A NON-LINEAR TECHNIQUE FOR ROOM IMPULSE RESPONSE ESTIMATION

Tim Collins

Department of Electronic, Electrical and Computer Engineering
University of Birmingham, Edgbaston, Birmingham, UK
T.Collins@bham.ac.uk

ABSTRACT

Most techniques used to estimate the transfer function (or impulse response) of an acoustical space operate along similar principles. A known, broadband signal is transmitted at one point in the room whilst being simultaneously recorded at another. A matched-filter is then used to compress the energy in the transmission waveform in time, forming an approximate impulse response. Finally, equalisation filtering is used to remove any colouration and phase distortion caused by the non-uniform energy-spectrum of the transmission and/or the non-ideal response of the loudspeaker/microphone combination.
In this paper, the limitations of this conventional technique will be highlighted, especially when using low-cost equipment. An alternative, non-linear deconvolution technique is proposed which will be shown to give superior performance when using non-ideal equipment.

1. INTRODUCTION

The basic principles behind most techniques for measuring the impulse response of an acoustical space are very straightforward. A known excitation signal is transmitted from a loudspeaker at one point and is simultaneously recorded using a microphone at another. If the excitation signal is an impulse function then, under ideal conditions, the recording will be the impulse response of the room plus any ambient noise. In practice, in order to suppress the effects of noise, a more energetic (longer duration) signal would normally be used. Common examples are pseudorandom sequences such as MLS (maximum-length sequences) or IRS (inverse repeated sequences) and frequency swept sinusoids [1]. In any case, the receiver processing attempts to recover the impulse response of the room using a linear deconvolution filter (a band-limited inverse filter). If the energy in the excitation signal is uniformly distributed in frequency, the band-limited inverse filter is identical to the matched filter for the signal, so optimal noise suppression is achieved. Non-linearly frequency-swept sinusoids do not have a uniform energy spectrum and their inverse filters are not their matched filters. The inevitable compromise in the output signal-to-noise ratio must be balanced against the benefit of the suppression of interference caused by harmonic distortion (usually at the transmitter) [2].
Using ideal transmission and reception equipment, this technique should yield an estimate of the response of the acoustical space to a band-limited impulse function. In practice, the responses of the loudspeaker and microphone will not be perfectly flat and will, to an extent, colour the estimate. An estimate of the frequency response of the equipment made in an anechoic room can be used to derive a linear equalisation (or deconvolution) filter and invert any colouration. Although effective, this simple filtering has two notable drawbacks:

- The equalisation filter will amplify the parts of the spectrum where the signal strength is weakest, inevitably increasing the noise level.
- A linear deconvolution technique cannot estimate the frequency response of the room outside the bandwidth of the equipment.

The second point is particularly relevant when using low-cost equipment. Using a domestic hi-fi system and a budget microphone, the lower cut-off frequency can be as high as 200 Hz or more. There can also be several deep notches in the frequency response, particularly at the crossover regions between the different drive units. The most common solution to these drawbacks is to ensure that broadband equipment with as flat a frequency response as possible is used and to tolerate the bulkiness and expense that this implies.
In this paper, an alternative non-linear deconvolution technique will be described that is capable of estimating the frequency response over a broad bandwidth using possibly incomplete information from non-ideal equipment. It will be shown that small notches can be eliminated and the bandwidth of the estimate extended well beyond the limitations of the measuring equipment. A description of this deconvolution algorithm with experimental results will follow a short recap of conventional transfer function measurement techniques.

2. CONVENTIONAL TECHNIQUES

Regardless of the excitation signal chosen, the basic outline of most transfer function measurement systems runs along the lines of the diagram shown in figure 1. The excitation signal, $s(t)$, is inevitably convolved with the impulse response of the transmitting loudspeaker, $g_{TX}(t)$, before further convolution with the response of the channel under test, $h(t)$. The received signal $r(t)$ is the channel output after further modification by the receiver response, $g_{TX}(t)$. $r(t)$ can, therefore, be expressed as:

$$r(t) = s(t) \otimes g_{TX}(t) \otimes h(t) \otimes g_{RX}(t) \qquad (1)$$

This is, in fact, an approximation. Ambient noise added to the received signal has been omitted for clarity. Providing the excitation signal is sufficiently energetic, this is a valid approximation.

