# THE HISSTOOLS IMPULSE RESPONSE TOOLBOX: CONVOLUTION FOR THE MASSES

*Alexander Harker, Pierre Alexandre Tremblay*

CeReNeM
University of Huddersfield
Huddersfield, HD1 3DH, United Kingdom
ajharker@gmail.com / p.a.tremblay@hud.ac.uk

**ABSTRACT**

This paper introduces the *HISSTools* project, and its first release, the *HISSTools Impulse Response Toolbox* (*HIRT*); a set of tools for solving problems relating to convolution and impulse responses (IRs). Primarily, the aims and design criteria for the *HISSTools* project are discussed. The elements of the *HIRT* are then outlined, along with motivating factors for its development, underlying technologies, design considerations and potential applications.

## 1. THE HISSTOOLS PROJECT

### 1.1. The HISS

The term *HISS* (Huddersfield Immersive Sound System) refers simultaneously to the multi-channel loudspeaker system, the academic community using it and the research lab attached to it. The common thread is a focus on the composition and performance of electronic music. More details can be found on the *HISS* website [1].

### 1.2. HISSTools

#### *1.2.1. Aims*

The *HISSTools* project aims to create powerful modular tools to address specific issues related to the composition, performance and presentation of electronic music. In particular, this project aims to bring complex technologies to mid-level users, who would not otherwise have access them. Pre-existing tools may not be accessible or suitable to many users. Typically, this applies to tools that exist only in an uncompiled code format (e.g. C / C++), or have an interface that is minimal or cumbersome (e.g. a command line interface). Where existing tools are inappropriate or non-existent, but literature on the technique *is* available, issues arise because; either the potential user has no time to deal with low-level technical implementations alongside creative work; or the user lacks the expertise to implement the technology from scratch.

The *HISSTools* project addresses such users by providing flexible tools for use in appropriate environments, such as objects for audio programming languages (such as *MaxMSP*, *pd* and *SuperCollider*, and plug-ins for use in the context of DAW software.

#### *1.2.2. Availability*

Although the focus of the *HISSTools* reflect the research activities of *HISS* members, they engage with problems that affect the wider community. The goal is to further understanding of DSP techniques, and widen the creative possibilities available to practitioners of electronic music, Thus, these tools are licensed freely for use, and are fully open-source.

### 1.3. *HISSTools* Design Criteria

#### *1.3.1. Lightweight / Comprehensive Modules*

The *HISSTools* are intended to be used in a variety of contexts, as part of musical workflows that cannot necessarily be predicted in advance. Thus, minimal, lightweight implementations, comprehensively dealing with a well-defined and coherent task are called for. Solutions should be sufficiently *powerful* but also *flexible* so that the user can concentrate on their specific musical needs, rather than on underlying technical details. Modularity is key to supporting a variety of musical and technical needs with minimal development time. Modules should place minimal requirements on the user in terms of setup and learning curve.

#### *1.3.2. Environment Appropriate Implementation*

It is desirable to maintain consistency with the behaviour of the host environments, in order to facilitate easy integration into a pre-existing workflow. This concerns core engineering decisions such as memory-access/state management models as well as more superficial issues; such as object and method/message naming conventions.

#### *1.3.3. Efficient*

For real-time applications, peak CPU usage (rather than average or amortised values) is the limiting factor; even a momentary overload can result in audio dropout. Thus, all possible steps should be taken to minimise the use of resources. These include careful algorithm choice/design, minimising unnecessary recalculations, use of SIMD instructions, algorithmic approximation (where appropriate) and avoidance of branching within loops.



_ICMC2012 NON-COCHLEAR SOUND LJUBLJANA _9.–14. SEPTEMBER

Memory efficiency is also a concern as even 64-bit spaces do not prevent excessive paging in and out due to a lack of physical RAM. The consequent data starvation can be disastrous for real-time performance.

#### *1.3.4. Stable and Reliable*

Stability is such an obvious concern that one should not need to raise it, yet unfortunately it is common to encounter tools for electronic music that are suffer significant deficiencies in this area. Lack of stability can be catastrophic in performance, amounting to absolute and unrecoverable failure. For the *HISSTools* we distribute pre-release tools to a set of international 'power user' testers and carrying out rigorous internal unit testing, in order to ensure both application stability, as well as reliable, accurate results.

#### *1.3.5. Pragmatic*

A pragmatic approach has previously been advocated by the authors in [26]. We favour only solutions that are suitability straightforward for the end use, and for which the benefits significantly outweigh any drawbacks. This may mean rejecting theoretically optimal solutions in cases where they prove unsuitable for quick and practical deployment in a musical context.

