Reducing comb filtering on different musical instruments using time delay estimation

Alice Clifford and Josh Reiss

Queen Mary, University of London

alice.clifford@eecs.qmul.ac.uk

Abstract

Comb filtering occurs when a signal and a delayed version of the same signal are mixed, for example when the signals from two microphones reproducing a single audio source are summed. This effect can be reduced by applying a compensating delay so there is ultimately no delay between the audio signals. This can be made automatic by using time delay estimation. This paper explores the effect on the accuracy of the time delay estimation when using bandwidth limited source signals, such as a variety of musical instruments with different frequency content. It is found that the smaller the bandwidth of the source signal, the less accurate the time delay estimation and comb filter reduction.

1   Introduction

When recording with a microphone it is important to choose the correct microphone for the task. Different makes and types of microphones have different frequency responses, causing spectral changes to the input source, and different polar patterns which determine the directivity. The choice can be creative, for example choosing a Shure SM58 because of its well-known peak at the vocal frequencies. It can also be a technical decision to use the microphone that will cause the smallest spectral change to the input signal. Even the aesthetic of the microphone plays a role in the decision process.
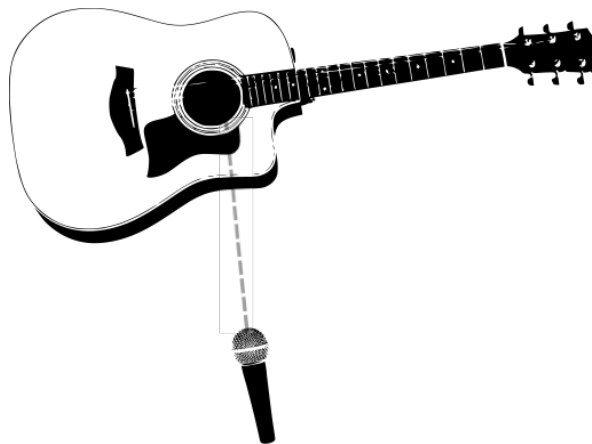

Figure 1 A single source being reproduced by a single microphone with the path of the direct sound indicated by a dashed line.
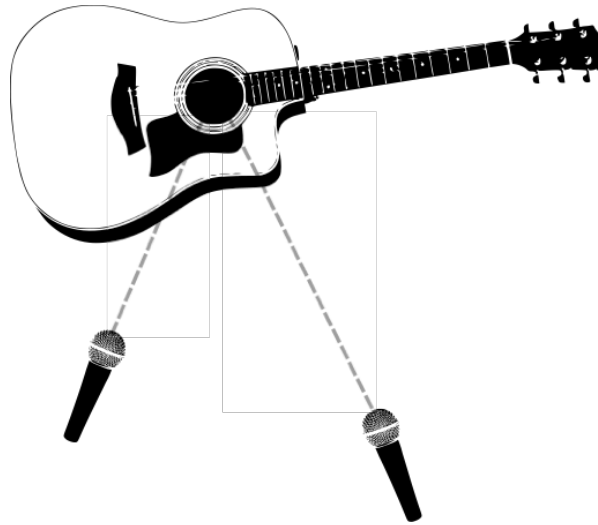
Figure 2 A single source being reproduced by 2 microphones with the path of the direct sound indicated by dashed lines.

The next task is then to place the microphone, which is also a creative and technical decision. The microphone may be placed to reproduce certain characteristics of a sound source but the placement can also cause other, unwanted artifacts to occur on the output of the microphone. The simplest configuration is a single microphone placed to reproduce a single source as shown in Figure 2. The source is placed within the microphone's pickup area, defined by its polar pattern. The microphone picks up the source with possible interference from reflections, exhibited as reverberation, or noise from other sources (Eargle, 2004).

A source, such as a musical instrument, will often reproduce very different sounds from different parts of the instruments. For example, the sound of an acoustic guitar from a microphone placed next to the sound hole will be different to the sound from a microphone near the neck. For this reason, multiple microphones can be used to reproduce different aspects of an instrument and mixed to produce the desired sound for that instrument. An example of this configuration can be seen in Figure 2.

Once more than one microphone is used in any configuration, other artifacts apart from usual reverberation and noise can occur. It is difficult, and often not desired, to place multiple microphones equidistance from the sound source. The result of this is that the time taken for the sound to travel from the source to each microphone, otherwise known as delay time, will be different. If these microphone signals are summed, either to create a desired sound for the source or mixed to a stereo output, the difference in delay time will cause the resulting output to be effected by comb filtering.

