oticon

The AUDMOD auditory model:

A critique and revisions

Report No: 43-8-5

Date: September 15, 1994

Author: Graham Naylor

OTICON A/S
RESEARCH UNIT "ERIKSHOLM"

Kongevejen 243
DK-3070 Snekkersten
DENMARK

Contents

Introduction

AUDMOD is a software implementation of an auditory model including the effects of sensori-neural hearing loss, developed by LBN [Nielsen (1993)]. AUDDISP is another program (also written by LBN), providing an interface 'shell' within which AUDMOD may conveniently be run. AUDMOD is a DOS program and AUDDISP is a Windows program.

This report covers work undertaken to establish appropriate choices of analysis parameters when using AUDMOD and to uncover limitations and errors in its implementation. In addition, a few errors and practical points regarding the use of AUDDISP have been investigated. During the course of this study, revisions have been made to both AUDMOD and AUDDISP.

The work described here was undertaken as a preliminary to the application of AUDMOD in the BINFIT project. This has naturally influenced the course of the work, especially in the disproportionate amount of effort spent on relatively high-frequency effects above the bandwidth of most current hearing aids.

The first section contains notes concerning the practical use of AUDDISP and revisions made to it. Section 2 uses analyses of chosen signals to illustrate potential pitfalls when choosing analysis parameters for AUDMOD, to draw conclusions regarding appropriate choices of parameters, and to expose errors and demonstrate their correction. Section 3 discusses shortcomings in AUDMOD regarding the conversion from excitation pattern to specific loudness, concluding that the conversion seems to be untrustworthy for high levels and large hearing losses. Suggestions are made as to how the believability of this transformation might be improved. In the final section, all the significant revisions made to AUDMOD are listed.

Where revisions to the software are described, they are accompanied by a vertical bar in the margin, as shown here.

1 AUDDISP: Interface, operation and revisions

This section contains notes regarding the AUDDISP user interface. It is not intended to serve as a user manual. Some (though not necessarily all) aspects of operation where care is required are mentioned, and revisions to AUDDISP are listed.

1.1 Interface and operation

1.1.1 .CSV output file format

This is the only output format which can be displayed as a 'spectrogram' in AUDDISP. When using this format, the data is output in semicolon-separated form. The first line contains various header information. Thereafter, each line contains specific loudness values for one frame.

The first line contains band numbers (0-N, where N will by default be 32), *overlaid* in the first seven cells with header information. For example, after reading into a spreadsheet:

Cell	A	B	C	D	E	F	G
1	1.41	20161	256	128	1	3	35
=	AUDMOD version	Sample rate	FFT length	FFT overlap (samples)	Spectra to average (0=all)	E-number of first band	E-number of last band

This format is a minimal one, best for importing data into other programs. There are no band frequencies and no total loudness. Note that 'Band number' is not the same as ERB index, since it always starts at 0.

1.1.2 .TXT output file format

This format is better suited to 'detective work', and is correspondingly expansive (see table 1.1). The data is output in tab-separated format, and after import into a spreadsheet it has an appearance as shown below: Lines 1 and 2 contain header information about the files for input and output and the analysis parameters. Lines 5, 6 and 7 specifiy band numbers in the analysis and their corresponding E-numbers and centre frequencies. Following are the outputs of the ERB filters (level in each ERB) and Roex filters (level at output of each Roex filter) for inputs at 0 dB SPL free field and at hearing threshold. The Roex filter output for input at UCL is also given, although it is not used in the model at present. After these calibration values, the outputs of the ERB and Roex filters, and the specific loudness values, are provided, for each N frames, where N is the number of spectra to average. At the end of each row of Roex outputs or specific loudnesses, the total excitation and total loudness respectively are given.