## 2. THE HISSTOOLS IR TOOLBOX

### 2.1. Overview

*HISSTools* first public release is a set of tools for working with convolution and IRs in *MaxMSP*. This set of objects addresses tasks such as measuring IRs, spectral display from realtime input/buffers, and non-realtime convolution, deconvolution and inversion. Although the primary platform for development has been *MaxMSP* (chosen for its overwhelming popularity amongst practitioners), much of the code-base is generic, so as to support porting to other environments.

These tools enable the design of solutions to a number of specific issues of concern to *HISS* members. The toolbox can be downloaded from the *HISS* website [12].

### 2.2. Toolbox Concept

The primary motivation for the *HIRT* was to provide tools for improving the concert presentation of electronic music (especially when combined with acoustic instruments). Any combination of room, loudspeaker, speaker position, and listener position will affect the sound heard by the listener. Acknowledging this was the focus of earlier research [28], in which the approach was to simulate the effect of the concert hall within the studio, so as to produce music that would not suffer unexpected and potentially highly detrimental alterations in concert. The reverse of this approach is to treat the system in question so as to achieve results closer to those heard in the studio. This is the first goal of the *HIRT* project: to allow for loudspeaker/room corrections to be generated and applied rapidly and efficiently for use in a concert environment. By extension, such corrections can also be employed in a studio setting.

The second motivation for the *HIRT* was to improve the frequency balance of close microphone or pick-up capture. This work was inspired by work by Alban Bassuet [3], who advocates a system based on the relative frequency profiles of two microphones at different distances from the instrument. The idea is to take a close capture of the instrument that is relatively free from issues of spill and feedback, and improve the frequency balance prior to electronic processing, thus achieving better tonal balance.

Whilst these motivations are well-defined, it was felt that specific solutions would be both inflexible for the given tasks, and also limited in terms of re-application. Thus, a highly modular design was preferred, suited to a large number of convolution and impulse response related problems. The *HIRT* is thus suitable for use in relation to several general problems, including multichannel IR measurement, multichannel convolution, FIR Design and crosstalk cancellation.

### 2.3. Underlying Technical Approaches

#### *2.3.1. FFT-based Processing*

The *HIRT* takes an FFT-based approach to convolution and deconvolution problems, using a 64-bit precision FFT for all frequency domain processing. This is an efficient means of calculating convolutions, but it also has important implications for deconvolution. Firstly, one must be aware that the circularity of the FFT means there is potential for time-aliasing, in which energy wraps around the end points of the FFT window. Some user awareness of this issue is thus required when using the *HIRT*.

This problem might be avoided by using a least-squares time domain deconvolution, which produces optimal results in the time domain for a given filter length. However, calculation times are impractical for even moderate filter lengths, whereas it is quite practical on modern computers to calculate FFT-based deconvolutions on inputs of several tens of seconds within a relatively short time (a few seconds at most). Furthermore, whilst a time domain approach results in filter that is optimal in the time domain sense, rather than in a frequency domain. On modern machines minimising filter length is not of great concern. Typically, we over-specify the filter length and control the optimality in the frequency domain, subsequently truncating and fading the resultant IR to remove near-zero coefficients. This gives direct control over the frequency specification of the filter, which is of highest concern for our applications.

#### *2.3.2. Fast Deconvolution with Regularisation*

Using an FFT-based approach, deconvolution between time-domain inputs is possible simply by taking the FFT of each input, performing complex division, and transforming the result back into the time domain.[1] It is often also appropriate to introduce a delay to the output in order to deal with non-minimum phase components, which result in a non-causal output.

However, deep nulls in the divisor can result in extremely long output filter lengths, often with significant time-aliasing if the FFT is not large enough to hold the result. Additionally, filter blow-up may result, in which the output is unusable due to very large coefficients In fact, there is no guarantee that the output will be finite or causal, even given that the input is an impulse response that is both finite and also a measurement of a causal system. In order to ensure useable results, it is necessary to circumvent such problems by 'regularising' the deconvolution [17]. Using frequency dependent regularisation it is possible to reduce the output of the deconvolution in frequency extremes, or other deficient frequency regions. Thus it is possible to control the length of the resultant filter, as well as its accuracy in different frequency ranges, according to the area of interest.

Regularisation as proposed in [17] is equivalent to convolution between the direct result of the complex division and a linear phase IR (which adjusts the amplitude response, but not phase). As linear phase FIRs are symmetrical, they exhibit equal amounts of pre- and post-ring. Thus, the implicit convolution can result in undesirable pre-ring in the output. A rearrangement of the process is proposed in [4] which makes this implicit convolution explicit. The linear phase convolution can then be replaced by a minimum phase convolution of the same amplitude response to reduce pre-ringing. The *HIRT* allows both linear and minimum phase forms of regularisation.