Comb filtering occurs when a signal is summed with a delayed version of itself, such as in the multiple microphone configuration. Cancellation and reinforcement of frequencies occur periodically between the two signals causing a comb shaped frequency response. The frequency response has distinctive peaks and troughs where the sound is either reinforced or cancelled, as can be seen in Figure 3. The resultant sound can be described as 'thin' and 'phasey' and is the basis of flanging and phasing effects (Zölzer, 2002). A

study in (Brunner et al., 2007) has shown that in subjective listening tests comb filtering can be heard when the delayed signal is as much as 18dB lower than the original.
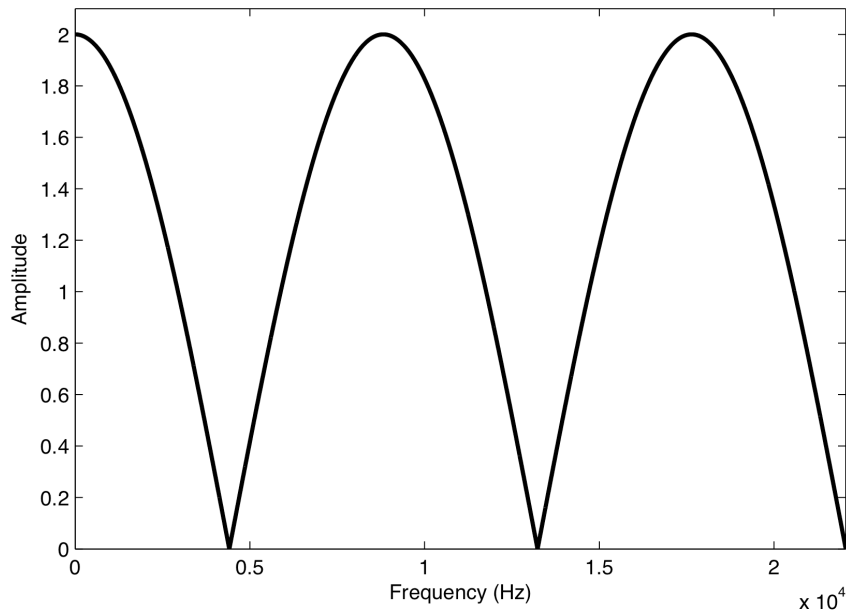

Figure 3 Typical frequency response of a comb filter.

Any configuration that will result in delayed versions of the same signal being mixed can cause comb filtering; such as when recording an electric guitar and using both a direct-input signal and a microphone recording of an amplifier, or if using external effects and mixing the effected signal with the original. An emerging technique in sound production is to parallel process audio, for example duplicating a track and sending one track through a compressor and mixing the two audio tracks together. In this case comb filtering can occur if there is latency in the processing applied to the duplicate track.
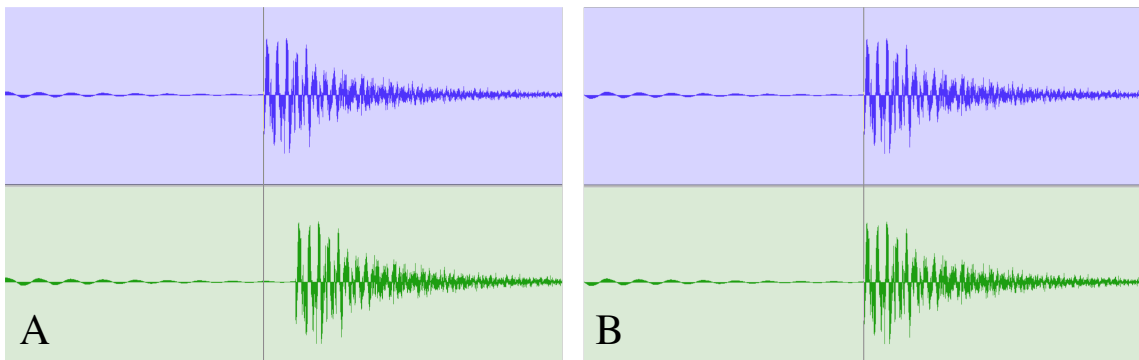

Figure 4 An example of nudging audio regions in an audio editor to compensate for delays that cause comb filtering showing delayed (A) and compensated (B) signals.

A difference of as little as 1 sample can cause comb filtering to the source signal. At a sample rate of 44.1kHz and taking the speed of sound at 343m/s, a difference in time of 1 sample is equal to a difference

in distance of just 8mm. Summing a signal and a delayed version of that signal by 1 sample causes a simple low pass filter and therefore high frequencies will be attenuated.

