlaps the signal. When the strength of the gate is increased, a point will be reached where the noise has been removed completely. Reducing W beyond this point will not result in a stronger gate. There are differences in the release parameter at gentle gate settings ($W=0.8$). Some of the synthesized signals have a long release time e.g. $R_{Z1} = 1.5s$ whilst others have a short release time e.g. $R_{Z2} = 10ms$. Whilst these differences seem stark, there is no difference in the effect of the gate, as it is completely open in each case, allowing all signal and noise to pass through.

6. CONCLUSIONS

The parameters of the noise gate have been reduced to a single parameter, W . A low value of W corresponds to a strong gate, and a high value corresponds to a gentle gate. It has been demonstrated that the clean sig-

nal and the regions of noise can be approximated, and that the improvement in SDR found when using the approximations is close to that found when the clean signal and noise are known exactly. The precise selection of W is a subjective matter, but indications of the best setting can be found. The method presented gives a bound for the maximum strength of the gate; the point where the discretised noise regions have been silenced. This would be a good suggested starting point. If the distortion of the signal is too strong then W can be increased gradually until it is at acceptable levels.

7. FUTURE WORK

It has been demonstrated that for a quantized drum beat a coarse method to approximate the clean signal and the noise regions is adequate. Further investigation is required for signals which do not conform to a quantized grid. For multitrack recordings, additional information could be extracted from other microphones to improve the accuracy of the estimate of the signal and noise regions. The weighting parameter is used to control the strength of the gate. To make the gate fully automated, the optimal weighting parameter for various types of drum recording should be identified. Future work will involve listening tests to determine suitable settings for this parameter.

8. REFERENCES

- [1] D. Dugan, "Applications of Automatrix Techniques to Audio Consoles", *Proceedings of the 87th AES Convention, New York, U.S.A., October 1989*.
- [2] G. W. Reichard, Jr., R. L. Breeden, "The 4A Speakerphone - A Hands-Down Winner", *Bell Labs. Rec.*, p. 233, September 1973.
- [3] Manufacturer's literature, Shure Brothers Inc., "Models M625, M625-2E, and M625M Voice-gate", *Data Sheet 27A1087*, 1977.
- [4] S. Julstrom, T. Tichy "Direction-Sensitive Gating: A New Approach to Automatic Mixing", *Journal Audio Engineering Society, Vol 32, No 7/8, July /August 1984*.
- [5] E. Vincent, C. Fevotte, R. G. L. Benaroya, "Performance measurement in Blind Audio Source Separation", *IEEE Trans. Audio, Speech and Language Processing*, 14(4), pp 1462-1469, 2006..
- [6] K. Yoshii, M. Goto, H. G. Okuno, "Automatic Drum Sound Description for Real-World Music Using Template Adaptation and Matching Methods", *The Proceedings of the 5th International Conference on Music Information Retrieval*, 2004.
- [7] D. FitzGerald, "Automatic Drum Transcription and Source Separation", *Ph.D Thesis presented to the Dublin Institute of Technology*, 2004.
- [8] G. Tzanetakis, A. Kapur, R. I. McWalter "Subband-based Drum Transcription for Audio Signals", *The 7th IEEE Workshop on Multimedia Signal Processing*, October 2005.
- [9] O. Gillet, G. Richard, "Drum Track Transcription of Polyphonic Music Using Noise Subspace Projection", *The Proceedings of the 7th International Conference on Music information Retrieval*, 2006.