### 2.3.3. IR Measurement

Currently, the preferred method of impulse response measurement for most applications is the exponentially swept sine (ESS) [6]. The ESS takes a constant amount of time per octave, with longer sweeps resulting in improvements to the signal-to-noise ratio. This method offers a very high SNR in comparison to previous technologies such as Maximum Length Sequences (MLS - [23]) or direct measurement. It is also has the benefit of separating out the linear component of a system from the non-linear ones, resulting in an IR for each harmonic of the system which together characterise the distortion characteristics of the measured system [5]. The relative merits of various technologies are discussed in [25].

The *HIRT* implements both ESS and MLS techniques for the purposes of completeness, despite the widespread adoption of the ESS technique. Historically deconvolution of recorded excitation signals has been performed through convolution with a suitable analytical inverse signal. Whilst previously this was important in terms of efficiency and memory resources, it is now viable to perform the deconvolution in the frequency domain using a large FFT without issue. Using this method, any known excitation signal can also be used to measure or estimate the IR of a system, although an arbitrary signal is unlikely to be optimal. We have achieved usable results using both coloured noise signals, and even a music programme (given a sufficiently long measurement period). Whilst the ESS signal offers very good measurement results, it is unpleasant to listen to, especially at high volumes, and thus it is not viable for use in occupied venues.

[1] Generally the FFT is quite large as it is required to be at least large enough to contain the longer of the two inputs, plus padding to deal with with phase shifts in the output.

### 2.3.4. Frequency Smoothing

Several applications require that the overall frequency profile of an IR be corrected (for instance, room equalisation, which should not correct every detail of a room response, as this is equivalent to dereverberation). In such instances, it is desirable to work on a smoothed approximation of the frequency response, rather than the response in raw form [13, 14]. The key formulation is convolution of the IR in the frequency domain, with a sliding von hann window, which increases in width as the central frequency increases. Thus, there is more smoothing (in a linear sense) at higher frequencies than at lower ones. This relates to perception in that the ear is more sensitive to small frequency differences in lower frequencies areas. In the *HIRT* the size of window at each frequency bin is controlled by the following formula: $ws = smooth_{lo} + \theta * (smooth_{lo} - smooth_{hi})$, where $ws$ is the window size, $\theta$ is the central frequency of the bin, and $smooth_{lo}$ and $smooth_{hi}$ are the smoothing amounts at 0Hz and the Nyquist frequency respectively. The unit in all cases is normalised frequency.

For this purpose we favour the dismissal of phase entirely, and smooth the power spectrum only. This is in contrast to the complex smoothing proposed in [14], which arguably conflates phase and amplitude information. Using large filter lengths we found this approach did not produce suitable results. This is also appropriate to our concerns which are biased towards correcting the amplitude spectrum over the phase spectrum.

### 2.3.5. Phase Alteration

For some applications it is desirable to create an IR with a known power spectrum, but with a controllable group delay, somewhere between the minimum and maximum possible values. The minimum and maximum group delay for a given power spectrum correspond to the minimum of maximum phase filters with the specified power spectrum. Exactly between these lies is a linear phase filter that has a constant group delay for all frequencies (and a symmetrical form). The *HIRT* employs homomorphic prediction [22] to produce minimum phase spectra from non-minimum phase input IRs. Using phase interpolation it is then possible to interpolate between minimum and maximum phase forms of a given filter [16].





## 2.4. Object Overview

### 2.4.1. IR Measurement

***irmeasure∼*** performs IR measurement of a Multiple Input Multiple Output (MIMO) system using a range of known signals. The ESS signal is the technically optimal method, but may not always be practical for use (see 2.3.3). Alongside the swept sine method, MLS and coloured noise signals are offered as more pleasant sounding signals, with close to optimal performance. Measurement parameters are fully customisable. The object internally handles the IR retrieval from the recorded measurement (either by deconvolution, or optionally by convolution with an appropriate inverse signal for swept sine measurements) and IRs are trimmed correctly from the internal buffer without the need for specialist user knowledge. With swept sine measurements non-linear components can be individually retrieved [5].

***irreference∼*** calculates the deconvolution between two recorded real-time inputs. The resultant IR represents an estimate of the convolution necessary to transform the reference signal into the measurement input. The assumption is that the two input signals have the same source, but are picked up through different system. Possible applications are estimating IRs with any broadband signal (for instance, music in an occupied room), and deriving relative IRs for systems such as microphones.

In situations where only the overall frequency profile needs to be modelled, *irreference∼* can smooth the spectra of the two signals before deconvolution. In this case phase information is dismissed, and two minimum phase spectra with the smoothed amplitude profiles of the recorded inputs are deconvolved to produce the output. This may be preferable to post-deconvolution smoothing (with the *iraverage∼* object - see below), especially in the case where the input signal is not particularly broadband.