AUDMOD Rev:	Date:	Parameter file:	Signal file:	Sample Rate:	Frame Size:	Overlap:	Averaging:	E_Beg:	E_End:
1.41	May 20 1994	sp85a.AUD	c:\signals\smallkem\eq3b\wn180p.TIM	25000	1024	512	-1	11	35
Channel no:	0	1	2	3	4	5	6	7	8
Ec (E):	11	11.25	11.5	11.75	12	12.25	12.5	12.75	13
fc (Hz):	520.21	540.66	561.68	583.27	605.45	628.24	651.65	675.7	700.4
0 dB SPL:									
ERB (dB SPL):	20	20	20	20	20	20	20	20	20
Roex (dB SPL):	3.82	4.79	5.34	5.67	5.89	6.05	6.19	6.31	6.41
HTL:									
ERB (dB SPL):	53.81	53.81	53.91	54.02	54.12	55.4	55.47	55.52	55.56
Roex (dB SPL):	52.87	53.89	54.5	54.9	55.2	55.43	55.62	55.8	55.96
--UCL--									
Roex (dB SPL):	154.12	154.96	155.35	155.49	155.49	155.41	155.28	155.11	154.92
Frame:	80								
ERB (dB SPL):	86.62	87.72	88.55	89.47	91.29	92.03	92.58	93.07	93.52
Roex (dB SPL):	81.61	83.25	84.58	85.62	86.47	87.3	88.12	88.92	89.74
N' (sone/Bark):	1.21	1.27	1.33	1.4	1.46	1.52	1.59	1.67	1.74

Table 1.1: *Format of .TXT output files.*

1.1.3 Spectra to average

When 'Average ... spectra at a time' is set to 'All', AUDDISP interprets it as meaning a single frame of output. When it is set to something else, e.g. 4, the result is 4 times fewer output frames than input frames, with the time per output frame in the AUDDISP display not being expanded to compensate. Hence the output data appears to refer to a shorter time interval.

1.1.4 Display of filenames

The filenames displayed at the bottom of the graphics screen do not always correspond in an intuitive way to the files one imagines oneself to be working with. They are the names of the files displayed, which are not necessarily those currently being used.

1.1.5 Entry of numerical values

When the user types numbers into menu fields, no check is carried out to ensure that the resulting string of text actually represents a number. Making a typing error here can lead to a program crash, whereby one is thrown right out of the program. Quite apart from this, there is no check concerning the allowable range of values.

1.1.6 RMS values of signals

The notional SPL calibration of signals input to AUDMOD is defined by relating a defined SPL to a defined RMS value of the numerical signal data. When a series of signals have been recorded with the same physical calibration, the SPL-to-RMS relation will be the same for all of them, in which case the same definition should be used for all the signals when input to AUDMOD. The appropriate RMS value may be found either by running LBN's DOS program 'rms' on a signal file, or by using the 'Open signal file' function in AUDDISP. This calculates the RMS value, presents it for the user, and (unless instructed not to) pastes the value into the relevant field of a subsequent 'New parameter file'.

If one is carrying out identical processing on a series of signal files by simply changing the input and output file names (using the Modify Parameters function), any revisions to the RMS value must be made manually.

1.1.7 Opening of signal files

The 'Open signal file' function under the File menu is tediously slow (due to limitations of the BASIC language), and need only be used when the RMS value of a signal is unknown (and required) or when the signal is to be displayed above the 'spectrogram'.

1.2 Revisions to AUDDISP

Default analysis parameters

The default FFT size in AUDDISP has been revised from 256 to 1024. For reasons, see section 2.1.3. Otherwise the default parameters appear satisfactory.

Parameter file and audiogram

AUDDISP has been revised to remove some minor but dangerous bugs relating to the treatment of the audiogram when reading a previously-defined parameter file (.AUD), and when creating a new default set of parameters.

Choice of number of channels

AUDDISP has been revised, so that instead of specifying the number of channels (Roex filters) to be used, the user now specifies their density on the E-scale. The default is 1 channel per ERB.

Choice of recording coupler

In addition to the couplers available under LBN's implementation, a 'Diffuse field' 'coupler' has been added. For reasons, see section 2.1.6.

