EVALUATING LOUDSPEAKER QUALITY AT LOW FREQUENCIES: OPTIMISATION OF A MUSIC-FOCUSSED MODULATION TRANSFER FUNCTION TECHNIQUE

LE Harris ISVR, University of Southampton, UK
KR Holland ISVR, University of Southampton, UK

1 INTRODUCTION

A PhD research project is underway to investigate a novel method of assessing the low frequency quality of loudspeakers during music reproduction; the work is focussed on professional applications such as mix monitoring in recording studios where replay becomes a matter of absolute accuracy rather than subjective preference. An earlier MSc project had produced results indicating that the method, based on the Modulation Transfer Function (MTF), might be a useful indicator of subjective quality. Several 'virtual' loudspeaker models with differing bass characteristics were simulated using digital signal processing; subjective ratings to evaluate the bass quality of these models were obtained through carefully controlled listening tests and compared to objective MTF data generated for each model.

The MTF is commonly used to calculate the Speech Transmission Index (STI) for evaluation of speech intelligibility inside listening spaces. The standardised test frequencies and bands used in this method have been shown to work well for the intended application i.e. speech assessment, but it was discovered that they led to unacceptable errors when applied to evaluation with musical stimuli at low frequencies. The following paper describes how the algorithm was optimised for this objective.

2 THE BASIS FOR INVESTIGATION

2.1 The MTF and Low Frequency Assessment

The Modulation Transfer Function was first adopted from optics into acoustics by Houtgast and Steeneken[1] in the early 1970s as a way to evaluate speech intelligibility inside a listening space. Along with several other authors, they reviewed and developed the technique (e.g.[2, 3]), and the MTF is now known within the audio industry as the basis for the *Speech Transmission Index* (STI), a standardised method of gauging intelligibility[4]. Despite its prevalent use in the assessment of speech quality degradation inside spaces such as sports arenas and performance halls, there has been little work to date on applying the MTF to evaluation of music reproduction.

Newell et al.[5-7] conducted an in-depth comparative study of popular recording studio monitors, analysing measurements in both the frequency and time domain. They found that those approaching the ‘ideal‘ (i.e. flat and extended) frequency response did not necessarily perform well in response to transients, exhibiting long low frequency decays in the time domain. This results in coloration of the sound as bass notes ring on after the high frequencies have died away. If a recording engineer is presented with this distorted spectrum as a result of poor transient accuracy in their studio monitors, it is likely that problems with the music going to tape will be overlooked, and/or incorrectly balanced mixes will be produced; these errors cannot be amended at a later stage. It is therefore critical that the pitch and relative levels between low frequency instruments such as bass drum and bass guitar can be assessed as accurately as possible. Though this may be widely acknowledged by recording engineers, the case for accurate timing and clear bass reproduction is not often documented in a convincing and straightforward way; however, two notable proponents are Newell and Colloms, having addressed these issues many times (for example, see [8-11]).

Though excellent transient behaviour may be critically important to accurate loudspeaker reproduction, the audio industry is still largely focussed on the frequency response as the primary indicator of sound quality. Further work by Holland et al.[12, 13] suggested that the MTF, incorporating both time and frequency behaviour, might be a better descriptor of low frequency performance. Harris[14, 15] pursued the idea and tried to establish a link between the MTF data and subjective quality scores for a range of loudspeakers with different low frequency characteristics. This formed the basis for the current work which attempts to optimise the MTF algorithm specifically for music so that it better describes the behaviour of a given loudspeaker and also correlates in some way to listeners' perception of the reproduction.

2.2 The Need for Modification

It was noted that the algorithm initially used to generate MTF scores had produced values somewhat lower than anticipated. After some investigation, it was discovered that a score of '1' was not being produced for a simulated 'perfect' system (a delta function, or perfect impulse, in the time domain, having a value of '1' at time $t=0$ and '0' at all others). This was evidence that the method of computation must be flawed, as in the absence of any external distortion (e.g. reflections due to room interaction), a perfect system must allow an input signal to pass through without any corruption. Hence, there must be no reduction in depth of a modulated signal and the MTF score must equal '1'. It was clear that the algorithm must be modified to produce a more accurate way of calculating the MTF in relation to low frequency musical stimuli.

3 EXPLORING THE PROBLEM

3.1 Possible Sources Of Error

The following factors were initially suspected of causing the reduced MTF scores:

a) **Distortion due to ‘brickwall’ filtering** of the frequency response function (FRF), introduced when selecting a band of frequencies for processing. It was proposed that windowing each section of the FRF (e.g. with a Hamming or Raised Cosine function) might ameliorate these effects if shown to be causing errors.

b) **Insufficient frequency resolution** of the FRF, leading to an inaccurate representation of the loudspeaker low frequency response and/ or slightly incorrect selection of the frequency band to be processed.

c) **Higher modulation frequencies being too large** in relation to the bandwidth of the FRF section under investigation: inspection of the individual MTF matrix scores indicated frequency dependent errors, with the performance gradually worsening as modulating frequency (*mf*) increased within a given band. This was particularly true for the very low frequencies where band size was small (bands were defined logarithmically in fractions of an octave).