***irsweeps∼*** generates exponential sweeps to a *buffer∼* (and inverse sweeps for IR retrieval by convolution if desired). Sweeps can be saved to disk and IRs measured using any system capable of simultaneous PCM playback and recording in cases where it is not practical to run *MaxMSP*. The *bufconvolve∼* object can later be used to do the appropriate deconvolution using a copy of the original sweep or its inverse.

### 2.4.2. Offline Convolution and Deconvolution

***bufconvolve∼*** performs both convolution and deconvolution of buffered audio. In the later case, there are options for different kinds of regularisation.

***irinvert∼*** performs IR inversion. In the single-channel case, this is essentially deconvolution of a Dirac delta function (a single sample spike in the digital domain) by an input IR. This object offers the same regularisation options as the *bufconvolve∼*. The object also implements MIMO deconvolution (as proposed in [17, 18]).

### 2.4.3. IR Processing

***iralign∼*** performs a simplistic time alignment of any number of mono IRs (stored in buffers). *iralign∼* does this by finding the maximum sample value in each IR, and front-padding all but one IR with zeros so that these samples are aligned to the same time index (corresponding to the latest position of the maximum amongst the inputs).

***iraverage∼*** can calculate one of two types of average given a set of input IRs. The time domain method averages corresponding samples of the inputs (which are ideally pre-aligned). This approach retains phase information, but can suffer from meaningless constructive and deconstructive phase interference between IRs. The frequency domain method discards phase information and averages the power spectra of the IRs. Optionally, the resultant power spectrum may then be smoothed, and a fixed phase result created.

***irphase∼*** takes an input IR, and creates an IR with the same amplitude spectrum, but a different deterministic phase spectrum. The phase of the IR can then be controlled from minimum phase, through to maximum phase (with linear phase in the centre). Useful for creating minimum latency filters (i.e. minimum phase), or creating custom filters with a specific phase specification.

***irtrimnorm∼*** a utility for performing trimming and normalisation of recorded IRs. It can be used either to crop IRs according to set in and out points, or to trim the IRs according to RMS level. This allows the removal of excess from the start and end of a set of IRs in a uniform manner. Normalisation is applied such that the relative level between inputs is maintained. The ability to operate over a set of related IRs is crucial for multichannel work.

***irnonlin∼*** the non-linear components retrieved from an ESS measurement relate to the individual harmonics of a system. If these are transformed appropriately, it becomes possible to model the non-linear system by a Hammerstein model which operates on consecutive powers of the input signal [2, 21], rather than harmonic transpositions. N convolutions are necessary to model the first N harmonics. The *irnonlin∼* object, takes N input IRs for the harmonics (taken with the *irmeasure∼* object) and returns N new IRs for use in a Hammerstein model.

### 2.4.4. Realtime Convolution

***multiconvolve∼*** is a single convenient object for zero-latency convolution using a fixed partitioning scheme. This object combines time domain convolution for the early portion of an IR with more efficient FFT-based partitioned convolution for the latter parts of the IR.[2] This object also supports multi-channel behaviour directly. There are two modes, the first of which is a parallel multi-mono

[2] This object is based on code from the pre-existing *timeconvolve∼* and *partconvolve∼* objects from the *AHarker Externals* package [11] The overall technique is similar to that in [8], and combines the benefits of zero-latency given by the time domain approach, with the efficiency of FFT-based convolution. The use of SIMD instructions throughout results in an efficient real-time implementation.

mode in which each input is convolved once only to drive the respective output. In true multichannel mode there are up to $N^2$ convolutions (one for every combination of individual inputs and outputs). This is appropriate for scenarios such as true multichannel reverbs, or crosstalk cancelation networks. For efficiency it possible to specify IRs for only certain routings within the convolution network.

#### *2.4.5. Visualisation*

***spectrumdraw*∼** is a comprehensive spectral display for either real-time audio, or buffered audio. A single display can be used to display up to four real-time audio streams simultaneously, or the spectral contents of buffers, (e.g. the spectral shape of an IR), or a combination of the two. For real-time uses FFT size and windowing parameters dynamically alterable, and there are several display modes, such as peak hold and time averaging.

#### *2.4.6. Misc*

***iruser*∼** generates fixed phase IRs (minimum through maximum phase as desired) from a user-defined amplitude response. This is specified as a series of $x$, $y$ points in Hz/dB. The final amplitude response is then created by connecting these points linearly in the log-log domain, and optionally smoothing the result,.

