RETHINKING THE BOX: APPROACHES TO THE REALITY OF ELECTRONIC MUSIC PERFORMANCE

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 article extends certain methodological elements of the work presented in the Thinking Inside the Box Project [26], by exploring efficient software-based methods for improving sound reproduction within the concert hall. The key problems discussed are: (1) the detrimental effects of room acoustic and/or sub-optimal loudspeaker design on the frequency response of amplification systems for concert use; (2) the non-ideal frequency responses typically encountered when using close-microphone techniques or contact transducers (the methods most suitable for live applications, including the presentation of pieces utilising live processing). The need for pragmatic, musician-centric software addressing these issues is identified, along with a set of criteria relevant to musicians working in the fields of live electronic performance and interactive technologies.

These problems, along with proposed solutions and software tools, are investigated practically in both controlled conditions and real world scenarios, and the outcomes of experimentation and testing are discussed in detail. Real world testing of the methodologies and tools is considered essential in order to ensure that any developed tools and correction procedures are robust and viable for use within the constraints of a typical concert performance of electronic music.

Finally, a generic procedure is presented for rapidly generating and applying inversions to speaker/room combinations and close audio capture using software developed to satisfy the requirements outlined earlier (The *HISSTools Impulse Response Toolbox* [12].

1. CONTEXT

The research presented here builds on previous work by one of the authors into a number of issues relating to the concert presentation of works combining electronic and instrumental forces [26]. It extends this research by taking a complementary approach to issues involved in the live performance of electroacoustic music.

One strand of the prior research focused on recreating the sound of the concert hall when auditioning in the studio, thus addressing the dichotomy between the sound of the studio and that experienced in the concert hall. This study takes the opposite approach to the issue of disparities between studio and concert hall conditions by exploring efficient software-based methods for improving sound reproduction within the concert hall. Here we attempt to deal with key issues in sound reproduction and sound capture to allow more predictable results in the concert hall, with a better correlation to those experienced in the more ideal listening conditions of the studio. Whilst using a virtualisation of the concert hall might be considered as an accurate way of evaluating the realities of concert presentation, it implies a bias towards a certain venue type whilst composing. In reality there is often a need to present works in previously unknown spaces of differing size and design. Thus, this new approach seeks to ensure the highest level of success when considering portability between venues.

Practical limitations and requirements were key in the design of appropriate tools for accommodating the realities of presentation within concert hall scenarios. Concerts of electronic works are often highly constrained in terms of time and resources [6, p. 205-7], and these constraints must be taken into consideration in order to reach workable solutions. The proposals presented here have been implemented not only as test cases, but also deployed successfully in a number of professional concert scenarios, both in-house and at external events.

2. PROBLEMS

The starting points for our research are two distinct but related problems encountered when presenting electronic music in a concert hall environment.

2.1. Problem #1: System Frequency Response

Studio listening conditions imply a controlled, dry acoustic and monitoring with a similarly controlled frequency response. Ideally the interaction between room and loudspeaker result in a relatively flat frequency response, allowing for accurate judgement of balance and frequency content. In practice, the concert hall scenario is often far from flat, due to the effects of both the loudspeaker and the acoustics of the room. The latter results not only in the complex time characteristics that are heard as the reverberation of the space, but also the overall coloration of the sound. Particularly problematic are prominent room modes that result in boosts within narrow frequency bands [28]. These can effectively destroy sound judgements taken at the mixing stage, rendering previously correctly balanced materials inconsistently with misbalanced chords, non-fluid lines and inaccurate timbral balance.

Less ideal loudspeaker models can also have a significantly detrimental effect to the reproduced sound due to poor frequency response. However, regardless of the source of the issue, the combination of loudspeaker and room acoustics act as a system that affects the frequency response experienced by the listener. This system is in practice complex, with each individual path from loudspeaker to single-point listening position having its own characteristic frequency response. However, we can expect many common aspects to the frequency responses of these paths [11].

It should be noted at this point that, of course, any sound made within an acoustic space (such as instrumental sources) will be similarly affected by the room acoustics. When combining instrumental and electronic forces this can become a relevant consideration (as discussed in **section 4.3.1**). However, whereas instrumental performance allows a high degree of flexibility in instrumental balance and response to the acoustics, electronic materials rarely offer the same level of flexibility, if in fact they offer any flexibility at all, an issue which complicates the successful presentation of works mixing instrumental and electronic forces [8, p. 108]. Additionally, the reverberant qualities of any concert hall space are part of the experience of presenting work in that environment, and are not to be seen as purely 'problematic'. In fact, for some purposes (such as in the practice of diffusion), these form a key aspect of the performance. Thus, our goal is to improve practical listening conditions in the concert hall, mitigating the most significant and problematic aspects of the system, and not to negate all acoustic effects of the space.

2.2. Problem #2: Close Microphone Capture

When presenting pieces that require processing of a live instrument or voice, close instrument capture with a directional microphone or pickup is often used to drive the processing. This approach minimises spillage from other sound sources (instruments or loudspeakers), and reduces the potential for feedback. Choices of close capture may also be a matter of practical convenience (for instance the use tie-clip type wireless vocal microphones to allow mobility). However, the tonal balance of close capture is often non-ideal, due to the proximity effects and instrumental radiation patterns [7]. Given that the reproduction system may also be non-ideal (as discussed above), the resulting sound (before any processing) is often significantly coloured, causing issues of blending with any acoustic sound. Here, the amplified/processed sound takes on an artificial quality due to formantic deformation resulting from the combined frequency response of the capturing and reproduction systems.

An ideal solution would allow for minimal feedback and spillage, whilst representing more faithfully the sound of the instrument or vocalist.

3. PROPOSED SOLUTIONS

3.1. General Approach

We take the approach of modelling both of these problems as inversion problems, with the assumption that the unwanted effects of any system can be modelled as one or more linear time-invariant filters that can be inverted to form correction filters for the reproduced or captured audio. The effect of these correction filters is to mitigate the problematic effects of the system. Furthermore, assuming the filters (and associated inversions) are finite, the correction filters can be applied using convolution. Such a model views the effects of loudspeaker and room as a single convolution operation. As convolution is commutative, correction can be applied *before* the loudspeaker/room system with the same result as applying the correction post system (which would be impossible in any case).