Point a) was shown to have little effect on the overall MTF scores, despite experimentation with a number of different windowing functions. Regarding *point b)*, it was concluded that the resolution of 1.35Hz (*sampling frequency/N-point FFT = 44100/32768*) which had been used throughout was the best compromise between accuracy and computational load. Discrepancies in selecting the desired frequency bands were found to be small with virtually no impact on the resulting scores. Thus, it was concluded that any errors resulting from insufficient frequency resolution were negligible in comparison to the large deviations from expected results that were under investigation. Analysis of *point c)* proved to be more revealing; a significant rise in MTF scores was observed when the size of each frequency band was increased. This verified that the problem with the MTF algorithm (or at least the most significant one) was frequency dependent.

3.2 Existing Evidence

Following a review of MTF-related literature, some evidence was found to suggest that the previous low scores were not as irregular as they had at first seemed. Linkwitz[16], Rife[17] and the British Standard relating to speech intelligibility[4] all stated that due to the filtering of the system FRF, the modulation depth (and hence the MTF score) *should* decrease with increasing modulation frequency. Rife states that the effect *"is not an artefact of measurement error, but is the result of band-limiting the transmission channel*".

Despite this information, no account could be found of any attempt to correct for these filtering errors, though Linkwitz did allude to the necessity for some sort of post-processing normalisation. Rife simply stated that the individual MTF values (i.e. scores at each *mf-frequency band* combination) were not important; only the averages in each band and the subsequent STI score were significant, and errors due to filtering with the bands would be reduced during the averaging procedure. This did not entirely answer the question as to why, despite its widespread use, no one had attempted to find (or at least did not routinely apply) a numerically accurate method of correcting the MTF data.

It was concluded that the answer lay with the application in which the technique is primarily used: assessment of speech intelligibility. For this kind of work, the standardised band centre frequencies range from 125Hz to 8kHz with a maximum modulation frequency of 13Hz. A 13Hz modulation in an octave-wide band centred on 125Hz will produce a noticeable error, but this will be small and certainly negligible in an 8kHz band. However, at the low frequencies (below 125Hz) associated with the present investigation of music reproduction quality, modulation frequencies are comparable to the test bandwidth. In this case, the reduction in MTF score due to the *mf / bandwidth* ratio *is* significant and becomes a critical problem in the lowest bands. It was therefore necessary to either find an alternative algorithm, or to develop some method of 'filter error' correction with which to amend the MTF scores.

4 OPTIMISING THE ALGORITHM

4.1 Alternative Methods

The algorithm used up to this point had implemented the equation derived by Schroeder[18] in which only the impulse response of a (noise free, linear time-invariant) system is required to compute the MTF. Schroeder derived this expression from the method initially proposed by Houstgast and Steeneken[1, 2] whereby broadband noise is cosine-modulated at a range of individual frequencies; the reduction in modulation depth of this signal after passing through a system indicates how much degradation in quality has occurred between input and output. Conducted formally, this method would be extremely time consuming, requiring physical readings of the modulated noise reproduced through a transmission channel (e.g. a loudspeaker in a room). The analysis could be carried out reasonably quickly if the system's impulse response data was readily available but the process would still be somewhat cumbersome and much slower than the Schroeder method, primarily because it relies on measurement of the output waveform amplitude and therefore requires many averages to obtain a clear result.

Thus, the Schroeder technique is clearly a desirable method to use, being fast, simple to implement and requiring only one (albeit accurate and noise-free) measurement of the system response. However, the inherent errors described in *section 3.2* make it unsuitable for assessment at low frequencies unless some form of correction is applied to the resulting data. For this reason, an alternative approach was devised in which the MTF results would be free of such filtering errors because the system impulse response did not need to be band-limited. The three approaches to computing the MTF are summarised here:

i) Schroeder Expression

The Fourier Transform of the squared impulse response divided by its total energy:

$$m(f) = \frac{\int_0^\infty h^2(t) e^{-j2\pi ft} dt}{\int_0^\infty h^2(t) dt} \qquad (1)$$

Where *m(f)* is the modulation index, *h(t)* is the system impulse response, bandlimited to the frequency range of interest, and *f* is the modulation frequency.

ii) Formal MTF technique

Using cosine-modulated white noise with a band-limited system response:

- Generate **broadband** (white) noise
- Amplitude modulate it using a cosine function
- Convolve with a **bandlimited** impulse response of the system under test
- Inspect the envelope of the resulting signal to determine the modulation depth (compared to the original function)

iii) Alternative MTF technique

Using modulated band-limited noise with a full system response:

- Generate **bandlimited** white noise
- Amplitude modulate it using a cosine function
- Convolve with the **full** impulse response of the system under test
- Inspect the envelope of the resulting signal to determine the modulation depth (compared to the original function)

It was discovered through experimentation that methods *i* and *ii* produced virtually identical results i.e. it was verified that the Schroeder method is a valid substitute for the formal method, and being considerably more time consuming, method *ii* was subsequently discounted as a possible algorithm.

Method *iii* was also eventually rejected because although it avoids the filtering errors and *does* produce the theoretical MTF score of '1' at all frequencies for a perfect system, there are two notable drawbacks: firstly, it is time consuming because, as with method *ii,* many averages are required to produce a sufficiently smooth envelope from which the modulation depth can be measured. The second problem is the inherent nonlinearity due to band-limiting of the broadband (white) noise *before* it is modulated. Sum and difference frequencies are produced during modulation and the band-limited noise spectrum spreads and reduces in magnitude; widening of the band indicates that frequencies have been added i.e. nonlinear distortion has occurred.