***irplapprox*∼** approximates the amplitude response of an IR with a set of linear segments using a recursive algorithm. Although the segments are nominally linear, they are derived from a log-log representation of the input. (i.e. linear in relation to an octave/decibel plot). This linearity in the log domain is appropriate to the nature of an audio signal, as with the similar approach taken in the *iruser*∼ object. This object is intended to facilitate both data reduction, and automated regularisation schemes, where the regularisation function is derived from a simplified representation the input IR.

***irstats*∼** estimates statistics of an IR, such as onset, reverb time, clarity and mixing time (the transition between early and late reflections of a room impulse response). See [7] for technical details of relevant acoustic measures.

***irtransaural*∼** - calculates impulse responses for stereo transaural crosstalk cancellation for a given system for the playback of binaural material over loudspeakers.

## 3. APPLICATIONS AND EVALUATION

### 3.1. Loudspeaker and Room Correction

#### *3.1.1. General Problem*

As discussed previously, any loudspeaker and room in combination will have an effect on the sound heard by a listener. In theory the system can be measured, and a correction impulse response derived to exactly compensate for the effects of the system (albeit for a single combination of speaker and listener positions) [20]. This correction IR can be applied as a convolution before output to the loudspeaker; as convolution is commutative, the effect of the system can be 'pre-corrected'.

Naive implementations of this technique are in practice rarely satisfactory for the exact listening position measured, or even nearby listening positions (where the impulse response of the system can be noticeably different), never mind more distant listening positions. The results can suffer from extreme ringing caused by deep nulls in the measured impulse response, long latency induced by a need to delay the impulse to allow correction of a mixed phase response.[3] and pre-ring from inaccuracies in the measurement [15].

The correction generated by direct convolution is also designed to correct the measured system to a Dirac delta IR. For a given combination of room, loudspeaker and listener position, such a correction attempts to dereverberate the signal exactly, as if the sound were playing in a perfect anechoic chamber. In practice this attempt is both unsuccessful, and undesirable; a better explanation of our aim is that of more uniform sound reproduction between venues, but not to entirely remove the effect of the room, which would produce extremely artificial-sounding results even if it were technically viable.

At this point in the discussion the situation might appear hopeless. However, by relaxing the requirements for exact correction the results in fact improve dramatically. The general problem is thus to generate corrections that work over a wide area (especially for concert hall presentation), do not suffer from overly distracting artifacts (the improvement should outweigh any negative effects - if indeed there are any), and are viable for multichannel systems (multichannel sound reproduction being of particular concern).

#### *3.1.2. Proposed Approach*

Our approach is a multi-mono one, in which each output channel is measured and corrected separately. It is most practical to do so in parallel so as to maintain the correct relative volume between channels, and ensure that the same set of parameters is applied to each channel during the correction procedure. The output of each loudspeaker is measured from a number of relevant positions in the venue, according to the application (for concert presentation we measure a wider area than for studio applications, but we use a minimum of two measurements per output channel in all cases) using the *irmeasure*∼ object. In larger venues we have found it necessary at this point to crop the IRs (using *irtrimnorm*∼) to avoid correction of the late reflections rather than the more direct sound). For each channel, the results are averaged and smoothed in the frequency domain before single channel inversion (using *iraverage*∼ and *irinvert*∼ respectively). Regularisation is required so as to avoid overcorrection at either end of the spectrum. Typically the regularisation values are first set

[3] Although loudspeakers are often close to minimum phase, rooms are more commonly mixed phase, so the inversion of the measured impulse response is likely not to be causal (i.e. it will depends upon future input) To make the filter causal it is necessary to delay the inverted response.





to suitable defaults, and then hand-tweaked according to the viable frequency range of the system in question. Finally, the resultant filters are converted to minimum phase to reduce latency. Optionally the filter may be truncated, faded and normalised at this stage (using *irtrimnorm*∼). **Figure 1** shows the proposed approach in graphical form.

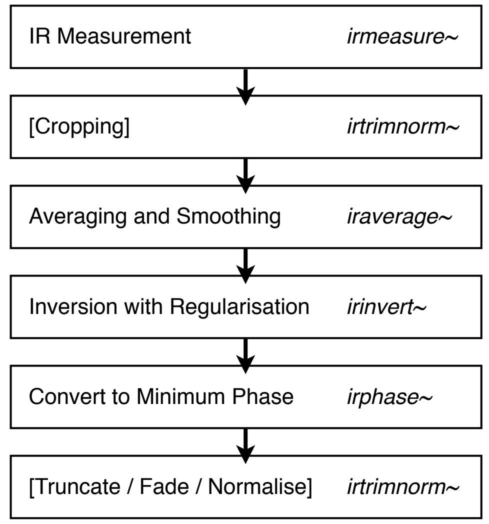


**Figure 1**. Loudspeaker/Room Correction

The measurement microphone is ideally totally flat, as the full system being measured (and corrected) includes the microphone. This is likely to become noticeable in the case that the microphone is more coloured than the loudspeaker/room combination. For our tests we have used a *DPA 4006* as our reference microphone, which exhibits an exceptionally flat frequency response.