Thus, the procedure follows a general three-part process:

1. The system in question is measured to provide a set of frequency responses for inversion.

2. The measured responses are processed to form practically usable correction filters.

3. The correction filters are applied from impulse responses (IRs) via convolution (either pre-output stage in the case of loudspeaker/room correction, or immediately post-input in the case of close capture).

In order to measure the effects of close capture we follow a variation of the approach proposed by Bassuet [5], who suggests that the effect of close capture can be modelled by comparing the close capture with a microphone that has been placed to provide optimal tonal balance. The instrumentalist or vocalist is then required to play across the range of their instrument (Bassuet suggests two full-range chromatic scales, one loud, one soft). We also propose the addition of a set of noise-based instrumental techniques, in order to produce significant frequency content that may not be provided by conventional playing, but nonetheless may be musically relevant to the presentation of some works. The difference between the close and more distant captures gives the filter to be inverted[1].

As electronic musicians are typically computer-based, we propose in both cases a software-based solution that can be applied using readily available equipment that already forms part of the musician's toolset.

[1] Here the term 'difference' implies the division of the two spectra, rather than a subtractive operation

3.2. Existing Tools

3.2.1. Literature

The subject of room and loudspeaker correction has been explored in some detail, although not without some disagreement regarding effectiveness/viability (see [10, 20, 17, 21]). The existing literature forms the basis for many of the algorithms and approaches taken here. The area of microphone correction is less explored, although there is significant technical overlap in our approaches to the two identified issues. Importantly, as musical practitioners, our goal is to develop a solid set of tools that can be deployed in practice for concert hall presentation, rather than to explore only theoretical, or experimental results. Additionally, our interests are in determining practically viable solutions combining the most robust and potent approaches from the pre-existing literature. Specific algorithms and techniques are discussed below where relevant.

3.2.2. Pre-existing Software and Hardware Solutions

These fall into two categories:

The first category addresses only parts of the technical problems explored (such as deconvolution or realtime convolution), suggesting that a user might be able to create a chain of tools to create and implement appropriate creation filters. These include open-source tools such as FScape [24](for batch deconvolution and other spectral processing), or any of the many available commercial or non-commercial convolution plug-ins (such as Audio Ease's Altiverb [4], Wave's IR-1 [27] or Lernvall Audio's LAConvolver [15]). In the latter case, many of these are designed specifically for reverb usage. This means that the user cannot be certain to apply a IR-based correction filter exactly as calculated, due to possible additional unwanted automatic IR processing (trimming/automatic gain matching etc.). However, the key issue when considering this category of available technologies is finding a suitable tool for performing filter inversion appropriate to room, speaker and microphone correction tasks. Whilst there may be technical tools available that enable expert users to create correction filters (such as via MathWork's signal processing toolbox for MATLAB [19]), these offer a very generalised interface to generic signal processing tools, rather than a rapid and viable solution to a specific problem. The use of tools in this category is impractical due to both the high level of technical knowledge required, and also the need for a simple and speedy process in the concert hall.

The second category of solutions are fully-formed room correction software or hardware solutions (such as Real Sound Lab's CONEQ [23], IK Multimedia's ARC [16] or the open source DRC [25]). These, are commonly targeted at studio or home listening usage and tend to offer an all-in-one black box solution to the problem of room correction (the exception being the somewhat experimental DRC - an open source project), with little flexibility. This leads to a set of practical limitations including restrictively short FIR lengths, induced system latency (not suitable for realtime/live work), inflexible and lengthy measurement procedures and channel number limitations.

3.3. Musician-centric Solutions

To suitably address these problems requires that any tools are suited to the realities and limitations of concert hall performances. As previously advocated in [26], pragmatism is of prime importance. We therefore take into account a number of concerns specific to the musician presenting electronic works in a live context.

3.3.1. Noticeable Subjective Sonic Improvements

As musicians we must insist on the primacy of the ear for judging the quality of results. Improvements must not introduce noticeable detrimental effects. Our goal is a practical gain, rather than a theoretical one.

3.3.2. Rapid Deployment

Rehearsal and preparation time for electronic music concerts is often tight, and necessarily the focus must be firmly on musical, rather than technical concerns to ensure a successful presentation of the work. Lengthy or overly complex calibration procedures are therefore not viable.

3.3.3. Ease of Use

There should not be a requirement for in-depth technical knowledge on the part of the user. Many composers and performers of electronic music are non-specialists from a technical viewpoint, and should be able to achieve good results without having to acquire substantial additional knowledge or expertise.

3.3.4. Flexibility

Tools should be able to accommodate variable numbers of channels/measurement procedures and offer the ability to fine-tune results quickly for a particular task or space.

3.3.5. Low Latency for Live Use

Particularly in the case of interactive or live work (as opposed to playback), additional latency is undesirable within the system. Ideally, corrective filters should add no (or negligible) latency.

3.3.6. Appropriate Environment

Any software tools developed should fit into the workflow and toolset of a typical electronic musician.

4. DEVELOPMENT AND CASE STUDIES

4.1. Process

4.1.1. Empirical Methodology

The starting point was to implement and test a variety of known techniques from the room correction and spectral processing literature, in order to develop a combination of steps that would produce practically usable, sonically convincing results. Our approach was to start from the least sophisticated techniques and increase sophistication as necessary in order to determine a minimally complex procedure of practical application, rather than assuming that the technically most elaborate solutions would be optimal (which here should be read as the most suited to the design criteria, rather than simply as indicating technical optimality). Throughout the process, successive developments were informed by practical evaluation and audition; both in the selection of algorithms, and in refining steps in the correction processes.

4.1.2. Synthetic Tests

Algorithms were first tested using predictable and known signals to ensure correctness and basic sanity. For instance, the convolution of single unit value sample (a dirac delta signal) with itself should yield a third identical signal. Any deviation reveals deficiencies in the coding of the algorithm.