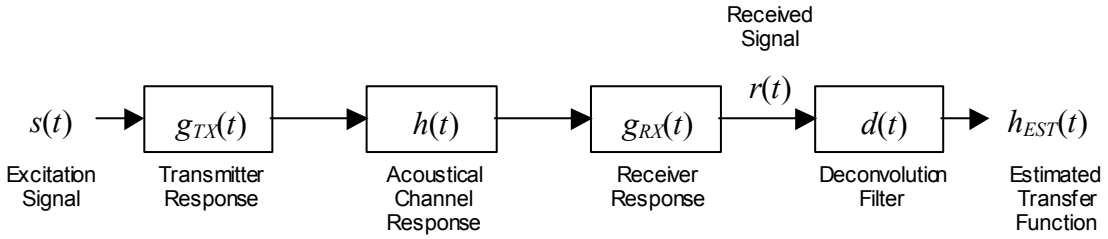

Figure 1. *Block diagram of a typical impulse response measurement system.*

From equation (1), it is clear that in order to obtain an estimate of the channel response alone, the combined effects of the excitation signal, the transmitter and the receiver must be removed from *r(t)* using deconvolution. The conventional method is to use a linear deconvolution filter, *d*(*t*), whose frequency response is, ideally, the inverse of the cascade of the excitation signal, transmitter and receiver responses:

$$d(t) \leftrightarrow D(\omega) = S^{-1}(\omega).G^{-1}(\omega) \tag{2}$$

where,

$$S(\omega) \leftrightarrow s(t)$$
$$G(\omega) \leftrightarrow g(t) = g_{TX}(t) \otimes g_{RX}(t)$$

So the estimated transfer function should be:

$$\begin{aligned} h_{EST}(t) &\leftrightarrow R(\omega).D(\omega) \\ &\leftrightarrow \left[S(\omega).G(\omega).H(\omega)\right]\left[S^{-1}(\omega).G^{-1}(\omega)\right] \\ &\leftrightarrow H(\omega) \end{aligned} \tag{3}$$

In practice, the excitation signal is limited to a finite frequency band. Outside this band, where $S(\omega)$ falls to zero, the actual inverse signal, $X^{-1}(\omega)$, would be infinite. The 'inverted' signal that is actually used for the deconvolution filter is limited to the same bandwidth of the transmission and is set to zero elsewhere.

3. LIMITATIONS OF INVERSE FILTERS

In order for the method described above to give an accurate estimate of the transfer function of the channel, it is essential that the bandwidth of the excitation signal equals (or exceeds) the bandwidth of the channel. In practice, it is only the audible band that is of interest and so a band-limited estimate of the channel response is made between around 20 Hz and 20 kHz. To obtain a good signal-to-noise ratio over the entire band, it is necessary that the bandwidth of the cascaded frequency response of $S(\omega)$, $G_{TX}(\omega)$ and $G_{RX}(\omega)$ exceeds (or at least equals) the limits of the audible band. This requirement is trivial in the case of $S(\omega)$, achievable for $G_{RX}(\omega)$ but can cause problems for $G_{TX}(\omega)$. If one wishes to make measurements using easily portable and/or inexpensive equipment, a lower cut-off frequency of 20 Hz can be difficult to achieve. The only options are to attempt to correct the transmitter response in the deconvolution filter or to accept a reduced band-limit. If the former option is adopted, the deconvolution filter will have a large gain at low frequencies and will increase the noise level accordingly.

This same problem can arise at the upper cut-off frequency of the system and at one or more frequencies in between. Dips and notches in the frequency response of low-cost loudspeakers are common and will demand extra gain in the same way, increasing the noise level undesirably.