#### *3.1.3. Case Study*

In concert, the system has been applied to a work for bass clarinet and electronics [27] in a large reverberant venue (St. Paul's Hall in Huddersfield) with which the authors are very familiar. The loudspeakers were three *Meyer UPJs* in LCR configuration. In this scenario, the best results were achieved by truncating the measured IRs before inversion to reduce correction of later reflections. In a work where the high frequency content is particularly detailed and direct, the results of the correction were notably clear and penetrating for the venue in question. As the score calls for the clarinet not to be amplified, the only issue was that the acoustic clarinet sounded a little distant in comparison to the electronics.[4] It is possible that increased regularisation would have reduced this problem, at the expense of a less noticeable correction. However, this issue would not arise during for the presentation of purely electronic music under the same conditions.

#### *3.1.4. Discussion and Evaluation*

Room equalisation is a controversial issue. There is no consensus on the best approach, or even whether its application is beneficial. Despite its detractors, it is our experience that the proposed approach produces noticeable improvements in all tested cases[5], albeit with some need to fine-tune results. We have tested the approach in a number of concert venues, in large general purpose rooms and at a number of institutional and home studios.

The smoothing parameters act as a means by which to control the tightness of the correction to details of the system response. This is mostly a matter of preference, and can be done by ear. The perceptual impact of using specific values is consistent across different rooms and loudspeaker combinations, and thus it is easy to choose suitable defaults. Regularisation parameters are somewhat harder to optimise, as appropriate values are highly dependent on the amplitude spectrum of the IR in question. In particular, overcorrection at high frequencies can be fatiguing for the listener; and care is required, as the rolloff of the different systems measured varied substantially, so a single default set of values cannot be used effectively.

Some discussion of phase issues is relevant, especially as a minimum phase approach has been taken. This effectively rules out problems of pre-ring, but could arguably create phase issues between channels if the corrections for each loudspeaker were substantially different. This has not been an issue to date. We verified this in a studio setting by switching both channels in a stereo set up to use the same mono correction IR rather than two separate correction IRs. Auditioning the two versions would reveal any phase issues in the stereo correction if present.

The problem of compensating for phase issues 'correctly' is a also complex one, and arguably one that has no clear definition across a wide listening area. In this case the phase spectrum is likely to vary wildly.[6] One approach is to use a linear phase correction (thus inducing a latency of half the filter length), but at best that only leaves phase relationships between channels the same as with no correction. In the case that a suitable mixed-phase average impulse for correction could be generated, artificial tests suggest that the resultant filter is likely to exhibit an unacceptable and distracting pre-ring. Thus, whilst not theoretically optimal, our approach remains to use minimum phase correction of the amplitude spectrum only.

### 3.2. Close Acoustic Capture Correction

#### *3.2.1. General Problem*

For the presentation of pieces involving processing of a live instrument, it is often desirable for the capture for processing to be taken from close to the instrument, often with a directional microphone. This reduces the pick-

[4] Note that the effect described is not an issue of *balance*, but rather of sonic *clarity*.

[5] Except a singular case at IRCAM, Paris, where the system was already very flat in amplitude response. Even in this case the correction procedure did not decrease fidelity in any way.

[6] Which phase spectrum should be corrected, and how does one make a meaningful average of phase spectra?

up of microphone spillage from other instruments, loudspeakers, extraneous noises, and room reverb, as well as reducing the potential for feedback. Unfortunately, the sound from an instrument is rarely well-balanced at short distances, and proximity effects from the microphone tend to exacerbate this problem. The proposal then (as in [3]), is to correct the close pick-up to emulate the results from a tonally optimal microphone type and position. This means a microphone with a flat frequency response placed such that the radiation patterns of the instrument have become balanced in a natural sounding manner. The ideal result retains the benefits of close pick-up, but improves the tonal balance noticeably, better reflecting the expected sound of the instrument from a normal listener position.[7]

### *3.2.2. Proposed Approach*

The proposed approach for close acoustic capture correction is very similar to that for loudspeaker and room correction. However, only a single measurement is necessary. This is taken by recording the instrumentalist through two microphones or pick-ups to generative a relative IR (using the *irreference*∼ object). We follow [3] in asking the instrumentalist to play chromatic scales in loud and quiet dynamics across the range of the instrument. Before inversion smoothing is applied, either by setting the *irreference*∼ object to smooth the two recorded signals prior to deconvolution, or by using the *iraverage* ∼ object upon retrieval of the IR (post the internal deconvolution of the *irreference*∼ object). Early results suggest that the differences are subtle, but pre-smoothing is slightly preferable, and the workflow for this approach is also simpler. After inversion with the *irinvert*∼ object, the resultant filter is converted to minimum phase. Again, optionally the filter may be truncated, faded and normalised. **Figure 2** outlines the full process for the correction of close acoustic capture.