In other situations, synthetic tests are a good measure of the limits a real world solution might hope to achieve. For instance, from a theoretical viewpoint, if a signal is convolved with an impulse response (such as that of a room/speaker combination), and then the result convolved with the inverse impulse response, the result should be an exact replica of the input. However, in reality the impulse response is unlikely to be exactly invertible (in a numerical sense) [21], and typically a method to avoid filter blowup must be employed, alongside the use of a modelling delay to generate a causal inversion filter (both issues are addressed in [18]). Utilising such techniques it is possible to evaluate the efficacy of direct impulse response inversion given a completely stable impulse response (this is equivalent to an attempt to dereverberate a signal). Here we found significant issues with audible pre-ring or post-echo, with noticeable low-level 'ghosting' of the input, despite the theoretically optimal conditions of a time invariant response. In practice, the frequency response(s) of a room/speaker combination will vary at least slightly over time and space, and real world test results of such an approach gave totally unusable results, not even approaching convincing from a listening position coincident with the microphone. However, given that the results from a synthetic system produced audible artefacts, it was clear that such an approach would not be capable of producing a viable solution in practice, even given a significant narrowing of the gap between real world and synthetic results.

4.1.3. Module Design and External Testers

MaxMSP was chosen as a relevant and dominant platform for musicians working with instruments and electronics, or interactive media. It has a wide user-base with varying levels of technical expertise. The aim in design was to follow the message format and naming conventions set by the standard library of objects in MaxMSP. This aspect of the process was also aided by an international set of expert testers, in a range of musical and technical fields, all well-versed in the MaxMSP environment. The process of feedback on interface (as well as functionality) was invaluable in ensuring that messages and object properties were consistent both with the core library, and across different objects.

4.2. In-house Testing

4.2.1. University of Huddersfield Studios and Initial Prototyping

Initial tests focussed around a controlled studio setup. Although these do not provide a realistic model for a concert hall setup, the University of Huddersfield studios provided a good reference point for basic algorithm tests, and providing a consistent and known setup to audition improvements in a more controlled environment.

Early prototyping focussed on the integration of measured power levels across the spectrum using a known noise signal (typically pink or white). By comparing the relative power spectrum of the two signals (output - the noise signal, and input - the same signal picked by a microphone in the space) it is possible to determine a inversion signal. It is important to note that ideally a flat microphone is required to make measurements, as the microphone is part of the system under measurement. We used a *DPA 4006* omnidirectional microphone for all speaker correction tests, as it is a class one microphone design. The approach is a similar approach to the one suggested by Bassuet for performing microphone correction [5], where the power spectra are estimated using time-integration of successive STFFT frames. In our case, we also averaged measurements from a number of listening positions to represent a wider listening area more accurately, rather than that of a a single listening point. Whilst reasonably convincing results were obtainable, such an approach suffers several practical and theoretical limitations:

Measurement Time - This method of time integration is highly approximate, and results vary wildly between subsequent frames. Longer measurement times significantly improve the reliability of the measurement as the variance of the power spectra mitigated by the averaging of a number of frames, with larger numbers of frames resulting in less variance. However, the results do not compare particularly favourably with more advanced methods of impulse response measurement. As rapid usage is a paramount concern, we did not consider the time/reliability trade-off to be acceptable for practical purposes.

Limited Frequency Resolution - This method depends on averaging successive STFFT frames in order to calculate the power spectra. Larger FFT sizes result in higher variance for the same measurement length, which makes the technique impractical for long filter lengths. Thus, achieving acceptable measurement reliability in a reasonable time frame necessitates a limited frequency resolution.

Frequency Smoothing - As mentioned, typically the results of this kind of measurement are highly noisy across the frequency range, due to both the complex nature of room frequency responses, and the approximate nature of the technique. Therefore it was necessary to apply smoothing (see [13, 14]) to the power spectra prior to inversion. This results in spectra that are less noisy and better represent the overall frequency response of the room/speaker combination, rather than the details of reverberation. We verified that corrections based on smoothed spectra were both substantially more convincing aurally than those without smoothing (which suffered many detrimental peaks/notches), and more applicable to a number of different listening points.

Time Alignment - The assumption made here is that the two power spectra equate to the exact same period of the known signal, ideally allowing each frame of the STFFT for each signal to be correctly aligned. However, the latency induced by the audio IO means that the signals are not aligned in realtime. One option here is to estimate the time delay of arrival before averaging, in order to correctly align the signals. However, this adds an additional measurement requirement, for which a naive method (such as cross-correlation) may return an incorrect result, thus reducing the reliability of the main power spectra measurement. Ideally we would avoid both the necessity for another calibration (which is necessary at each new measuring position) and the possibility of significant error in the measuring process.

Lack of Time/Phase Information - This method discards any time or phase information. Thus, it it gives equal weighting to all reflections in determining the frequency response, regardless of how late in the system's impulse response. As concert halls may exhibit long reverb tails this means the correction is based on the entire length of the impulse response, which may be a poor representation of what the listener hears as the direct sound. Reverb tails tend to increase in coloration over time (due to the cumulative filtering effect of repeated reflections off similar surfaces), and thus including the late reflections in the inversion potentially results in an over-correction of the perceived frequency response, which better corresponds to the earlier part of the impulse response.

Minimum Phase Corrections - As latency was a key concern, it was decided to convert filters to minimum phase, giving the same amplitude response in the frequency domain, but with the energy of the filter placed as early as possible to minimise latency. Before this, the filters (in the absence of phase information) were linear phase, requiring a latency of half the filter length. Minimum phase filters were found to provide negligible latency.

4.2.2. A second approach

Due to the restrictions of an approach based on time integration of power spectra, it was decided that a direct method of impulse response measurement would be preferable. For this, we employed the exponentially-swept sine wave method (or ESS - see [9]), along with less accurate methods based on the direct deconvolution of known noise signals (which may be preferable for occupied spaces, although less accurate results). As before, we propose averaging measurements from several listening positions, and smoothing the overall result (for the same reasons as given above). As well as providing a more accurate measurement of the system, without the need for explicit time alignment, this approach also allows for the truncation of the impulse response (with appropriate fading out), in order to treat only the early part of the response, ignoring the longer reverb tail. This was found to be of particular use in larger venues, where a truncated response of around 300ms gave more perceptually pleasing results. Inverting the whole impulse response tended to overemphasise the upper frequency range. This is not unexpected, as most venues will exhibit a significant rolloff of high-frequencies over time.