4. NON-LINEAR DECONVOLUTION

Similar problems to those described above occur in many related areas. For example, a linear adaptive equaliser used to deconvolve a communications signal from a channel response will only perform well if the channel does not exhibit deep spectral nulls [3]. The solution in such cases is to use a non-linear equaliser or deconvolution algorithm. It should be stressed that the channel is still a linear system but the signal processing algorithms used for the deconvolution may not be.

There have been many non-linear deconvolution algorithms proposed in the past. The basic idea is to iteratively find a vector, $h_{EST}(t)$, that minimises the cost function:

$$E = \overline{\left[s(t) \otimes g(t) \otimes h_{EST}(t) \otimes s^*(-t) - r(t) \otimes s^*(-t)\right]^2} \tag{4}$$

This cost function is the mean-squared difference between a matched-filtered version of the received signal and an estimate of the same signal synthesised using *a-priori* knowledge of the excitation signal and the transmission/receiving equipment's combined impulse response as well as the current estimate of the channel response. Theoretically, deconvolution could be performed directly using the received signal. In practice, however, matched filtering is performed first for several reasons:

- Pulse compression will reduce the effective duration of the received signal to approximately the duration of the channel response. A shorter time-window of interest can therefore be extracted, reducing the dimensionality of the deconvolution problem.
- Ambient noise within the time-window of interest will be suppressed.
- If a logarithmically swept sinusoid is used for the excitation signal, interference caused by harmonic distortion will be separated from the window of interest [4].

To simplify the notation, equation (4) can be rewritten as:

$$E = \overline{\left[x(t) \otimes h_{EST}(t) - y(t)\right]^2} \tag{5}$$

where,

$$x(t) = s(t) \otimes s^*(-t) \otimes g(t)$$
$$y(t) = r(t) \otimes s^*(-t)$$

Assuming $g(t)$ has been measured in anechoic conditions, the only unknown function in equation (5) is the impulse response estimate, $h_{EST}(t)$.

In the early stages of this research, attempts to minimise equation (5) involved the use of a simple LMS algorithm similar to one that might be used in a decision feedback equaliser [3]. Although this did successfully converge and minimise the error function, it would not manage to convincingly estimate the channel response outside the bandwidth of the transmission signal. On reflection, it is not surprising that this happens. The training of most adaptive algorithms is based on the error vectors calculated in previous iterations. These error vectors will be coloured by the frequency response of the system, $G(\omega)$, encouraging the channel estimate to converge towards a solution occupying the same bandwidth.

A more fundamental problem is that the band-limiting effect of $G(\omega)$ means that there will be a potentially infinite set of plausible solutions for $h_{EST}(t)$, all of which minimise the cost function in (5) to the same degree. Outside the bandwidth of the measuring equipment, where $G(\omega)$ tends to zero, the spectrum of the estimated channel response is undefined and any value will produce an equally low error. Figure 2 illustrates an example of this problem generated by simulation. The estimated impulse response (top) was formed using a band-limited excitation signal (lower cut-off frequency, 100 Hz, upper cut-off frequency, 15 kHz) and is clearly not the same as the actual impulse response (bottom). If either function is convolved with $g(t)$, however, the results are indistinguishable.

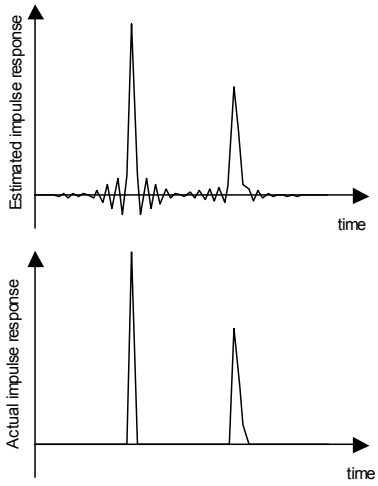

Figure 2. *Estimated impulse response using a band-limited limited signal.*

Although it is impossible to uniquely deduce the channel response outside the bandwidth of $x(t)$, it is proposed that one can, at least, derive a solution to (5) that is more plausible than the band-limited estimate produced by a linear deconvolution filter. We would like to not only minimise equation (5), but also to do so using the most likely estimated impulse response from the infinitude of possible candidates based on any additional *a-priori* information available. Simple modelling of room acoustics suggests that the expected impulse response will have a sparse nature, consisting of a number of discrete early reflections which will become more densely distributed as time progresses, but whose average amplitude should decay exponentially with time. We expect the impulse response to have a sparse nature (at least during the first few hundred milliseconds) rather than being a continuous function of time. It is hypothesised, therefore, that the most appropriate deconvolution algorithms will be those that attempt to produce a sparse solution.