The comb filtering caused by a delay of a signal can be reduced by applying a compensating delay to give the illusion that the source is arriving a both microphones at equal time. This can be done either by measuring the distances of the microphones to the source and calculating the difference in delays that occur and applying an additional delay to the microphone signal that has initially the least delay. Compensating delay can also be applied "by ear" until the comb filtering is reduced. In a studio situation where the audio can be post processed, the audio regions can be nudged so the signals visually line up. This is shown in Figures 4(a) and 4(b). Figure 4(a) shows a snare drum recorded by two microphones. The bottom waveform contains the delayed signal. Figure 4(b) shows the same recording after the waveforms have been manually "nudged" in line. Both snare drum signals now occur at the same time. Many audio production softwares include some form of delay compensation to compensate for latencies that occur when using inserts.

These methods can be inaccurate and do not attempt to apply the precise delay that is occurring. These methods also apply a static delay. Therefore if the source or the microphones are moving, comb filtering may still occur.

2 Delay Estimation

Methods in signal processing exist for estimating the delay between two signals, known as time delay estimation (TDE), with no prior knowledge of either the microphone or source positions. An overview of current TDE methods can be found in (Chen et al., 2006). A concise and widely used method is the Generalized Cross Correlation (GCC) (Knapp and Carter, 1976). This method can also have weightings applied to improve the accuracy of the delay estimation against noise and reverberation, in this case the Phase Transform (PHAT). This sets all frequency amplitudes equal to 1, preserving the phase information, before performing the inverse FFT. The delay is calculated using the method outlined in Figure 6 and is equivalent to calculating the impulse response between the microphones and weighting with the PHAT (Meyer, 1992). Both methods produce the output as seen in Figure 7.
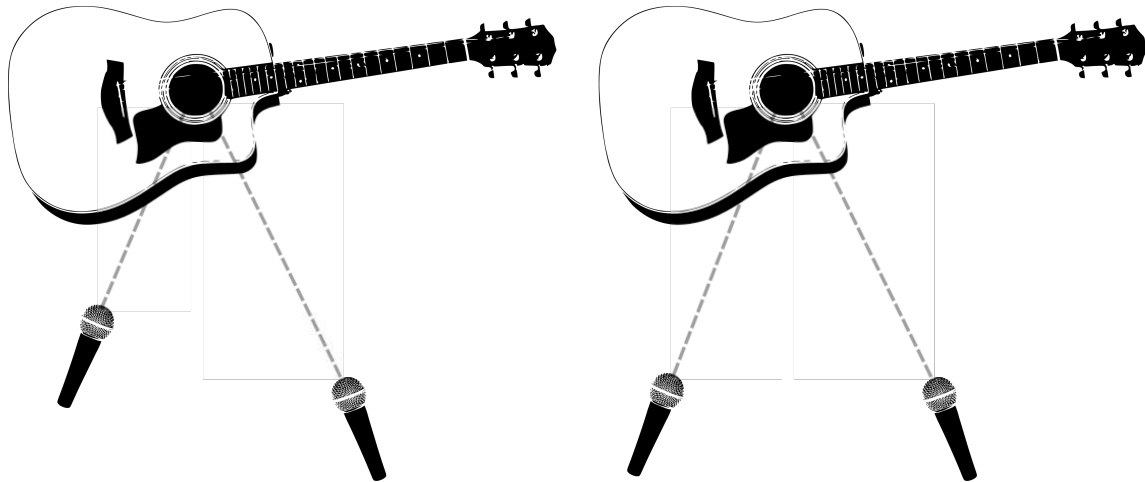
Figure 5 A representation of the effect of using delay compensation to reduce comb filtering.

When a delay is present, the position of the peak in the GCC-PHAT output is the estimate of the delay. This estimation of the difference in delays can subsequently be used to apply accurate delay compensation. In an ideal situation this will remove the comb filtering, and will boost the gain of the source as all frequencies are reinforced, causing a doubling in gain. A method for automatically calculating and compensating for comb filtering caused by delays using the methods mentioned is implemented in (Perez Gonzalez and Reiss, 2008) and it has been shown that multiple delays can be calculated from a single GCC-PHAT calculation in (Clifford and Reiss, 2010).
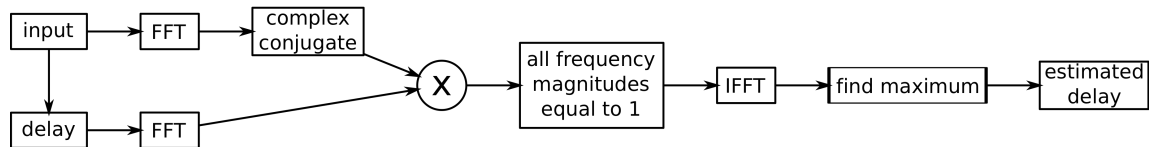
Figure 6 Block diagram of the GCC-PHAT method to estimate delay where FFT denotes Fast Fourier Transform and IFFT denotes Inverse Fast Fourier Transform