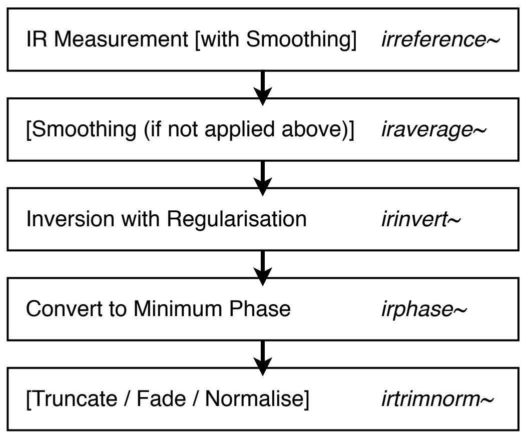


**Figure 2**. Close Capture Correction

As with loudspeaker/room measurements we use the flattest available microphone as a reference, as the full system measured includes the microphone.

[7] It is extremely rare for an audience to listen to an instrument from a distance of less than 30cm!

### *3.2.3. Discussion and Evaluation*

To date we have tested this approach on a number of files taken from microphone recordings of clarinet and bass clarinet and also on recordings of an acoustic bass guitar, captured by both a room mic (a *DPA 4006* for the reference signal) and an internal piezo pick-up to be corrected. In the case of the clarinets, the close capture was made by a *DPA 4060* clipped onto the music stand.[8] Two reference signals were tested, from a *DPA 4006* and *DPA 4011*, both placed 1 metre away from the source.

It is apparent that the amount of smoothing applied is crucial in this application, as lower levels of smoothing result in the reverberant characteristics of the room being simulated, and a tendency towards excessive filter ringing. It is necessary to spend time adjusting the smoothing parameters for the most preferable results. For acoustic bass guitar lower levels of smoothing also resulted in more simulation of the body of the instrument, which was captured by the room reference mic, but not the piezo pick-up.

The regularisation parameters were also more sensitive to fine-tuning than for loudspeaker/room correction, especially in the case of the bass guitar, where poor choice of parameters at low frequencies resulted in either a lack of apparent low-end, or an overly boomy correction.

Despite the issues encountered, overall it was possible to reach good results quickly (within minutes), with the corrected signal being both sonically preferable to the uncorrected signal, and audibly closer to the sound of the reference signal. In this scenario, where latency is crucial and both microphones are expected to be minimum phase systems there is no compelling reason to use a linear phase filter, unlike in the case of room and loudspeaker correction.[9]

## 4. FUTURE DEVELOPMENTS

Currently, correction applications require some manual adjustment of parameters. Whilst this may in future remain the best way to produce optimal results, it is also desirable to have methods of auto-regularisation that can estimate appropriate regularisation parameters, especially as setup time for concerts is often time-pressured. Preliminary work has been undertaken in this area, but as yet without any firm outcomes.

Further improvements for correction applications might come from improvements to the smoothing algorithm (for instance a perceptually-based algorithm for calculating smoothing widths at a given frequency, using Bark or ERB scales [24]), a common pole modelling approach in which the aim is to analytically identify poles for correction(see [10, 9, 19]) or a more relevant approach to phase issues. The addition of time domain deconvolution techniques to the toolbox is another avenue for potential exploration.

[8] Alternatively, one might use an instrument mounted microphone, to alleviate issues arising from the performer's movement.

[9] Note that the inverse of a minimum phase systems is also a minimum phase system.





In terms of end-user access, plans include an advanced set of tutorials on correction techniques, porting to other environments (*pd*, *supercollider* and *csound*), and the provision of compensation IRs for common mid-range microphones (such as the *AKG 414*). The aim of the latter would not be primarily for real-time correction for concert usage, but rather as a means to compensate for the microphone during measurement, thus producing more accurate results when a less coloured microphone is unavailable. AudioUnit and VST plug-ins are under development for applying corrections in a DAW environment.

# References


[1] HISS Website. [Online]. Available: http://www.thehiss.org

[2] E. Armelloni, A. Bellini, and A. Farina, "Not-linear convolution: A new approach for the auralization of distorting systems," in *Audio Engineering Society Convention 110*, 5 2001.

[3] A. Bassuet, "Analyses du spectre de puissance des instruments de musique et application pour restitution d'une scene sonore," Master's thesis, L'Ecole Nationale Supérieure Louis-Lumière, June 1999.