4.1. Matching Pursuit

Several deconvolution/channel-estimation algorithms that have proved successful with sparse channels in the past have been based on the matching-pursuit principle [5, 6]. In fact, matching-pursuit can be used for continuous channels as well although the computational complexity increases in such situations. The basic principles of matching pursuit are very simple and can be most easily described by a pseudo-code algorithm:

Initialise $h_{EST}(t) = 0$

Loop:

By cross-correlating $x(t)$ with $y(t)$, find the time delay τ where $x(t-\tau)$ best matches $y(t)$. Also, find the amplitude, A, that minimises the mean-square-difference:

$$\int \left[y(t) - Ax(t-\tau) \right]^2 \, \mathrm{d}t$$

Subtract $Ax(t-\tau)$ from $y(t)$. Add $A\delta(t-\tau)$ to $h_{EST}(t)$.

Repeat until the residual signal in $y(t)$ is zero (or below a preset threshold)

At the end of the iterations, $h_{EST}(t)$ will contain a sampled estimate of the impulse response of the channel. With sparse channels, this simple algorithm will deduce the correct estimate in very few iterations. With arbitrary channels, convergence to the optimally sparse solution cannot be guaranteed theoretically [7]. In practice, however, it has been found to give a reliably close approximation.

4.2. Improving Matching Pursuit Performance

With simple simulated channels containing multiple discrete delayed impulse functions, matching pursuit rapidly converges to the exact solution. When the impulses are closer together than the effective duration of $x(t)$, each one may require several iterations to be correctly resolved, but convergence is still very reliable.

Taking a slightly more realistic channel model containing delayed, time-smeared impulses (simulating reflections from extended surfaces), the sparsity preserving nature of the simple matching pursuit algorithm can cause problems affecting the high-frequency parts of the channel estimate spectrum. This can be most easily demonstrated by example, as in figure 3.

When faced with a time-smeared impulse response, the matching pursuit algorithm's first assumption is to approximate that response with a single best-fit impulse function. The amplitude of this first impulse will always be an over-estimate as it contains energy from adjacent time-bins spread by the band-pass nature of the equipment response, $G(\omega)$. As a consequence, when the algorithm comes to estimate the amplitude of the adjacent samples (in iterations 3 and 4 in this case), these will tend to be under-estimates. Within a few iterations, the channel estimate evolves into a function that 'zig-zags' around the actual response. The error between the estimated and actual response occupies higher frequencies and may take many iterations to

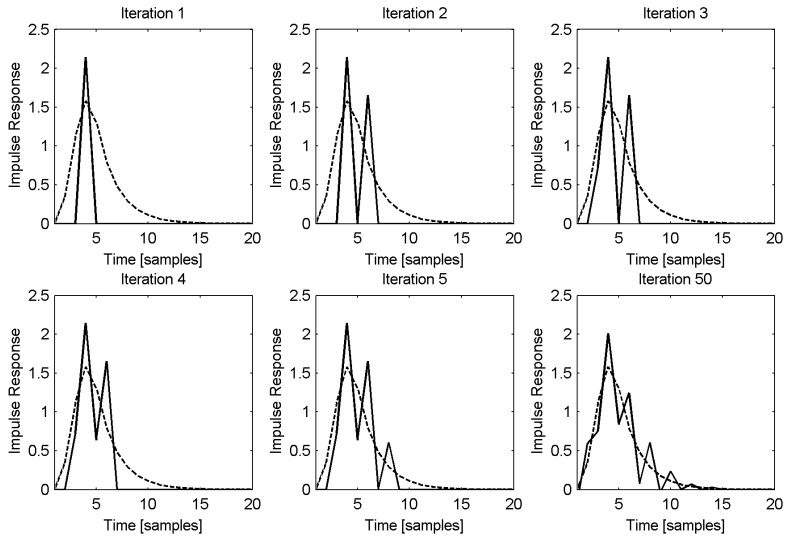

Figure 3. *Evolution of the estimated channel response (solid line) compared with the actual channel response (dashed line)*

correct (if ever) due to the limiting upper cut-off frequency of $G(\omega)$.