In a studio situation a single delay can be calculated by performing the calculation on the entire signal. A delay can then be set for the whole track. In some instances, the delay between sources will not be static but will be constantly changing. This can be due to the performer, the instrument or the microphone shifting during the performance. For this reason the audio recording can be split into blocks and the delay calculated for each block allowing the estimated delay to change over time. This information can then be used to automate a delay over time over the whole track. Performing the delay estimation using blocks also enables the delay estimation to be performed in real time and used in live sound. The problem with performing delay estimation in blocks rather than a whole audio track is that the amount of data available to perform the calculation decreases and therefore the accuracy decreases. The method can also be extended to incorporate more than 2 microphones. In this case, the microphone with the longest delay is identified and all other microphones delayed to be in line with this microphone.
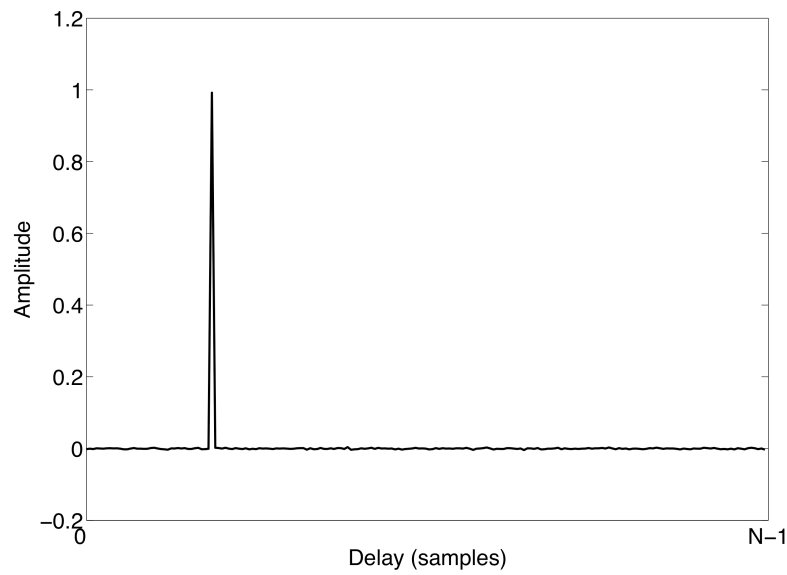
Figure 7 Output of the GCC-PHAT

3   Bandwidth

A test was performed to investigate how using real recordings affected the accuracy of the delay estimation. A single loudspeaker was used as an input and two microphones placed at different distances from the loudspeaker in the listening room at Queen Mary, University of London, an acoustically treated studio control room. The input signals were dry recordings made with either very close miked or recorded directly into the microphone pre amp and were of a vocal performance, bass guitar and snare drum. It was found that the accuracy of the time delay estimation varied depending on the input source.

Figure 8 shows the output of the estimation, with the block number shown against the calculated delay. The horizontal dashed line indicates the correct delay. It can been seen that the delay estimation of the bass guitar very rarely estimates the correct delay, whereas the delay estimation of the snare drum more often estimates the correct delay, the majority of the estimates alternating between 0 and the correct delay. This indicates that different types of instrument input signals will produce different accuracies of delay estimation. The results suggest that the frequency content of the input signals is important in the accuracy of the delay estimation. A bass guitar signal will have frequency energy concentrated in the low frequencies, where as a snare drum, which is similar to random noise, will have energy spread over the whole frequency spectrum. Approaching this in terms of frequency bins of an FFT calculation, a bass guitar has bins at low frequencies with high energy and other bins with low energy, whereas a snare drum will have the energy spread over the all bins.
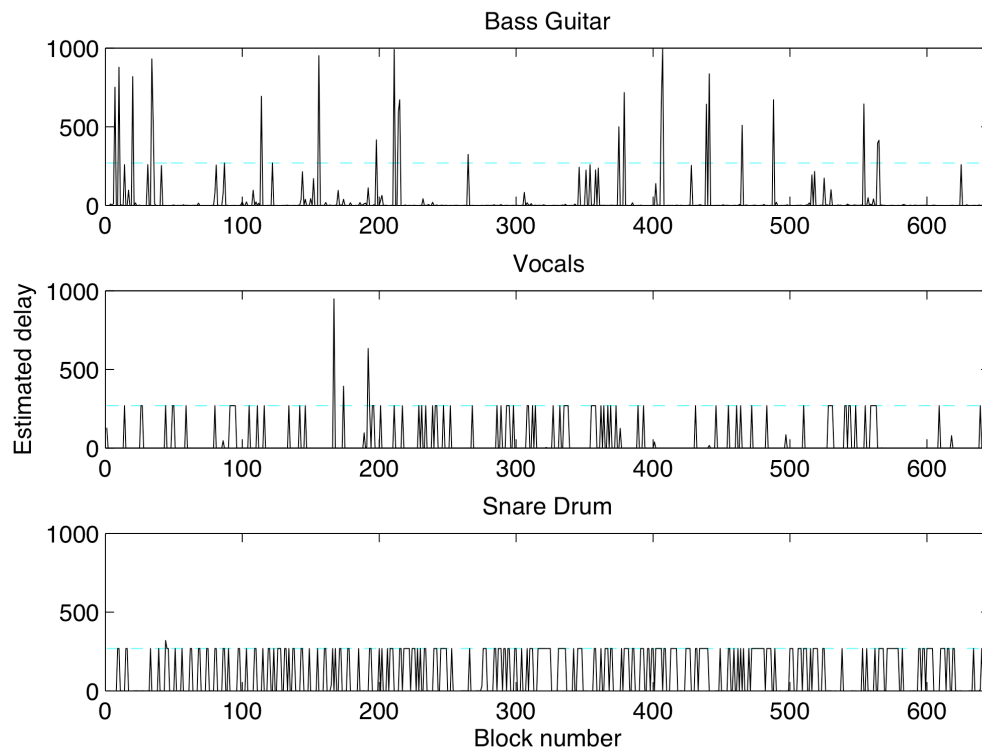
Figure 8 Results of experiment with real recordings showing the percentage correct delay estimates per block.