[4] M. Bouchard, S. G. Norcross, and G. A. Soulodre, "Inverse filtering design using a minimal-phase target function from regularization," in *Audio Engineering Society Convention 121*, 10 2006.

[5] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio Engineering Society Convention 108*, 2 2000.

[6] ——, "Advancements in impulse response measurements by sine sweeps," in *Audio Engineering Society Convention 122*, 5 2007.

[7] A. Gade, "Acoustics in halls for speech and music," *Springer Handbook of Acoustics, ISBN 978-0-387-30446-5. Springer-Verlag New York, 2007, p. 301*, vol. 1, p. 301, 2007.

[8] W. Gardner, "Efficient convolution without input-output delay," *Journal of the Audio Engineering Society*, vol. 43, no. 3, pp. 127–136, 1995.

[9] Y. Haneda, S. Makino, and Y. Kaneda, "Modeling of a room transfer function using common acoustical poles," in *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on*, vol. 2, mar 1992, pp. 213 –216 vol.2.

[10] ——, "Multiple-point equalization of room transfer functions by using common acoustical poles," *Speech and Audio Processing, IEEE Transactions on*, vol. 5, no. 4, pp. 325 –333, jul 1997.

[11] A. Harker. (2011) AHarker Externals. [Online]. Available: http://alexanderjharker.co.uk/Software.html

[12] A. Harker and P. A. Tremblay. (2012) HISSTools Impulse Response Toolbox (version 1.0). [Online]. Available: http://www.thehiss.org

[13] P. D. Hatziantoniou and J. N. Mourjopoulos, "Generalized fractional-octave smoothing of audio and acoustic responses," *J. Audio Eng. Soc*, vol. 48, no. 4, pp. 259–280, 2000.

[14] ——, "The application of complex-spectral smoothing in the analysis of room acoustics," in *ASME Greek Section National Conf. on recent advances in Mech. Eng., Patras, Greece,(2001 September)*, 2001.

[15] ——, "Errors in real-time room acoustics dereverberation," *J. Audio Eng. Soc*, vol. 52, no. 9, pp. 883–899, 2004.

[16] M. J. Heylen, Robert L. M.; Hawksford, "Interpolation between minimum-phase and linear-phase frequency responses," in *Audio Engineering Society Convention 98*, 2 1995.

[17] O. Kirkeby, P. Rubak, and A. Farina, "Fast deconvolution using frequency-dependent regularization," *to be submitted to IEEE Trans. Speech and Audio Processing*.

[18] O. Kirkeby, P. Rubak, P. A. Nelson, and A. Farina, "Design of cross-talk cancellation networks by using fast deconvolution," in *Audio Engineering Society Convention 106*, 5 1999.

[19] J. Mourjopoulos and M. A. Paraskevas, "Pole and zero modeling of room transfer functions," *Journal of Sound and Vibration*, vol. 146, no. 2, pp. 281–302, 1991.

[20] S. Neely and J. B. Allen, "Invertibility of a room impulse response," *The Journal of the Acoustical Society of America*, vol. 66, p. 165, 1979.

[21] A. Novák and, L. Simon, F. Kadlec, and P. Lotton, "Nonlinear system identification using exponential swept-sine signal," *Instrumentation and Measurement, IEEE Transactions on*, vol. 59, no. 8, pp. 2220 –2229, aug. 2010.

[22] A. Oppenheim, G. Kopec, and J. Tribolet, "Homomorphic prediction," in *EASCON '75*, 1975.

[23] D. D. Rife and J. Vanderkooy, "Transfer-function measurement with maximum-length sequences," *J. Audio Eng. Soc*, vol. 37, no. 6, pp. 419–444, 1989.

[24] J. Smith III and J. Abel, "Bark and ERB bilinear transforms," *Speech and Audio Processing, IEEE Transactions on*, vol. 7, no. 6, pp. 697–708, 1999.

[25] G.-B. Stan, J.-J. Embrechts, and D. Archambeau, "Comparison of different impulse response measurement techniques," *J. Audio Eng. Soc*, vol. 50, no. 4, pp. 249–262, 2002.

[26] P. A. Tremblay, "Pragmatic consideration in mixed music: a case study of La Rage," in *Proceedings of the International Computer Music Conference 2006*. ICMA, 2006, pp. 527–530.

[27] ——, *La rupture inéluctable, for bass clarinet and electronics*. Unpublished, 2011.

[28] P. A. Tremblay and S. McLaughlin, "Thinking inside the box: A new integrated approach to mixed music composition and performance," in *Proceedings of the International Computer Music Conference (ICMC 2009), Montréal, Canada, August 16-21, 2009*. ICMA, August 2009, pp. 379–386.