As our concern is the correction of the amplitude spectrum, we favour the dismissal of phase entirely, and thus smooth the power spectrum only (after truncation). The complex smoothing proposed in [14], arguably conflates phase and amplitude information, and we found this conflation to produce undesirable and nonsensical results with long filter/measurement lengths.

It is important to note that individual channels are inverted separately in a multi-mono approach, although this process is carried out in parallel, so as to preserve relative level differences between channels (thus correcting for them during the inversion process).

For the inversion several methods of combatting overcorrection were employed, including regularisation [18], which in practice we found to be sufficient for our purposes. However, the other methods explored are also available in the final software for applications where they might be relevant.

4.2.3. HISS System

Both the initial prototypes and the revised approach were tested using *HISS* (Huddersfield Immersive Sound System [1]); a multichannel sound system for the presentation of electronic music.

Listening was carried out from a range of on- and off-axis positions within the concert hall. Tests of recorded material were played back over a range of loudspeaker types, both with correction and without for A-B comparison, taking care to match the levels of both versions in order to avoid bias due to loudness. This balancing was performed manually, as it is the perceptual level that must be matched, rather than the numerical amplitude level (such as RMS), which will most likely not result in perceptually matched results.

Speakers tested were:

1. Meyer UPJ-1P with UMS-1P subwoofer

2. Bose L1mk2, with B1 subwoofer

3. Bellecour 360Sound

Typically, results were clearer with the corrected audio, and improvements were noticeable, once appropriate parameters had been chosen. The speaker design was an important factor, both on the efficacy of the procedure, and suitable choices for parameters. In particular, correcting the *Meyer UPJ-1P* setup gave more subtle results than the other speaker designs, which have significantly more coloured frequency responses (both in their published specifications and the measurements taken). **Figure 1** compares the frequency responses of the *Meyer* and *Bose* models (responses shown with 1/3 octave smoothing). It is apparent that the *Bose* system has noticeably larger deviations in frequency response (for instance the obvious dip around 4.8kHz) and also less low and high frequency extension than the *Meyer* setup. **Figure 2** shows the *Bellecour* speaker also. The more coloured *Bose* and *Bellecour* models exhibited a more dramatic and obvious sense of correction. However, whereas with correction of more coloured speakers the improvement was a noticeably flatter frequency response, for the *Meyer* speakers, subtleties of the mix became more apparent, such as stereo imaging, and depth perspective.

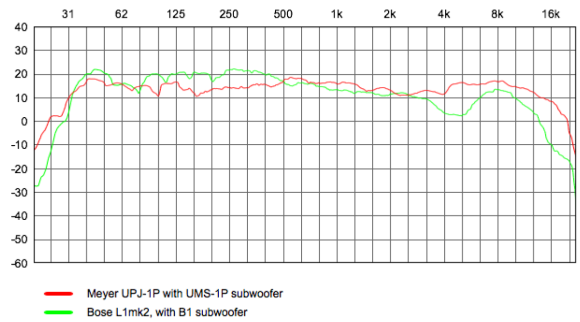

Figure 1. *Meyer* and *Bose* Frequency Responses *(Hz / db)*

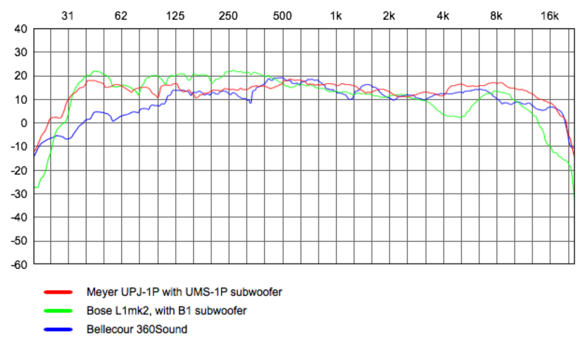

Figure 2. Loudspeaker Frequency Responses *(Hz / db)*

Extremes of frequency were a particular listening focus, as both very high and low frequencies tended to be the most obviously affected by parameter changes in the correction. Typically, regularisation can be minimal for the majority of the frequency range (having almost no noticeable effect on the results of the inversion), but as the system rolls off at either end more regularisation is necessary to prevent the system from becoming overdriven. Before the onset of distortion, overcorrection results in a harsh and fatiguing sound, the result of driving the extremes of the spectrum beyond the viable limits of the speaker/amplifier designs. Thus, it is necessary to exercise care in choosing both the amount of regularisation at the extremes, and also the frequency points at which to start applying noticeable correction. We employed trapezoidal regularisation shapes in the log-log domain (log frequency vs. dB level). As the more coloured speaker designs tested had less low/high frequency extension than the *Meyer* models, the trapezoid needed to be narrower and steeper in order to produce useable results for these speakers. Obviously, this also implies that the result has less low- and high-end content in these cases.

The earlier time-averaging approach (see **section 4.2.1**)was less convincing in dealing effectively with low frequencies. This is probably due to the reduced frequency resolution resulting from shorter filter lengths (for the direct IR measurement approach, filters of arbitrary length are possible, potentially lasting for several seconds), and the approximate nature of the measurement; which is liable to be more representative at higher frequencies. This is due to a greater degree of smoothing at high frequencies, and thus each high frequency bin value is result of the averaging of more data points, reducing the variance of the measurement. With this earlier approach, issues of a perceived lack of bass were apparent. Using the ESS measurement technique as the basis for inversion resulted in a more even low end, in which balance between subwoofer and full-range speaker was improved, especially in the case of ‘hyping’ of the low end, which tends to sound more instantly attractive when quickly calibrating subwoofer levels by ear. The control of low-mid range frequencies, where the most problematic room modes tend to reside was also more precise using correction based on ESS measurements.