The problem with performing delay estimation with blocks of a signal is that the accuracy decreases for bandwidth limited source signals, i.e. source signals that have frequency content that extends over very little of the frequency range, in this case up to 22.05kHz. As the delay estimation is performed using the Fast Fourier Transform this can be thought of in terms of frequency bins. In theory, if the input source is random noise, all frequency bins will contain a value. As the audio is split into smaller blocks, the number of frequency bins per block also reduces.

To investigate this, a simulation was performed to test delay estimation accuracy of band pass filtered white noise. The band pass filters ranged linearly from 0Hz (low pass filter) to 22050 (high pass filter). For each centre frequency the bandwidth was increased linearly between 50 Hz and 11050Hz. The reason for this is that if more than half of the frequency bins contain significant energy then a correct delay estimate will be produced. The results of the simulation can be seen in Figure 9. It can be seen that as the bandwidth of the filter is increased, the accuracy increases. It can also be seen that this trend occurs for each centre frequency. This is shown by the dashed black line showing the average of all centre frequencies.
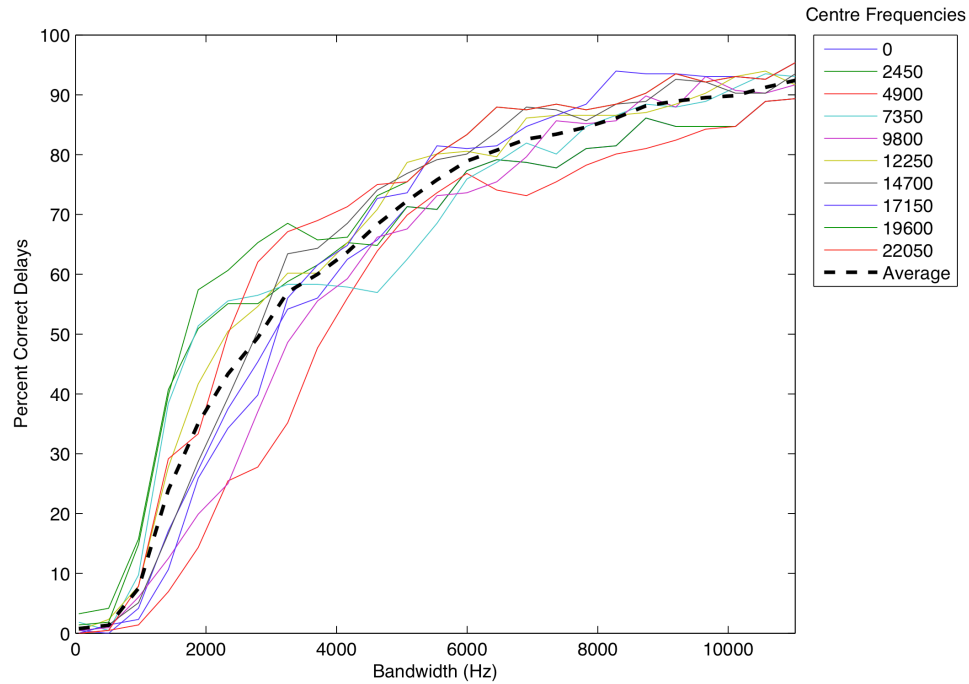
Figure 9 Results of a simulation to investigate how accuracy is proportional to bandwidth.