It should be noted that when using a band-limited excitation signal, some degree of error at high frequencies is inevitable. It is also clear, however, that the estimation formed in figure 3 is not a sensible estimate for a practical room response. In practice, we know that the reflections are expected to have a low-pass nature. At present, however, the deconvolution algorithm is unaware of this knowledge and assumes, instead, that the reflections will tend to occupy the entire spectrum equally.

The proposed solution to this problem is to increase the set of basis functions available to the matching pursuit algorithm. To begin with, the set of functions only contains just time-shifted replicas of the filtered excitation signal, $x(t)$. Equally valid additions to this basis are any function that can be formed by convolving a low-pass window function with $x(t)$. These additional functions are redundant in the sense that they can be formed by adding together a number of members of the original basis with appropriate weights. They may, however, provide much closer matches with time-spread channel responses than any member of the original basis. With appropriately chosen windows, this could prevent the problems illustrated in figure 3 as well as speeding up the convergence of the algorithm. These benefits are illustrated in figure 4 where the set of basis functions is enhanced with additional members derived by convolving $x(t)$ with cosine-squared windows of varying lengths.

In the first iteration in figure 4, the best match is found to be a version of $x(t)$ convolved with an eight-sample wide cosine-squared window. As a result, the convolved basis function is subtracted from $y(t)$ and the cosine-squared window is added to the current channel estimate. This is clearly a much closer first-order approximation to the actual impulse response than the single impulse function in figure 3. As a consequence, within only a few iterations the estimated impulse response evolves into a very close match to the actual response and after 50 iterations is near perfect (given the band-limited information available to the algorithm).

The effectiveness of this technique is highly dependent on the choice of low-pass filter windows used to form the additional basis functions. If, for example, the function was the same as the actual impulse response in figure 4, a perfect match would have been achieved on the first iteration. In experimental trials, it has been found that the larger the library of filter windows, the less iterations are required before convergence. In practice, however, the number of different windows should be limited. In particular, once the window length exceeds around 50 samples, the chances of coming across a good match become increasingly slim and the gains in convergence time become insignificant. Also, as more windows are used, the computational complexity increases. Convergence is indeed faster in terms of the number of iterations, but this gain is more than cancelled out by the increased processing time required per iteration.

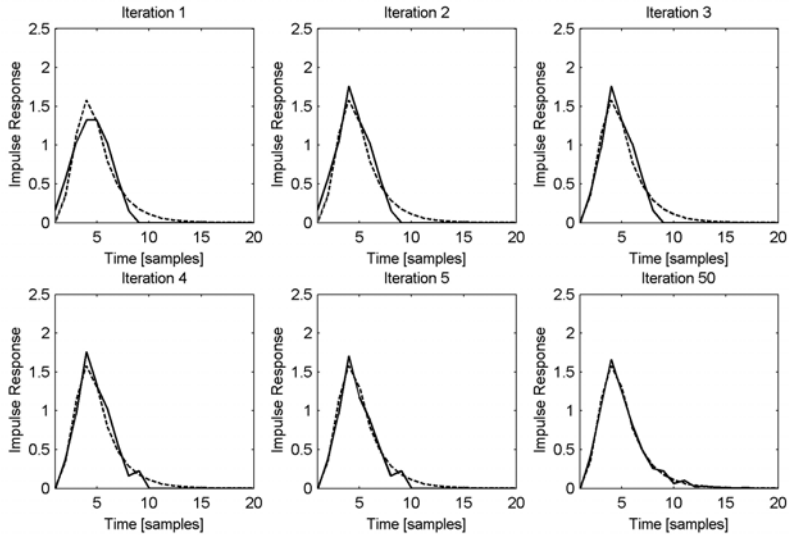

Figure 4. *Evolution of the estimated channel response (solid line) using the enhanced over-complete basis compared with the actual channel response (dashed line)*

In practical experiments, it has been found that a choice of three window functions is a good compromise. The windows chosen were a single impulse function (which must always be one of the windows) and two cosine-squared windows of four and eight samples duration.