Smoothing levels acted consistently on different models of speaker, and acted as a means to control the level of detail in the correction. A suitable choice is a matter of compromise between competing concerns of accuracy and listening area. Too little smoothing resulted in a correction that is problematic at positions far from the measurement positions, with the potential for noticeable filter ring. Excessive smoothing results in a very broadband correction, which corrects only the very general shape of the frequency response, but lacks the detail to deal with prominent individual peaks and troughs in the systems response accurately. Unlike with the regularisation parameters, it became clear that default parameters could be chosen without regard to the specific system, with predictable results in each case. The method of smoothing employed uses variable-width smoothing across the frequency range, with the user specifying the amount of smoothing at 0Hz and the Nyquist frequency in cycles/sample (units of normalised frequency). We found values of 0 cycles/sample at 0Hz and 0.08 cycles/sample at the Nyquist frequency to produce reliably useable results, and propose these as defaults, although tweaking to taste may be desirable for specific scenarios and applications.

We also auditioned linear phase filters against the minimum phase equivalents, as an A-B comparison. Even when individually correcting channels, no perceivable difference was noticeable when switching between phases, probably due to the similarity and smoothness of the correction filters. Although one might expect that the difference between filters may result in timing issues between channels, it would seem that the ear is simply not sensitive to the kinds of differences found in practice in the concert hall, and it was impossible to blindly identify the phase of the filter. We expect that if the low frequency components of two correction filters were significantly divergent, the phase differences between filters may result in noticeable timing issues; in practice we have not encountered this situation within a concert hall scenario. On the other hand, the latency when using minimum phase corrections can be considered negligible, making the correction filters suitable for realtime work.

4.2.4. Microphone Correction

As the problems of speaker/room correction and microphone correction are both fundamentally concerned with correcting system frequency responses, the methods for solving them practically can be substantially similar. Here the significant difference is that it is not possible to measure the impulse response of the system of interest directly (the difference between a close microphone and a more optimally positioned reference microphone) using a test signal. In part, this is because we wish to capture differences resulting from the radiation pattern of the instrument, which are typically highly frequency dependent. Thus, we must use the instrument itself as the 'test signal'. Whilst our starting point was based on time-integration of power spectra, we prefer to attempt to estimate the impulse response more directly. For this, we propose the direct deconvolution of two recorded inputs (one from each microphone or capture method). As these test files will typically be 40-60 seconds in length the required FFT size is relatively large, but modern computing power is sufficient to calculate the results of such a deconvolution in a few seconds. The resultant IR is a noisy approximation of the real impulse response. The noise will be dependent on the recorded material, which should give a good signal to noise ratio across the entire frequency spectrum (as in the 'ideal' swept sine wave used for accurate impulse response measurements of speaker/room combinations). Hence, we advocate the use of a glissando or chromatic scale across the full range of the instrument, as in [5], with the addition of noise-based techniques as discussed in **section 3.1**.

The rest of the proposed method remains similar to that for room/speaker correction. After deriving an impulse response, smoothing is applied followed by inversion with regularisation. Finally the result is converted to minimum phase, and optionally truncated (with an appropriate fade).

It should be noted that alongside microphone correction we advocate use of speaker/room technologies (as in this example) to avoid the potential situation in which the balance achieved by microphone correction alone is rendered void by deficiencies in the speaker system or room.

4.2.5. Bass Clarinet Microphone Correction

Recordings were taken of bass clarinet in three contrasting venues; one a very reverberant space, the second a dry concert hall, and the last a very dry studio space. In each space three microphones were used:

1. An omnidirectional DPA 4060 attached to the music stand
2. An omnidirectional DPA 4006 at one metre
3. A cardioid DPA 4011 at one metre

Recordings were taken of calibration scales, broadband noise techniques (breath sounds/tongue slaps etc.) and short passages of music. The resultant files were used to test the procedure, and to make a comparison between the use of cardioid and omnidirectional polar patterns for the reference mic. After deriving correction filters from the calibration recordings, the other recordings were then played back, with the close capture from the music stand corrected, and compared to either both of the captures at one metre.

Application of the correction resulted in a frequency response closer to that of the more distant captures, although with a higher ratio of clarinet to reverberation, again with a need to pick appropriate parameters for the procedure. For this application, the smoothing parameters appear more sensitive than for speaker/room correction, with lower smoothing values resulting in emulation of the room modes in the correction, as well as significant filter ring. Thus, as the reverberation will depend on the venue, proposing universal defaults is more problematic. As the method of measurement is inherently approximate, some small amount of low frequency smoothing was found to be beneficial in this scenario (although only in the range of 0.01 cycles/sample at 0Hz, as substantial smoothing in the low frequency range provides very poor accuracy of correction). The concerns are also different with regards to applicability over a wide area, as the variance between the position of the instrument and microphone will vary far less then the maximum distance between different listeners relevant to a speaker/room correction procedure. Therefore, lower levels of smoothing overall are viable, with the emulation of a more distant capture becoming more convincing as the smoothing level is reduced. For this application, more judgement is required by the user as to the balance between improvements gained from lower levels of smoothing as opposed to issues arising from reduced levels of smoothing due to the inaccuracy of the measurement technique (partly due also to the high level of variance in radiation patterns across the range of most instruments). Regularisation parameters behaved much as with the speaker/room correction application, with the caveat that a higher level of regularisation overall was necessary to combat the higher levels of variance in the measurement, and avoid excessive ringing.

The cardioid capture exhibited perceivable low frequency loss at a distance of one metre (although also a higher ratio of clarinet to reverberation). This is not unexpected, as the microphone is advertised as at is flattest at 30cm from the source, and deviations from this positioning will exhibit the proximity effect. Thus, the use of a cardioid microphone as a reference is problematic for instruments producing significant low frequency content. Closer placement of the reference microphone would detract from the aim of taking a suitably balanced sound as a reference.

4.2.6. Acoustic Bass Guitar Piezo Pickup Correction

This set of tests considered the correction of an acoustic bass guitar piezo pickup. In this case, two clear issues were present. One was the compromised frequency response of the piezo system, which is typically an issue even when using high-end transducers (e.g. [2]). The second issue is that the transducer system is designed to pick-up the vibration of the strings directly, and hence does not reproduce much of the resonances from the body of the instrument that are crucial to the acoustic sound.