This change in accuracy occurs because of how frequency is related to phase. Figure 10 shows the output of the GCC-PHAT and corresponding phase response of low pass filtered white noise and unfiltered noise, duplicated and a 10 sample delay applied to the duplicate. Plot C shows the expected ideal output of the GCC-PHAT, with a definite peak indicating the delay. When a time delay is applied to a signal, the same delay in time is applied to each frequency in that signal, but each frequency is changed by a different phase.

An ideal time delay will apply a linear phase change to each frequency. This can be seen in Plot D as the phase response is linear. The slope of the linear phase response will be equal to the time delay. Plot A shows the output of the GCC-PHAT using low pass filtered noise at 1000Hz. The maximum peak is at 0 but a secondary, less defined peak can be seen at 10 samples, the applied delay. In delay estimation, the delay would be estimated as 0 as this is the maximum peak. Plot B shows the phase response for the filtered noise. It can be seen that initially at low frequencies there is a linear relationship between phase and frequency. The phase then distorts as the cutoff point of the filter is reached. Above 1kHz, the cutoff point of the filter, the phase response then becomes static. If the slope of the phase response is equal to the delay in the output function, then the slope of the phase response in B will mostly be 0, as it is a horizontal line, with some of the slope being equal to 10 samples. This is shown in the output by a peak at 0 and a smaller peak at the actual delay.
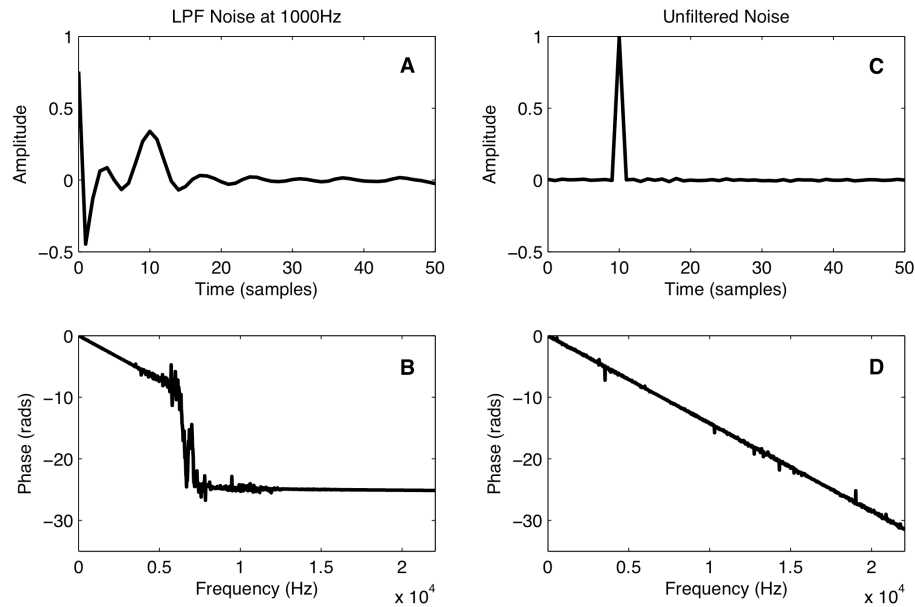
Figure 10 A) GCC-PHAT output of low pass filtered noise at 1kHz and B) the phase response. C) The GCC-PHAT output of unfiltered noise and D) the phase response.

The result of a smaller bandwidth is that some frequency bins will have little or no information and thus the phase response will also have little information. Ideally, the phase response of a delayed signal will have a linear phase, the slope of which is equal to the delay, as seen in Plot D of Figure 10. As mentioned previously, as the delay information is contained in the phase content of a signal, the frequency magnitudes are set to 1. If there is little or no information at certain frequencies, this will be exhibited as a horizontal line in the frequency response, as seen in Plot B. As the slope of this part of the phase response is 0, this is exhibited as a peak at the 0 delay position in the GCC-PHAT output, seen in Plot A, which may have a higher amplitude than the peak corresponding to the correct delay, thus making the delay estimation incorrect.