Whilst a modulated noise signal also forms the basis for methods *i* and *ii*, it is broadband i.e. contains equal amounts of all frequencies; it is therefore impossible to add elements to the spectral content by modulation, only change their magnitude, which is a linear process. It became clear that method *iii* also had an intrinsic flaw, an 'intermodulation error', and due to the fundamental differences between algorithms *i* (and *ii*) and *iii* their results differed quite markedly.

On the basis of these findings, the Schroeder method was chosen for the updated Modulation Transfer Function algorithm.

4.2 Developing A Musical Modulation Frequency Set

The initial set of modulation frequencies had been taken from the work carried out by Holland et al.[7, 12, 13] and was similar to those applied in speech intelligibility tests; noise modulated at these frequencies is intended to resemble the spectrum of speech but not necessarily that of music. The highest value used was 13Hz and as previously discussed, large modulations in comparatively narrow bands were producing sizeable errors in the MTF data. It was proposed that a new group of modulation frequencies could be developed through analysis of musical signals which would more closely represent the typical input to a loudspeaker at low frequencies; this would in turn produce MTF data more relevant to musical reproduction quality. It was also hoped that the maximum modulation found in music might be lower than that of speech so that the *mf / bandwidth* error would be reduced.

4.2.1 Music Analysis

For a given musical extract the objective was not to find its frequency content, but rather the frequencies of any short term changes within it. 'Short term' in this case was provisionally deemed to be anything less than approximately 2 seconds, as it is unlikely that variations slower than this would perturb the loudspeaker response by any notable amount. Rather than pick an arbitrary lower modulation limit, an MTF analysis was conducted on 10 'real' loudspeaker systems in a single band centred on 63Hz (given that this should cover the fundamental frequencies of instruments such as bass guitar and kick drum). The band was modulated with a range of frequencies; it was proposed that the frequency at which the MTF scores became identical would be the one which would no longer disturb the loudspeaker response. 0.5Hz was chosen to be this frequency, showing approximately a 1% difference in MTF (overall mean) score between systems.

The following technique was used to identify the musical modulation frequencies because it was simple to execute and extremely accurate:

i) *Select an extract of adequate length* – A two-minute section was arbitrarily extracted from each piece of music: long enough to include progression of the music without imposing too high a computational load. Fixing the excerpts to an identical number of samples also made batch processing easier within MATLAB. All recordings were *.wav* files taken from CD (16-bit, 44.1kHz) and extracted from a single channel.

ii) *Low-pass filter to attenuate frequencies outside the range of interest* – The focus was upon how a loudspeaker would respond to low frequency excitation and thus, frequency content above 150Hz was removed from the stimuli to prevent the introduction of unwanted modulation components (additional frequencies were observed when full-bandwidth extracts were analysed as compared to the same extracts after filtering).

iii) *Apply Hilbert Transform and take the absolute values* – The Hilbert Transform is commonly used in modulation applications to detect the envelope of a signal. When applied to a purely real signal, the result is in the same domain and complex, with the real part being identical to the original input, and an imaginary part which is the same but shifted by 90^o; taking the absolute values of the Hilbert Transform then gives the amplitude envelope of the original signal[19].

iv) *Calculate PSD* – As the amplitude envelope of the signal tracks changes in the music, it follows that the Fourier transform of this gives the modulation frequencies contained within it. To calculate the power spectral density (PSD), successive segments of the signal are extracted, Fourier transformed, and the results added. This leads to an averaged spectrum showing the relative power of frequency components across the duration of the extract. Segments 10 seconds long were used in this case, giving a frequency resolution of 0.1Hz – a precision slightly higher than the lower bound of the range of interest.

4.2.2 Interpreting the Envelope Spectrum

It may have been possible to pick out a set of key modulation frequencies via inspection for one or two different extracts, but this was not a practical solution for the analysis of a wide range of music; an algorithm was required which could process a large number of extracts and automatically pick out the main peaks. A program was subsequently developed which recorded the top '*N*' (e.g. 20) frequencies occurring in each spectrum. This method worked well but was based only on the number of times each frequency appeared, irrespective of its magnitude. It may therefore have been possible that a frequency occurring very often but at a comparatively low level would appear to be more significant than one which occurred less often but was extremely prominent every time.

A different method was adopted; this was simply an addition of the envelope PSD from a large number of musical extracts. It was anticipated that any dominant frequencies common to a majority of the spectra would thus be prominent in a single averaged spectrum; their mean relative magnitude would also be preserved which might allow some form of weighting scheme to be developed at a later stage.

173 extracts were used for the final analysis. Preliminary trials on material within specific genres showed a marked difference in the prominent modulation frequencies and thus, care was taken to ensure that a wide range of styles was represented without particular bias towards any one type. The resulting averaged spectrum indicated that music in general contains number of dominant modulation frequencies, all occurring below 10Hz. *Figure 4.1* shows the averaged spectrum labelled with an '**x**' at the frequencies chosen for inclusion in the musical modulation frequency set (MmfS): 0.80, 1.10, 1.50, 1.75, 2.15, 2.50, 3.65, 4.35, 5.75 and 8.55Hz. Thus, the original MmfS of 20 frequencies ranging between 3.5 and 13Hz had been halved and the maximum modulation frequency reduced.