Recordings were taken from the piezo pickup, as well as from a *DPA 4006* microphone placed at one metre from the source. Two passages were recorded: a set of quiet and loud chromatic scales for calibration/generation of the correction filters, and a passage of improvised over the range of the instrument. As with the clarinet microphone correction the regularisation parameters were more sensitive to fine-tuning, especially in the low frequency range, where poor choice of parameters at low frequencies resulted in either a lack of apparent low-end, or an overly boomy correction.

In this case the smoothing parameters were important in controlling not only the amount of room resonances simulated by the correction, but also the extent to which the resonance of the body (captured by the reference microphone) were simulated. Whilst very low smoothing levels resulted in excessive filter ring, and emulation of the reverberant characteristics of the space, high smoothing levels tended to remove the characteristics of the guitar's body. Whilst the frequency balance was clearly less coloured with this highly smoothed correction than with the raw piezo capture, it resulted in a sound with a more synthetic quality. Thus, in this circumstance, some compromise between correcting only the broad shape of the frequency response, and the finer details necessary to virtualise the instrument's body is necessary, according to requirements and taste. However, acceptable results were again achievable within minutes, and the results of parameter changes were predictable. Notably, if measurement is made in a well-treated acoustic environment (such as a studio) then the correction filters have the great advantage of portability, as the output from the piezo transducer will not be significantly affected by changes of venue.

4.2.7. Vocal Microphone Correction

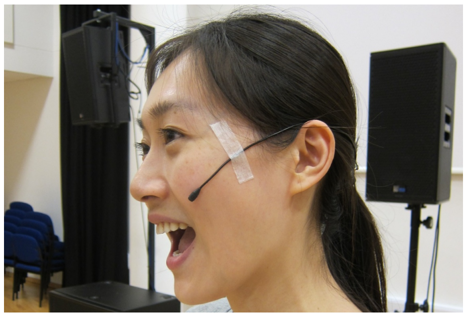

Figure 3. Vocal Mic Position

Further tests of the microphone correction procedure involved testing with tie-clip style microphone (a *DPA 4060*) taped to a soprano singer's face in a concert hall environment, typical for operatic amplification (see **Figure 3**). This mic'ing approach, enables both close-capture and mobility for the singer, but does not reflect well the tonal qualities of the voice as heard from a normal audience listening position. Here, the relevant comparison was between the amplified or electronically manipulated voice, and the sound of the singer unamplified. The reference microphone used was again a DPA 4006, and tests were performed of both amplification and various processing types (delay, distortion and additive resynthesis of the input). The reference signal was chosen as a set of noisy vocal techniques (to provide sufficient broadband energy, especially for very high frequencies), followed by several full-range glissandi using different vowel sounds.

The amplification system consisted of five *Meyer UPJ-1P* active monitors, which were also corrected using the speaker correction procedure. The frequency plot for the microphone correction is shown in **Figure 4**. Here we can see that a significant boost is necessary in the high and low frequency ranges, with a noticeably uneven curve in the mid-range showing, for instance, two clear peaks and troughs of around 5-6dB around 500Hz. A substantial correction in the low frequency range is also evident, although this is too low to affect the sinusoidal components of the soprano voice, and thus will be relevant mainly to transient and noise-based content.

Auditioning was carried out from a variety of both on-axis and off-axis positions within the audience space. From both on- and off-axis listening positions the corrected sound blended more naturally with the unamplified voice, and created a much more convincing musical hocket when using a delay with correction than without.

Notably, the frequency response of the uncorrected microphone sounded more clearly 'amplified', with a comparatively dull sound due to the comparative lack of high-frequency content.

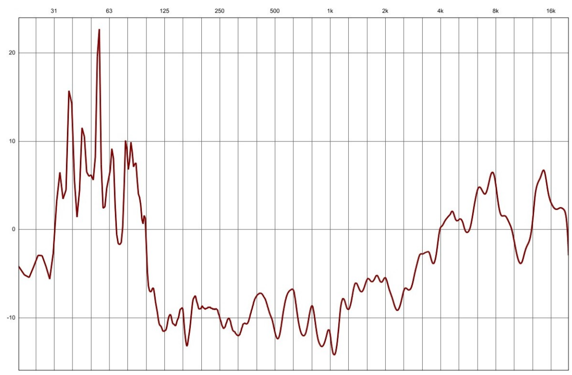

Figure 4. Vocal Mic Correction Curve *(Hz / db)*

4.3. Real World Case Studies

The techniques discussed above have been employed in a number of real world scenarios in order to test their practicality and robustness to the demands of the realities of the concert scenario.

4.3.1. La Rupture Inéluctable

La Rupture Inéluctable by Pierre Alexandre Tremblay is an interactive work for unamplified bass clarinet and *MaxMSP*. The electronics are presented over three speakers in LCR format behind the performer, and the majority of the electronic part is derived through processing of the live bass clarinet. Other materials in the electronic part are fixed soundfiles.

Composition of the piece was concurrent with the research into correction procedures, and the piece was designed to make use of both the microphone and the speaker/room correction. The piece has subsequently been presented in a number of venues of different types utilising both of these approaches. Notably, in concert the clarity and presence of the fixed parts is consistently improved by use of the correction, as high frequency rolloff is a typical characteristic of speakers and rooms, and the musical materials make use of digital distortion techniques producing a broad frequency content with significant high frequency content. At the premiere performance, in a large church-type venue with a 3-second reverb time, the speaker system consisted of three *Meyer UPJ-1P* loudspeakers, each with a dedicated *M1D-Sub subwoofer*. Four listening positions were measured (left and right from front row and several rows further back). Here, we found that truncating the impulse responses significantly increased the perceived improvement of the speaker correction, which otherwise was overly aggressive in the high frequency range, presumably due to cumulative high frequency rolloff over the duration of the impulse response. In this venue listeners from a range of seating positions noted that the electronics were present and penetrating, in a manner not usually experienced in that venue. However, in this particular case the acoustic, unamplified bass clarinet was not as upfront in sound, due to the acoustics of the room, and there was a slight mismatch between instrumental and electronic parts. This might point to the necessity in more reverberant venues to account for this issue, perhaps by correcting to a frequency response that is not flat. This could be achieved, for example, by convolving by the desired frequency response post correction. This effect was not replicated in any of the less reverberant performance venues.