Choice of fast or slow initialization

In LBN's implementation of AUDMOD, the initialization routine is fast, but strictly speaking incorrect. The errors arising from this shortcut are discussed in section 2.1.4. The errors will often (but not always) be acceptable. Therefore an option has been added to AUDMOD and AUDDISP allowing the choice of a fast or a correct initialization. The default is a fast initialization.

Spectra to average

Given an input signal n FFT frames long, input spectra may be averaged in blocks of 1 to n before being sent to the auditory model. AUDDISP's interface has been revised to allow the choice of 'All frames except the last one', in addition to the previous 'All frames' or a specified number. For reasons, see section 2.1.2.

2 Signal analysis with AUDMOD

This section describes the results of analyses made with the purpose of studying the effect of variations in the available parameters. Conclusions are drawn regarding advisable choices of parameters and AUDMOD's general trustworthiness.

2.0.1 Time-varying signals

AUDMOD does not model any time-dependent auditory effects. Therefore it is strictly speaking only applicable to steady-state signals. Nevertheless it is constructed in such a way that both input and output are regarded as functions of time. AUDMOD has previously been used to model the auditory response to strongly time-varying signals (speech and music) [Nielsen 1993], and no doubt will be again in the future. Such usage is not dangerous unless the kind of study being undertaken relates to auditory processes and signals in which temporal masking plays a major role. In any case, the analysis parameters (especially those relating to FFTs) must be chosen more carefully when analysing time-varying signals. This subject is not dealt with further here: suffice it to say that FFT length and overlap should be chosen in a reasonable relation to the rates of change of features of interest in the signal. Typically with speech signals, one should choose FFT windows of about 125 ms. As will be seen in the following sections, FFT lengths below 512 points are probably inadvisable.

2.0.2 Steady-state signals

The rest of this study concerns only steady-state signals. In this section, two types of signal have been used to investigate and illustrate the effects of varying different parameters. The two signals are:

1. Pink noise, 50-10000 Hz
2. Sum of seven equal-amplitude sinusoids at 125, 250, 500, 1000, 2000, 4000, 8000 Hz.

Both signals have sample rates of 20480 samples/second. Unless otherwise stated they are applied to AUDMOD at a level of 65 dB SPL in the free field 'coupler' and with no hearing loss.

2.1 The effects of varying the analysis parameters

2.1.1 Parameters available

The following analysis parameters can be adjusted:	Default
1. FFT length (256-8192)	1024
2. FFT overlap (0-<100%)	50%
3. Input FFT frames to average (1-All)	1
4. Transmission factor (3 alternatives)	Zwicker a0
5. Channels per ERB	1
6. Limiting E-band numbers (0-35)	3,35
7. Binaural loudness (Yes, No)	No
8. Initialization (Correct, Fast)	Fast
9. Recording coupler (Free field, 711 (Shaw), 711 (Mehrgardt & Mellert), IEC 303, Diffuse field)	Free field

Not all of these are considered here.

In the following, the default values are used for all parameters other than those explicitly varied, unless otherwise stated.

2.1.2 Spectra to average

It is useful to compare the mean specific loudness values obtained by (i) sending each FFT frame through AUDMOD individually and averaging afterwards ('post-averaging') against (ii) the overall specific loudnesses obtained by averaging all the FFT spectra together before the input to AUDMOD ('pre-averaging'). The comparison is carried out for four FFT lengths, 256, 512, 1024 and 8192 (corresponding to FFT frames of 12.5, 25, 50 and 400 ms).

The results are shown in figure 2.1 for the pink noise signal. For FFT lengths of 1024 or more, the difference is negligible, but for shorter FFTs post-averaging yields the more reliable result. This is presumably because pre-averaging the complex spectra from the short FFTs results in a poor estimate of the true spectral density at low frequencies, where there are very few FFT points per ERB.