When the MTF algorithm had been finalised, including the updated MmfS, it was used to analyse the response of a group of loudspeaker (woofer) models. It was seen that virtually identical results were produced with an even further refined frequency set of 0.8, 1, 2, 4, 6 and 8Hz. As this meant faster computation and less data to manage, the smaller set was subsequently used. *Table 1* summarises the MmfS for the three instances considered in this study: *BSI (STI)* – British Standard on Speech Intelligibility Index [4], *Holland* – the preceding investigation into application of the MTF to bass reproduction in music, and *Harris* – the current project.

Application	No. of Mod. Freqs	Mod. Freq. Range (Hz)	Stimulus
BSI (STI)	14	0.63 – 12.5	Speech
Holland	20	3.5 – 13	Music (low frequencies)
Harris	6	0.8 – 8	

Table 1 Summary of MTF modulation frequencies in three applications

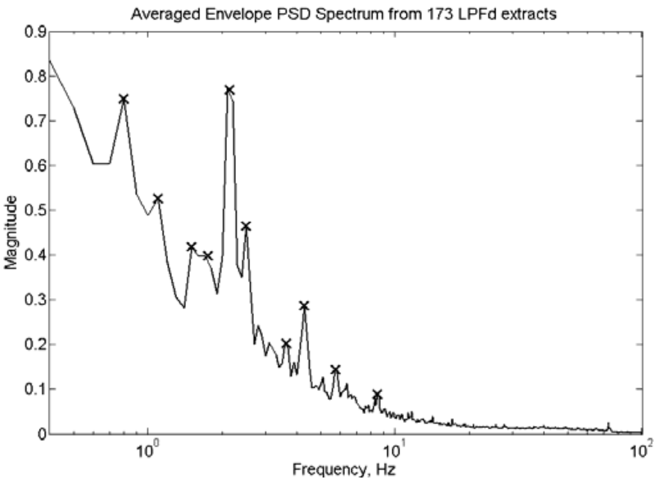


Figure 4.1 Music envelope power spectrum, averaged from 173 extracts. Peaks indicate prominent modulation frequencies in the music. Frequencies chosen for the algorithm modulation frequency set are marked with an 'x'

4.2.3 Modulation Frequency By Genre

To decrease processing time whilst trying to find a method of selecting the musical modulation frequency set, extracts were processed in small groups. On a few occasions, these extracts were all taken from the same genre due to convenience and it was noted that the averaged envelope spectrum was different for different types of music. A short investigation was subsequently carried out to observe the effect in more detail.

A number of musical extracts (in the same format as used for the main investigation) were grouped according to genre, classified as follows:

BLUES – 'Old' standard blues from the Chess label

RAP – American hiphop from the 1980's to present e.g. Dr Dre

EASY LISTENING – A selection of slow and laid back music

CHOPIN – Solo piano, composed by Chopin (Romantic era)

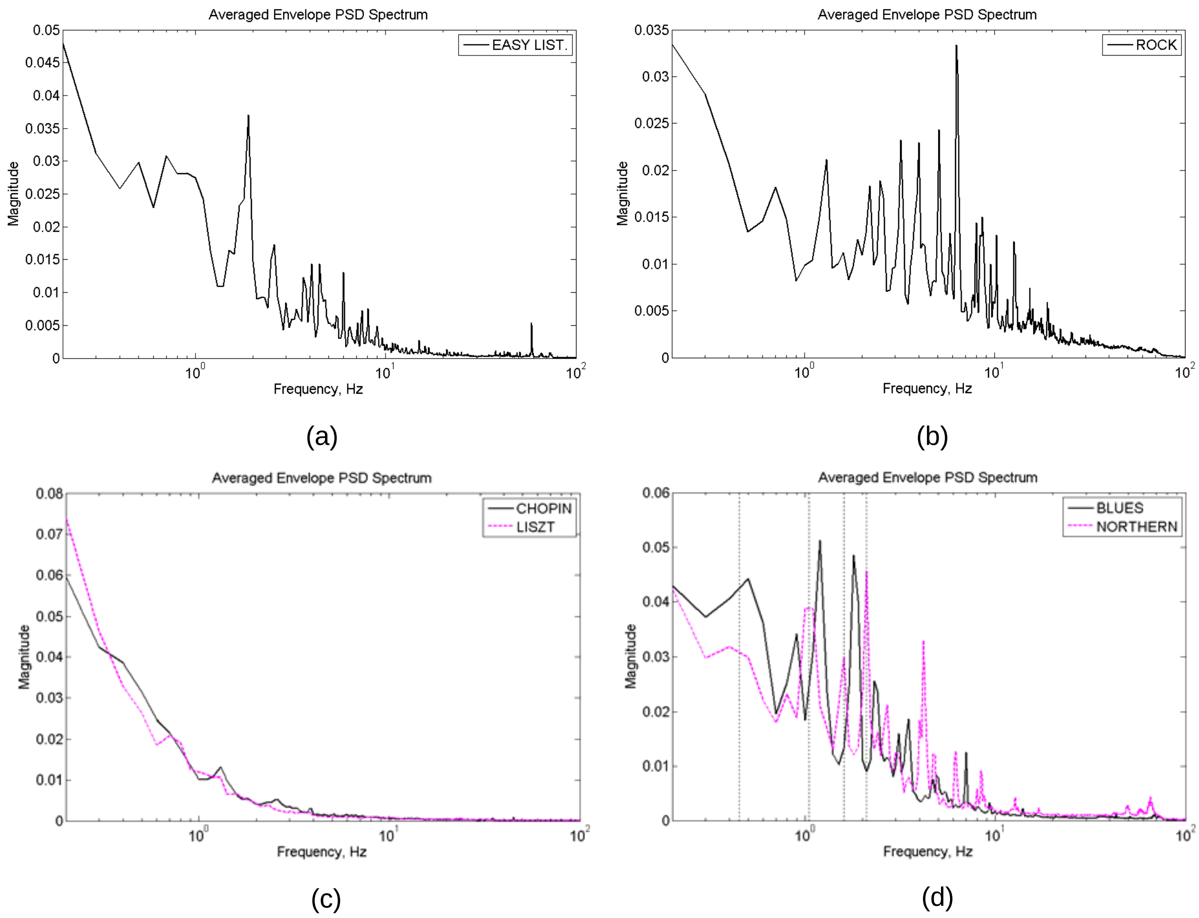
LISZT – Solo piano, composed by Liszt and Friedman (Romantic era)