The microphone correction was in all cases noticeable in maintaining the spectral balance of the instrument. A significant part of the piece relies on additive synthesis in which the amplitudes of each frequency component replicate those measured at the input to the processing. Thus, correcting the feed to the processing results in a more faithful and even balance between different sinusoidal components.

4.3.2. New York Counterpoint

A second case study involved the presentation of Steve Reich's *New York Counterpoint* in which the recorded parts are spatialised using a multichannel speaker system. The set-up was a semicircle of seven speakers, spaced symmetrical around a central speaker placed directly behind the performer. The concert hall was of a large design with a raised staging area, and a reverb time of c. 1.5 seconds. In this case, as the concert was an external event, the sound system was supplied by the organisers, and was far from ideal.

Initially, rehearsal took place without speaker correction, as preparation time was extremely tight for this and other pieces in the programme. However, during the initial rehearsal period it became clear that the system was creating significant issues of balance both between tape and live performer, and also between different lines in the recorded parts. Typically, this was made most apparent by individual notes sticking out of uniform musical textures according to their spatial positioning. Broadband equalisation was ineffective at significantly controlling such issues. This is a clear example of a situation in which manual treatment of the problems over seven channels would have been extremely time-consuming and would require a high degree of expertise and listening skills, with no guarantee of accurate results.

Thus, it was decided before the final rehearsal that the speaker correction would be necessary to achieve musically acceptable results in the concert hall. Four measurement points were used corresponding to laterally central seating positions in the two banks either side of the central aisle, at distances of approximately 5m and 10m from the stage. These positions were chosen so as to encompass the majority of the front-half of the audience seated within the concert hall floor space. Measurement was performed using two microphones in two passes. The best available microphones were a pair of *AKG 414s* in omnidirectional mode (with no LF rolloff). Although far from class one, it was assumed that the apparent issues with the speaker system significantly outweighed any deviations from flat in the microphone frequency responses.

The measurement procedure used a 15 second sweep for each speaker, thus totalling under 2 minutes for each pass. All aspects of the processing were automated so as to calculate immediately on the completion of the second pass, with the option to reprocess with different parameters after measurement.

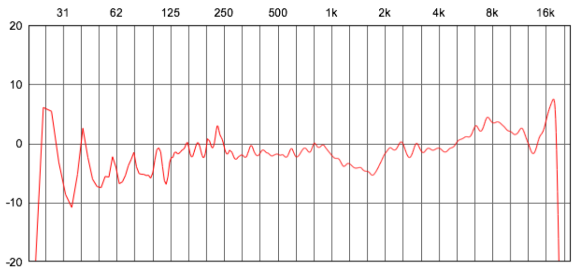

Figure 5. NYC Speaker Correction Curve *(Hz / db)*

Quick auditioning was carried out for the correction of each channel, using a known full-range audio recording. Immediately, issues of balance were noticeably improved throughout the seating area, and the tonal differences between different speakers became less apparent. Some slight harshness resulting from extreme high frequency content was noticed, and the regularisation parameters were adjusted to rolloff the high frequency correction at a lower point. This removed the sense of harshness. The correction filters were then loaded into the spatialisation patch for the playback of *New York Counterpoint*, and the piece auditioned. Balancing between notes within a chord was substantially improved, and lines rendered more fluid. **Figure 5** shows the final correction curve for one speaker, which shows a clear set of peaks in the low midrange, explaining the issues of chord balance in the uncorrected playback. The entire correction procedure, including rigging of microphones, tests and adjustments was carried out in under 30 minutes.

4.3.3. Other Pieces

The authors have now implemented speaker/room correction for a number of pre-existing pieces in live concert presentation, either combining instruments and electronics or for fixed media alone.

Retrofitting older pieces in this manner, or creating a suitable playback environment for fixed media pieces can be done in less than an hour, prior to the concert day. We have found this approach to be very beneficial to concert presentation in a range of venues. Results are most striking when the uncorrected system exhibits significant issues, and the level of user expertise is far lower than would be required for manual equalisation, and with more accurate results.

5. OUTCOMES AND CONCLUSIONS

5.1. Outcomes

5.1.1. The HISSTools Impulse Response Toolbox

The tools developed for correction applications form the *HISSTools Impulse Response Toolbox*, a set of modular tools for *MaxMSP* that deal with convolution, deconvolution and other impulse response-related tasks. The design of these tools reflects the concerns set out in **section 3.3**, with particular focus on maximising reusability and speedy deployment. As a result of this methodological research process, it became clear that providing a single black-box solution to deal with correction applications would reduce the flexibility of the resultant software beyond a desirable point for supporting a range of users with specific and variable requirements of system size and setup. Thus, the decision was made to create a set of objects addressing individual subproblems (e.g. IR measurement, filter inversion, realtime convolution etc.) that would enable straightforward combinations of objects to be used, but would also generalise the tools to a much wider range of applications. More on the specifics of the *HISSTools Impulse Response Toolbox*, and more technical details of the correction procedures can be found in [12].

As the design of the HIRT facilitates usage beyond the direct applications discussed here, since its public release the toolbox has been employed by a number of academic and non-academic practitioners to solve a range of convolution, deconvolution and measurement problems. This includes uses within psychoacoustics, acoustic virtualisation, impulse response measurement and reverb applications. Examples are the use of various parts of the toolbox inside of a set of reverb-related Max for Live devices released as part of Ableton's *Max for Live Essentials Pack* [3], as well as use of the externals as part of Rui Penha's *Spatium* package [22]. There is also third-party interest in porting to other environments, including pd and Supercollider.