This can be resolved by using a non-rectangular window, such as the Hann window as seen in Figure 11, before performing the FFTs in the GCC-PHAT calculation. Non-rectangular windows are commonly used to improve resolution in the frequency domain as they taper to almost 0 at the start and end points, reducing frequency anomalies that occur due to truncation of signal into blocks (Mulgrew et al., 2003). Non-rectangular windows also affect the phase of the signals being windowed. Using a non-rectangular window removes the 0 peak that occurs with bandwidth limited signals, therefore the correct delay becomes the maximum peak.
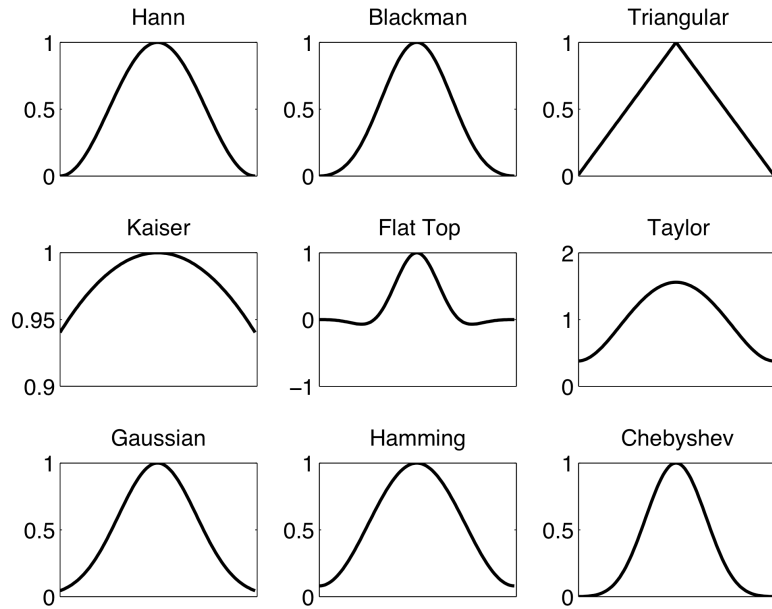
Figure 11 Examples of window shapes.

When the signal is split into blocks, as described previously, it can also be called windowing. The simplest of window is the rectangular window. This is simply taking the blocks of samples as they are, without any more processing, and performing the calculation. There are also other shaped windows that can be used, as seen in Figure 11. Different shaped windows are used to reduce distortion that occur due to edge distortions. The non-rectangular windows shown all taper at the edges, some down to zero.

Different windows can be applied when extracting blocks of a signal to perform the GCC-PHAT. The position of this process in the calculation is shown in Figure 12.
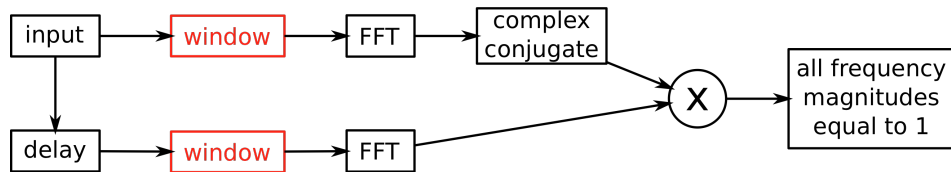

Figure 12 Block diagram of the estimation of delay using the GCC-PHAT using a non-rectangular window.

It was found that when a non-rectangular window was used the accuracy of the delay estimation improved. Preliminary experiments suggested that of the common non-rectangular windows, the Hann window produced the most improvement in accuracy therefore the real recordings were analysed again using a Hann window prior to the GCC-PHAT calculation. The results can be seen in Figure 13 which shows the percentage of frames for each input source which estimated the correct delay and the results

from using both a Rectangular and Hann window to window the signal for each block. The time delay estimation performed with the Rectangular window show the bass guitar had the worst accuracy and the snare drum the highest accuracy. The accuracy of all input sources is shown to improve when using a Hann window. It can be seen that the highest proportion of the delay outputs for the bass guitar is the correct delay value therefore if the results are accumulated and averaged an accuracy close to 100% will be reached.
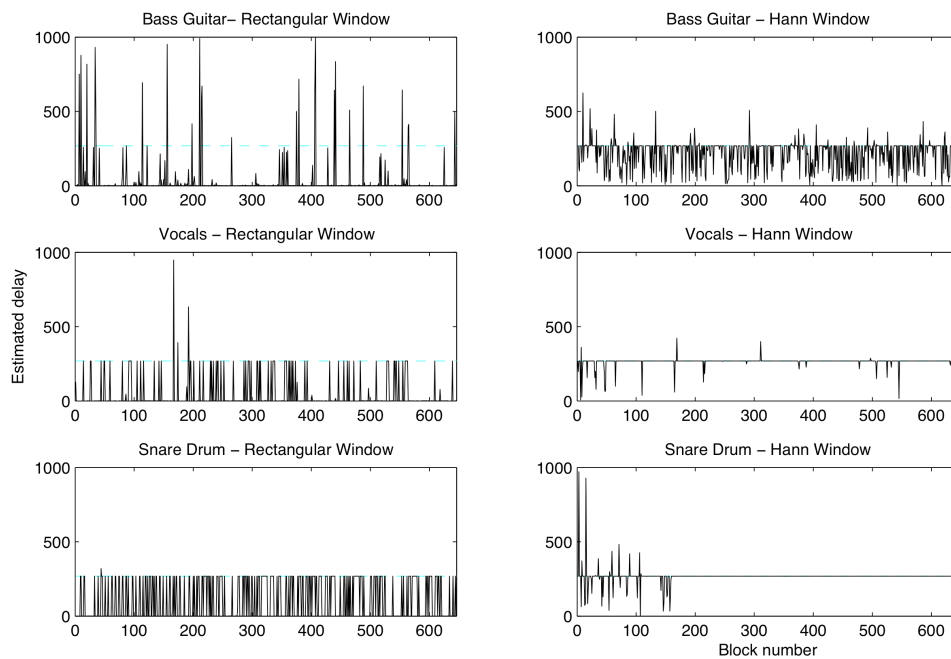