URBAN – Modern 'urban' music, mainly American-style new R'n'B

NORTHERN SOUL – Underground American soul music from the 1960's

ROCK – Rock, tilted towards metal, from the 1980's to present.

Each group contained 12 extracts and was processed individually according to the PSD method described in *section 4.2.2*. The resulting spectra are shown in *figure 4.2(a)* to *(e)*. The approximate locations of the first four dominant modulation frequencies are marked with a vertical dashed line in *figures 4(d)* and *(e).*



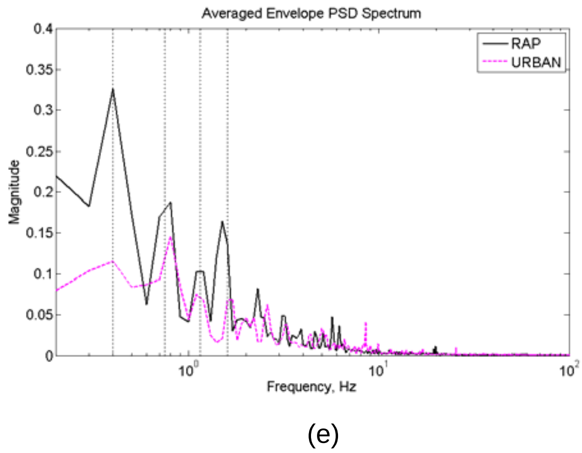


Figure 4.2 Averaged envelope power spectra for different musical genres:

(a) Easy listening
(b) Rock
(c) Solo piano (romantic era)
(d) Blues and Northern Soul
(e) Rap and Urban

The most obvious conclusion from observation of *figure 4.2* is that the envelope spectrum for each genre is markedly different, not only in the dominant frequencies, but also in their relative amplitude (note the varying y-axis scales; frequency on the x-axis is viewed over the same range for all of the plots). It is no surprise that the *Liszt* and *Chopin* spectra were very similar, essentially being the same genre; they were compared out of interest to see if there were any differences between the two composers' work. The lack of prominent frequency peaks reflects the fact that the extracts analysed had very little repetitive structure.

The related genres of *Blues* and *Northern Soul* did not exhibit the same spectral peaks but had a similar shape, appearing to be slightly offset from each other: one having a peak where the other had a dip or notch. This is in contrast the other related-genre case of *Rap* and *Urban*; these two *did* appear to share many peak frequencies, but with the *Rap* spectrum being on the whole much higher in magnitude. This featured the highest peak by far found in any of the genres, occurring at 0.4Hz, indicative of some element of the music repeating every 2.5 seconds. It is suspected that this may be due to the fairly regular meter of this type of music which is driven by a heavy kick drum at a regular pace, sustaining a beat over which the speech-like vocals can be performed.

The implications of these genre-related differences in modulation frequency may be significant for the proposed low-frequency MTF application, though they have not yet been investigated in any more depth than described here. It is reasonable to postulate that if an engineer or producer is trying to assess a pair of loudspeakers on which one type of music will primarily be monitored (not unreasonable, particularly on the urban or dance scenes), it may be more revealing to exchange the generic modulation frequency set for one tailored specifically to that genre.

4.3 Normalisation and Minor Tweaks

The MTF results could be influenced by several parameters within the algorithm, other than which modulation frequencies were used; the simulated replay level required to implement a noise-floor correction of the scores was retained (80dB SPL), and the band centre frequencies (18Hz to 80Hz spaced at 1/6th octave intervals) remained the same as those which had previously been used because the range of interest had not changed. Choosing the width of each frequency band and a noise modulating function required some experimentation in order to make a decision as to which settings might give an improved measurement.