5.2. Currently Proposed Methods and Technical Discussion

A generalised procedure for performing frequency response corrections is shown in **Figure 6**. For microphone/close capture correction, a single measurement suffices, based on a comparison of a reference microphone with the close capture source. For this comparison, broadband material evenly covering the spectrum is desirable, for which we suggest full-range chromatic scales or glissandi, with the optional addition of noise-based content that produce significant frequency content outside of the range produced by conventional playing or singing techniques. For room/speaker correction any number of measurement positions may be averaged (the long as they are well-balanced in the field of interest, for instance left/right for stereo reproduction). Our studies suggest that three positions are adequate for good results. Our approach for speaker/room is to treat all channels of a system sepa-

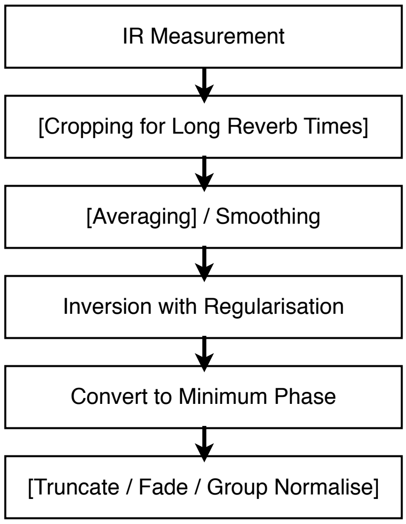

Figure 6. Generalised Correction Procedure

rately, but in parallel (preserving any level relationships between channels). Although we opt to dismiss measured phase entirely in our procedure this has been demonstrated as suitable to our needs through a number of practical tests and case studies. Likewise, we have not experienced practical issues with conversion to minimum phase filters, which enables latencies suitable for realtime work. However, should be noted that presentation formats in which phase relationships are paramount (such as in ambisonic presentation formats) may suffer under these conditions, and have not been tested.

The *HISSTools Impulse Response Toolbox* is not itself limited to implementing this approach, and is also suitable for inversion problems in which the measured phase information must be corrected (given a suitable modelling delay to deal with non-casual components of the impulse response's inversion), or for implementing variants on these steps. The software is thus open to future developments, and additions enabling more sophisticated correction techniques. However, our case studies show that significant benefits can be achieved rapidly from the current approach, which is both pragmatic and robust.

5.3. Future Work

Currently, some user guidance is necessary to fine-tune parameters for smoothing, regularisation and cropping lengths. Whilst in many cases default values will work sufficiently well, it is desirable to minimise any requirement for the user to oversee the correction process. Further investigation of methods for automating parameter choice would be beneficial, along with a more comprehensive understanding of the impact of the choice of cropping length for IRs prior to inversion. Although some preliminary investigation has been carried out in this area of auto-regularisation (based on the input IRs), and the toolbox supports further investigation in this area, we have not yet developed a method that can be shown to match the results found by hand reliably. As parameters can be tweaked and auditioned rapidly, it would be necessary for any automated process to work both speedily and with a high degree of reliability.

Better understanding of the conditions in which phase may become relevant is also desirable, especially for phase-critical applications (e.g. ambisonic reproduction).

Relevant improvements to the correction procedure would be to examine more sophisticated smoothing algorithms (such as perceptually-motivated smoothing based on ERB or Mel scales), and to review the averaging procedure to better represent the commonalities between measurements (for instance, by identifying common poles between impulse responses).

Several video tutorials for the toolbox are planned, two of which are available at the time of writing, to maximise the accessibility of the tools. Work on a set of native measurement and correction applications and/or plugins is also underway, to accommodate users outside of the MaxMSP environment.

References

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

[2] RMC Pickup Co. [Online]. Available: http://www.rmcpickup.com/

[3] Ableton. (2013) Max for live essentials. [Online]. Available: https://www.ableton.com/en/packs/max-live-essentials/

[4] Audio Ease. (2012) Altiverb 7. [Online]. Available: http://www.audioease.com/Pages/Altiverb/

[5] 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.

[6] S. Berweck, "It worked yesterday: On (re-) performing electroacoustic music," Ph.D. dissertation, University of Huddersfield, 2012.

[7] E. Boyer, "Le contrôle du rayonnement appliqué à la reproduction sonore d'un instrument de musique," Master's thesis, Université du Maine, August 2009.

[8] S. Emmerson, *Living Electronic Music*. Ashgate Publishing, 2007.

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

[10] L. D. Fielder, "Analysis of traditional and reverberation-reducing methods of room equalization," *J. Audio Eng. Soc*, vol. 51, no. 1/2, pp. 3–261, 2003. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=12249

[11] Y. Haneda, S. Makino, and Y. Kaneda, "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.

[12] A. Harker and P. A. Tremblay, "The HISSTools Impulse Response Toolbox: Convolution for the masses," in *Proceedings of the International Computer Music Conference*. The International Computer Music Association, 2012, pp. 148–155.

[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] http://audio.lernvall.com/. (2008) Laconvolver. [Online]. Available: http://audio.lernvall.com/

[16] IK Multimedia. (2012) Arc system 2. [Online]. Available: http://www.ikmultimedia.com/products/arc/

[17] M. Karjalainen, T. Paatero, J. Mourjopoulos, and P. Hatziantoniou, "About room response equalization and dereverberation," in *Applications of Signal Processing to Audio and Acoustics, 2005. IEEE Workshop on*, oct. 2005, pp. 183 –186.

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

[19] MathWorks. (2013) MATLAB Signal Processing Toolbox R2013a. [Online]. Available: http://www.mathworks.co.uk/products/signal/

[20] P. D. Mourjopoulos, John N.; Hatziantoniou, "Real-time room equalization based on complex smoothing: Robustness results," in *Audio Engineering Society Convention 116*, 5 2004.

[21] 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.

[22] R. Penha. (2012) Spatium. [Online]. Available: http://spatium.ruipenha.pt/

[23] Real Sound Lab. (2011) CONEQ v3.2.0. [Online]. Available: http://www.realsoundlab.com/large/5448/

[24] H. H. Rutz. (2010) FScape v1.0.0. [Online]. Available: https://github.com/Sciss/FScape

[25] D. Sbragion. (2012) DRC v3.2.1. [Online]. Available: http://drc-fir.sourceforge.net/

[26] P. A. Tremblay and S. Mc Laughlin, "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.

[27] Waves. (2005) IR1 Convolution Reverb v2. [Online]. Available: http://www.waves.com/plugins/ir1-convolution-reverb

[28] R. Wilson, M. D. Capp, and J. R. Stuart, "The loudspeaker-room interface-controlling excitation of room modes," in *Audio Engineering Society Conference: 23rd International Conference: Signal Processing in Audio Recording and Reproduction*. Audio Engineering Society, 2003.