Figure 13 Results of delay estimation of various input sources using both a rectangular and Hann window.

The accuracy of each instrument as a percentage of the blocks that estimated the correct delay can be seen in Figure 14. This also shows an overall improvement to each input signal. An improvement in delay estimation accuracy results in more accurate delay compensation and therefore more effective comb filter reduction.
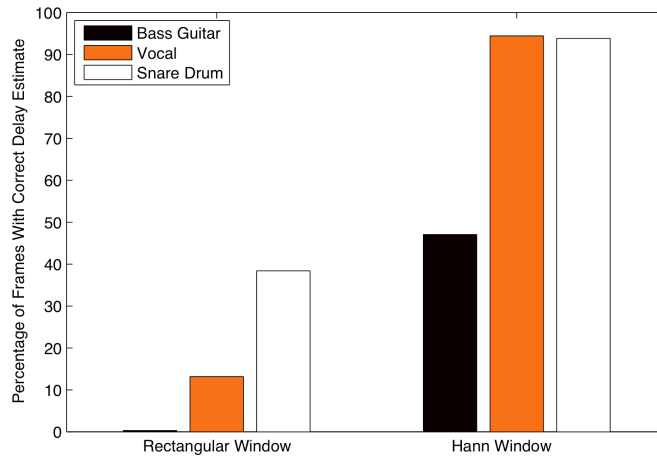
Figure 14 Results of delay estimation of various input sources showing the percentage correct delay estimates per frame with both rectangular and Hann (non-rectangular) window.

4   Conclusion

Comb filtering has been shown to be a detrimental effect to the perceived quality of the audio and is an undesired effect to the signal. Manual methods have been outlined for reducing this effect and it has also been shown that signal processing of a signal using time delay estimation can be used to reduce the effect of comb filtering automatically.

Delay estimation has been used to estimate the actual delay between signals and compensate for this delay accordingly to reduce comb filtering. It has shown to be applicable to live and studio productions, primarily for the compensation caused by delays that occur when multiple microphones are not placed equidistance from a source. Examples have also been given, such as when using external effects, where delay estimation and compensation can be used in studio production.

This paper has shown that the accuracy of the time delay estimation is dependent on the input signal and the length of the audio under observation, but the accuracy can be improved by using non-rectangular windows, such as the Hann window.

Extensions of this work include looking at calculating sub sample delay and how delay estimation can be used in the reduction of comb filtering in multiple source configurations.

Bibliography

Brunner, S., Maempel, H.-J., and Weinzierl, S. (2007). *On the audibility of comb-filter distortions*. In Proceedings of the 122nd Audio Engineering Society Convention, Vienna, Austria.

Chen, J., Benesty, J., and Huang, Y. A. (2006). *Time delay estimation in room acoustic environments: An overview*. EURASIP Journal on Applied Signal Processing, 2006:1 – 19.

Clifford, A. and Reiss, J. (2010). *Calculating time delays of multiple active sources in live sound*. In 129th Convention of the Audio Engineering Society.

Eargle, J. (2004). *The Microphone Book*. Focal Press, Oxford, UK.

Knapp, C. H. and Carter, G. C. (1976). *Generalized correlation method for estimation of time delay*. IEEE Transactions on Acoustics, Speech and Signal Processing, 24(4):320–327.

Meyer, J. (1992). *Precision transfer function measurements using program material as the excitation signal*. In Proceedings of the 11th International Conference of the Audio Engineering Society: Test and Measurement, Portland, Oregon.

Mulgrew, B., Grant, P., and Thompson, J. (2003). *Digital Signal Processing: Concepts and Applications*. Palgrave, 2nd edition.

Perez Gonzalez, E. and Reiss, J. (2008). *Determination and correction of individual channel time offsets for signals involved in an audio mixture*. In Proceedings of the 125th Audio Engineering Society Convention, San Francisco, USA.

Zölzer, U., editor (2002). *DAFX - Digital Audio Effects*. Wiley, UK.