4.3.1 Modulating Function

MTF results were compared for three different noise - modulating functions: Raised cosine, squared cosine, and absolute-value cosine. There appeared to be little difference between the MTF values generated by the three functions and so the raised cosine was chosen for the final algorithm because its use as a modulating function is established. It also produces the wanted frequency without the need for any multiplication factor (the others produce twice the frequency as they are changed from a waveform with positive and negative values to one which is positive only).

4.3.2 Frequency Band Width

Although 1 octave-wide bands are specified in the BSI standard relating to the STI and were used for the original MTF analysis, various other bandwidths were investigated to see how the results varied and if they offered any advantage. The effects of changing this parameter were most easily seen when analysing a 'perfect' system with the Schroeder algorithm whereby increasing the bandwidth gave higher MTF scores, and conversely, narrowing the bands led to larger errors.

Most of the bandwidths tried were logarithmic, such as 1/3rd octave i.e. the upper and lower limits were defined as fractions of an octave based on the specified centre frequency; the range of the bands therefore varied quite considerably. As it was believed that the Schroeder errors were largely dependent on the size of the modulating frequency relative to the bandwidth, the effects of using fixed-width linear bands were also investigated. Due to the lowest centre frequency being at 18Hz, the widest band possible was approximately 30Hz (18±15Hz). Although this linear-band approach was somewhat unconventional, it provided clear evidence that the error was a function of the *mf / bandwidth* ratio: for a perfect system it could be seen that the scores for a given *mf* were identical in each band but decreased with increasing *mf* i.e. the width of every band was fixed, so as the modulation frequency grew in relation to this, so did the errors in the resulting MTF scores.

Following the comparison of many sets of MTF results, it was decided that 1/3rd octave bands were unnecessarily narrow for use in the low frequency range of interest; even when using the conventional 1 octave bands, the MTF scores for a perfect system were quite low. It was eventually decided that a bandwidth resulting in an error less than 5% overall (i.e. a mean MTF of 0.95 or greater, before addition of a noise floor) should be employed; 2.667-octave bands were found through trial and error to be the minimum bandwidth that would meet this criteria.

4.3.3 The Band-limited Impulse Response

As described in *section 4.1*, a band-limited version of the system's impulse response (IR) was needed to compute each MTF. The filtering was performed in the frequency domain as it proved impossible to generate suitable digital filters that could perform this task in the time domain (the cut-off frequencies required were very low in comparison to the sampling frequency and the filters were therefore unstable). After applying the Inverse Fourier transform, the resulting I.R. was acausal i.e. contained information before time *t=0*.

If this full IR was used directly for computation, the MTF matrices had a 'stripy' appearance (an example is given in *figure 4.3*); it seemed that the data was being 'scrambled' due to the lack of causality. Two solutions were considered:

- Truncate the IR so that data in the second half would be discarded. This was the method used in the preceding MTF investigations.

- Apply a phase shift in the frequency domain to delay the signal and thus impose causality. In this case the entire IR would be used for the MTF calculation.

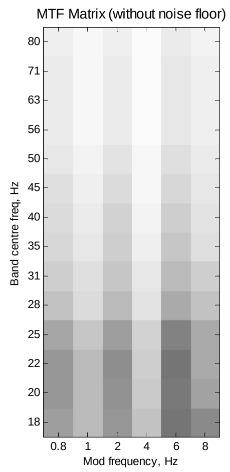


Figure. 4.3 Stripy behaviour seen in the MTF matrix when using the entire length of the acausal band-limited impulse response (no added noise floor)

When the MTF scores from each method (i.e. the acausal half-IR and the delayed, full-IR method) were compared, little difference could be seen; this was true when analysing real loudspeaker data as well as experimental model responses. Qualitatively the two techniques appeared to be the same, with the only difference being that the mean MTF scores from the full-IR method consistently came out with a lower value. This 'offset' was not regarded as a serious problem and thus, the half-IR algorithm was chosen as it could perform the MTF analysis in less than half the time taken by the phase-shifting equivalent.

4.3.4 Reassessing the MAF-Correction Figures

As in the preceding work, a noise-floor was added to the MTF results through a set of correction figures based on the Minimum Audible Field (MAF) or threshold of hearing at the low frequencies of interest. for the updated algorithm, the correction values were altered slightly to agree with the most recent British standards relating to equal loudness contours and the thresholds of human hearing[20, 21] which combine the results from a number of studies. The correction figures were estimated from the graphs by assuming a reproduction level of 80dBSPL; for the centre frequency of each band, the corresponding correction level was determined by estimating how far below the 80dB line the hearing threshold curve was at that point. If the curve lay above the line, the correction figure was zero. These MAF levels were used to modify the MTF scores using the following equation:

$$m_c(f) = m(f)\frac{1}{1+10^{\frac{L_N - L_S}{10}}} \qquad (2)$$

Where $m(f)$ and $m_c(f)$ are the uncorrected and corrected MTF scores respectively, L_N is the MAF correction value at the centre frequency of a specified band and L_S is the level of the loudspeaker frequency response in the same band, assuming a mean passband sound pressure level which is the same as that used to obtain the MAF levels (80dB in this case).