5. EXPERIMENTAL RESULTS

In order to demonstrate the deconvolution algorithm, a low-cost test system was assembled using a domestic hi-fi amplifier and speaker with a cheap condenser microphone. The first stage was to calibrate the algorithm for this equipment by transmitting and recording the excitation signal in an anechoic room. Under the assumption that the room response equals a time-shifted impulse function, the received signal, $r(t)$, should be just the convolution of $s(t)$ and $g(t)$. Passing this signal through the matched filter for the transmission gives an estimate of $x(t)$.

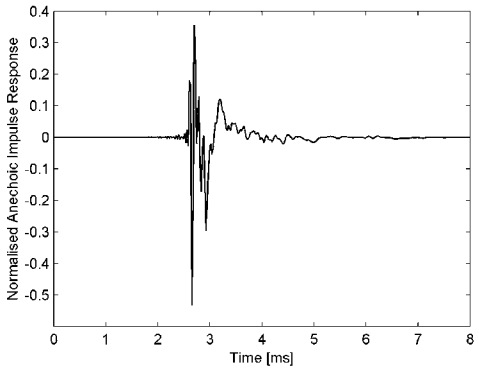

Figure 5. *Measured anechoic impulse response for the chirp transmission.*

Using a low-cost amplifier and speaker at high volume inevitably leads to a significant degree of distortion. In order to separate harmonic distortion components from the desired transmission signal, a log-sweep transmission was used [4]. Providing the duration of the sweep is greater than the reverberation time of the room, the matched filter for these signals can temporally separate energy from harmonic distortion in a way that is not possible with pseudo-noise type sequences. A transmission duration of 10 seconds was used over a frequency sweep range of 40 Hz to 20 kHz. Figure 5 shows the measured anechoic impulse response of the system.

For the sake of comparison with existing techniques, a linear deconvolution filter was derived from the inverse filter for the log-sweep [4] and incorporates the additional equalisation needed to correct the loudspeaker/microphone colouration and provide a uniform frequency response when presented with $x(t)$.

5.1. Simulations

When testing the algorithm with simple channels containing a few sparse impulses, the result is perfect recovery of the original impulse response. Although this compares favourably with the band-limited estimate given by a linear deconvolution algorithm, it is not an entirely fair test. Sparse channels favour the matching pursuit algorithm by its design and perfect deconvolution is almost trivial under such conditions.

For a more realistic, but controllable experiment, a synthetic channel response was obtained from a commercial reverberation effect software plug-in set for a reverberation time of 700 ms. By feeding an impulse to the effect, the actual channel response of the virtual room was obtained. Then, an estimate of $x(t)$ obtained using the test equipment in an anechoic room was fed into the effect giving a synthesised receiver matched filter output, $y(t)$.

Looking at the overall time-domain forms of the actual channel and the estimates formed by the linear deconvolution filter and

by the non-linear matching pursuit method, it is only possible to discern any difference under a close, detailed examination. Figure 6 shows a brief 2.5 ms window from the impulse responses. The difference between the non-linear estimation and the actual channel is clearly much smaller than the corresponding error for the linear estimation. This is due mainly to the absence of high frequency information outside the transmission bandwidth. This information is absent in the raw recording but is correctly inferred by the non-linear matching pursuit technique.