4.3.5 Normalisation of the Schroeder MTFs

Though the Schroeder method of MTF computation was believed to be the best of the three options considered, it still suffered from the inherent 'bandwidth error'. A method of correction was eventually developed, involving normalisation by 'perfect system' MTF scores. This process is crucial to the technique and is arguably the most important step in the development of the MTF algorithm for the low frequency applications of interest. The problem and the correction method is explained here in more detail:

In the Schroeder method, a modulated noise signal is applied to a system within a specified frequency band. When the modulation frequency, *mf,* is small relative to the bandwidth, the signal can be fully modulated and an accurate result for the reduction in modulation depth can be obtained. As the *mf* increases and becomes comparable to the size of the band it is modulating, there is a reduction in the maximum possible depth of modulation. Errors are therefore introduced because the input signal is not fully modulated to begin with, and the computed depth of the output progressively approaches zero i.e. total distortion; such results were always seen in the non-normalised Schroeder MTF matrices at locations corresponding to these conditions (e.g. band with centre frequency = 18Hz, cosine modulated by frequency *mf* = 8Hz). Thus, even a theoretically perfect system will show imperfections at certain frequencies in its MTF matrix because the desired modulations cannot all be correctly applied. Ideally, the same (full) modulation would be achieved in each band so that the bandwidth-dependent errors could be avoided.

In theory then, if the MTF matrix for a perfect system is obtained, any deviations from a score of ‘1’ must be due to these filtering or band-limiting errors (assuming that other inaccuracies such as those described in *section 3.1* are negligible in comparison). The perfect-system MTF matrix therefore becomes a set of correction values with which to amend the scores from any other system analysed using the same parameters (identical frequency bands, modulating frequencies etc). Normalisation by the perfect system data effectively restores full modulation depth in each of the bands so that the correct value can be computed; any deviations between modulation frequencies in different bands must then be attributed to the behaviour of the system response rather than the assessment technique.

Figure 4.4 illustrates the normalisation process; *(a)* and *(b)* are the MTF matrices (shown in 'intensity image' form[15]) for a loudspeaker and perfect system, both MAF corrected (i.e. with noise floor added). *Figure 4.4(c)* is the matrix after normalisation i.e. after the matrix in *figure 4.4(a)* has been divided by that in *figure 4.4(b)*. Note that the corrected mean MTF score is higher than that of the uncorrected system, but lower than for the perfect case.

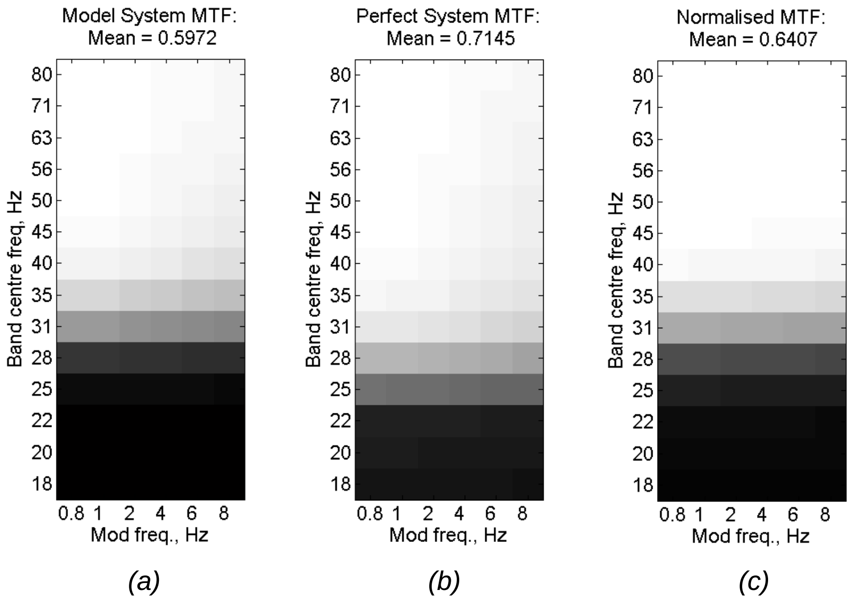


Figure 4.4 MTF matrices for a model system (a) *and the corrected result* (c) *after correction, i.e. normalisation by a perfect system* (b)

5 IMPLEMENTING THE OPTIMISED ALGORITHM

The original MTF algorithm had thus been revised and was believed to be an optimal, or at least more correct, way to assess a loudspeaker's ability to reproduce music accurately at low frequencies whilst also correlating well with subjective judgements of quality. The updated parameters are summarised below:

- Values generated using the Schroeder method would be normalised by a set of equivalent scores from analysis of a simulated perfect system.

- The modulating function would be a raised cosine at 6 successive frequencies ranging from 0.8 to 8Hz.

- 14 bands would be modulated, spaced 1/6th octaves apart with centre frequencies between 18 and 80Hz. Each band would be 2 and 2/3rd octaves wide.

- Band-limiting of the system impulse response would be performed in the frequency domain and only the first half of the data would be used for the computations after applying the inverse Fourier transform (impulse response truncated).

- A form of noise floor would be added to the MTF data by applying a series of correction figures derived from the threshold curves describing the limits of human hearing at low frequencies, assuming a sound reproduction level of 80dB SPL.