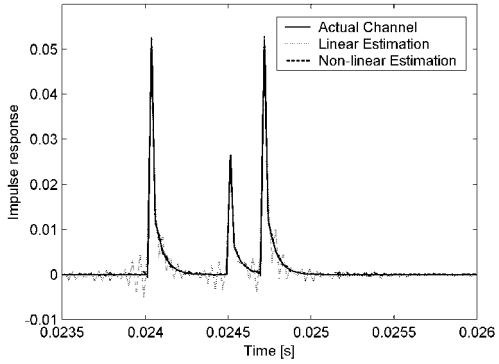

Figure 6. *Detailed section of the time-domain forms of the actual and estimated impulse responses.*

The errors illustrated in figure 6 will not normally be of great concern as they are above the high frequency hearing threshold for most people. At the opposite end of the spectrum, however, there is a similar deviation. This is most easily demonstrated by a frequency response plot as shown in figure 7. It is clear that whilst both methods work equally well above around 100 Hz, the linear estimation begins to deviate near the edge of the transmission bandwidth and falls off dramatically below 50 Hz due to the roll-on of the deconvolution filter. The non-linear matching pursuit method, by comparison, maintains a reasonably accurate inferred response all the way down to d.c.

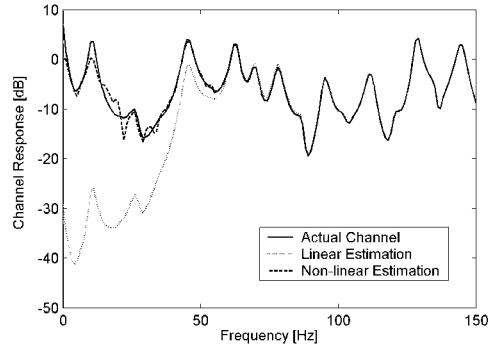

Figure 7. *Magnitudes of the low frequency responses of the synthetic channel and estimates formed by linear filtering and by matching pursuit.*

6. CONCLUSIONS

The transfer function estimation algorithm presented above can be thought of as an enhancement, rather than a replacement, for conventional techniques. As long as records of the excitation signal, the transfer function of the measuring equipment and the transfer function of any post-processing are available, this technique can be applied retrospectively to extend the bandwidth of existing measurements.

In simulation experiments, the algorithm has proved to be very effective. With sparse channels, complete recovery of the original transfer function can be achieved over the full audible band and beyond. With more densely distributed channels, the implicit ambiguity of the inverse problem means that 100% accuracy is impossible. Despite this fact, channel estimation outside the bandwidth of the excitation signal is still achieved with impressive fidelity.

In practical experiments in real acoustic spaces, the algorithm performs well again. Although it is impossible to quantify the improvements (since the actual channel response is not known), the subjective improvement compared with conventional techniques is obvious. Examples of the results from real spaces will be given at the conference presentation.

7. REFERENCES

[1] G. B. Stan, J. J. Embrechts & D. Archambeau, "Comparison of Different Impulse Response Measurement Techniques", *J. Audio Eng. Soc.*, vol. 50, pp. 249-262 (2002).

[2] S. Müller & P. Massarani, "Transfer-Function Measurement with Sweeps", *J. Audio Eng. Soc.*, vol. 49, pp. 443-471 (2001).

[3] J. G. Proakis, "Adaptive Equalization Techniques for Acoustic Telemetry Channels", *IEEE J. Oceanic Eng.*, vol. 16, pp. 21-31 (1991).

[4] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique", 108th AES Convention, *J. Audio Eng. Soc. (Abstracts)*, vol. 48, p. 350, preprint 5093.

[5] S. G. Mallat & Z. Zhang, "Matching Pursuits with Time-Frequency Dictionaries", *IEEE Trans. Sig. Proc.*, vol. 41, no. 12, pp. 3397-3415 (1993).

[6] S. F. Cotter & B. D. Rao, "Sparse Channel Estimation via Matching Pursuit with Application to Equalization", *IEEE Trans. Comms.*, vol. 50, no. 3, pp. 374-377 (2002).

[7] S. S. B. Chen, D. L. Donoho & M. A. Saunders, "Atomic decomposition by basis pursuit", *SIAM J. Sci. Comp.*, vol. 20, no. 1, pp. 33-61 (1998).