This algorithm will now be used as part of another subjective/ objective data comparison following a series of ‘ABX’ listening tests with a different set of simulated loudspeaker systems. Data from experiments of this type will be simple to process and absolute as listeners must give a simple ‘one or the other’ answer. Though this will not allow direct examination of the relationship between MTF scores and listener quality judgements, it will be possible to produce a ranking order for both types of data i.e. a given system may be shown to perform better than another, either through having a higher overall MTF score or by being judged as sounding closer to a designated reference system (but it will not be possible to say *how much* better). If it is shown through this kind of experimentation that the listener and MTF rankings are identical, it would be a strong indication that the choice of MTF algorithm was an appropriate one. Further tests can then be planned in order to try and establish a link between the objective MTF scores and subjective listener ratings for bass quality.

6 SUMMARY

It was known that the algorithm used to compute the modulation transfer function in previous studies contained an error. Thus, any further investigations using the MTF could not be carried out until this problem had been corrected or an alternative method devised.

Based on the findings from detailed and comprehensive investigation, a modified algorithm was produced which will be used in the next stage of research into the quality of loudspeakers reproducing music at low frequencies. This was based on information published in the literature, interpretation of investigative data and comparison with the results from subjective listening tests gathered during an earlier project.

It is anticipated that the efficacy of the optimised algorithm will be verified through comparison of data from further listening tests and MTF analysis of a wide range of different loudspeaker systems. However, if the proposed method gives little or no correlation between the subjective analysis of bass reproduction quality and the corresponding MTF data, it may be necessary to reassess the choice of algorithm and experiment with a number of others in the hope of finding a better match.

7 REFERENCES

1. Houtgast, T. and H.J.M. Steeneken, *The Modulation Transfer Function in Room Acoustics as a Predictor of Speech Intelligibility.* Acustica, 1973. **28**(1): p. 66-73.
2. Steeneken, H.J.M. and T. Houtgast, *A Physical Method for Measuring Speech-Transmission Quality.* Journal of the Acoustical Society of America, 1980. **67**(1): p. 318-26.
3. Houtgast, T. and H.J.M. Steeneken, *A Review of the MTF Concept in Room Acoustics and Its Use for Estimating Speech Intelligibility in Auditoria.* Journal of the Acoustical Society of America, 1985. **77**(3): p. 1069-77.
4. *BS EN 60268-16: Sound System Equipment – Part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index*. 2003, British Standards Institute.
5. Newell, P.R., K.R. Holland, and J.P. Newell, *The Yamaha NS10M: Twenty Years a Reference Monitor. Why?* Proceedings of the Institute of Acoustics, 2001. **23**(8).
6. Newell, P.R., K.R. Holland, and P. Mapp, *The Perception of the Reception of a Deception.* Proceedings of the Institute of Acoustics, 2002. **24**(8).
7. Holland, K., P. Newell, and P. Mapp, *Steady State and Transient Loudspeaker Frequency Responses.* Proceedings of the Institute of Acoustics, 2003. **25**(8).
8. Colloms, M., *High Performance Loudspeakers* 5th ed. 1997 John Wiley and Sons.
9. Colloms, M., *Basso Profundo*, in *Stereophile.* December 1991. *Available online: http://stereophile.com/reference/39/*. (Accessed 12/09/2007).
10. Newell, P., *The Great Low-Frequency Deception*, in *Resolution*. May/June 2003. p. 48-51.
11. Newell, P., *Audiophile Vs Professional Monitoring Priorities*, in *HIFICRITIC*. Issue 7, 2008. p. 11-14.
12. Holland, K., P. Newell, and P. Mapp, *Modulation Depth as a Measure of Loudspeaker Low-Frequency Performance.* Proceedings of the Institute of Acoustics, 2004. **26**(8).
13. Holland, K., et al., *Excess Phase Effects and Modulation Transfer Function Degradation in Relation to Loudspeakers and Rooms Intended for the Quality Control Monitoring of Music.* Proceedings of the Institute of Acoustics, 2005. **27**(8).

14. Harris, L.E., *Can the Modulation Transfer Function Be Used to Predict the Low Frequency Reproduction Quality of Musical Signals?* MSc Dissertation, 2007, University of Southampton.
15. Harris, L.E., K.R. Holland, and P.R. Newell, *Subjective Assessment of the Modulation Transfer Function as a Means for Quantifying Low-Freqeuncy Sound Quality.* Proceedings of the Institute of Acoustics, 2006. **28**(8): p. 195-203.
16. Linkwitz, S. *Investigation of Sound Quality Differences between Monopolar and Dipolar Woofers in Small Rooms*. in *AES 105th Convention*. 1998. San Francisco, CA.
17. Rife, D.D., *Modulation Transfer Function Measurement with Maximum-Length Sequences.* Journal of the Audio Engineering Society, 1992. **40**(10): p. 779-90.
18. Schroeder, M.R., *Modulation Transfer Functions: Definition and Measurement.* Acustica, 1981. **49**: p. 179-182.
19. Randall, R.B., *Frequency Analysis*. 1987, Bruel & Kjaer. p. 62-70.
20. *BS ISO 226:2003 Acoustics — Normal Equal-Loudness-Level Contours*. 2003, British Standards Institute.
21. *BS EN ISO 389-7: Acoustics — Reference Zero for the Calibration of Audiometric Equipment — Part 7: Reference Threshold of Hearing under Free-Field and Diffuse-Field Listening Conditions*. 2005, British Standards Institute.