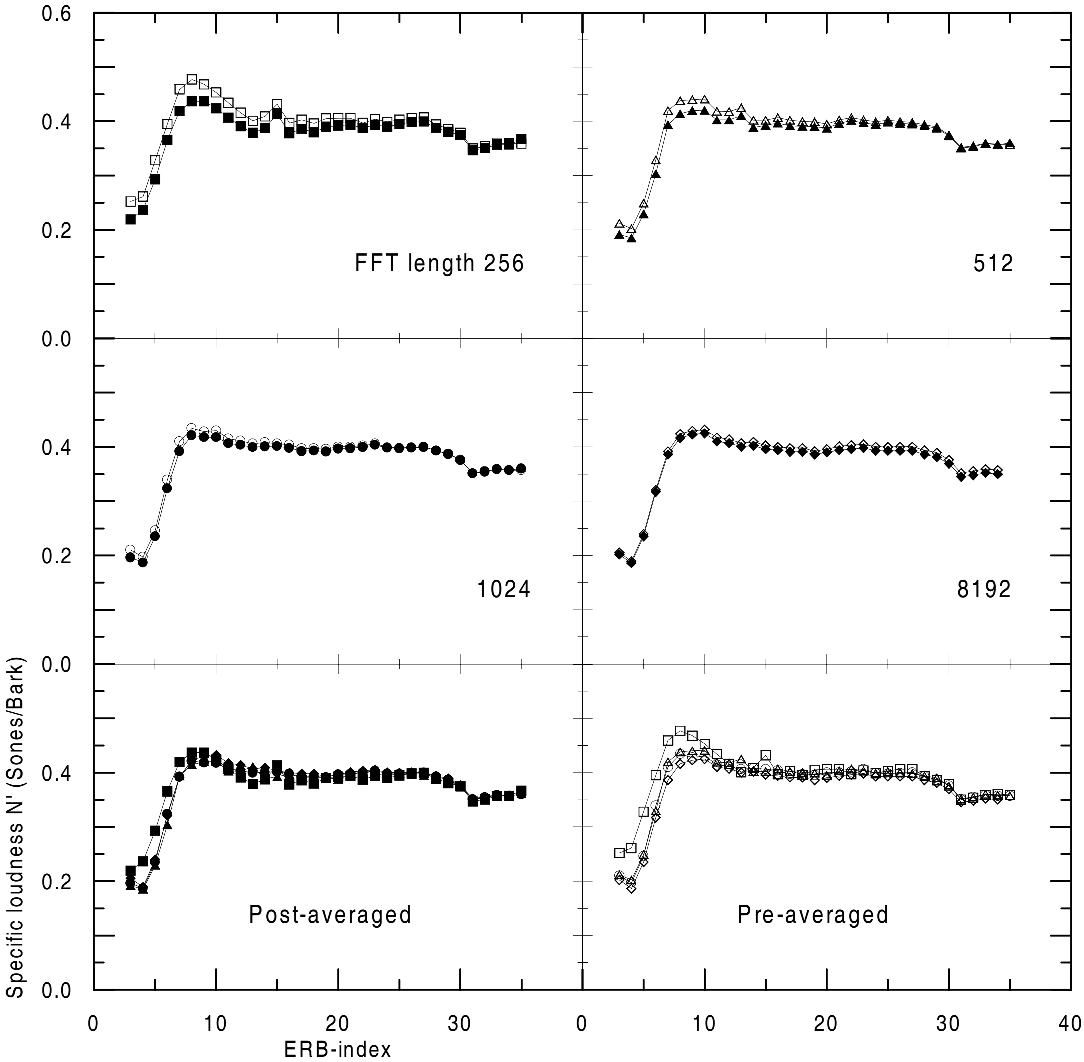


Figure 2.1: *Specific loudness for pink noise signal with varying FFT length and with averaging before and after AUDMOD. Open symbols: pre-averaged, closed symbols: post-averaged.*

Figure 2.2 shows the same comparison carried out with the sinetone complex. A difference between the results of the two methods only shows up for 256-point FFTs, and then only marginally, at low frequencies.

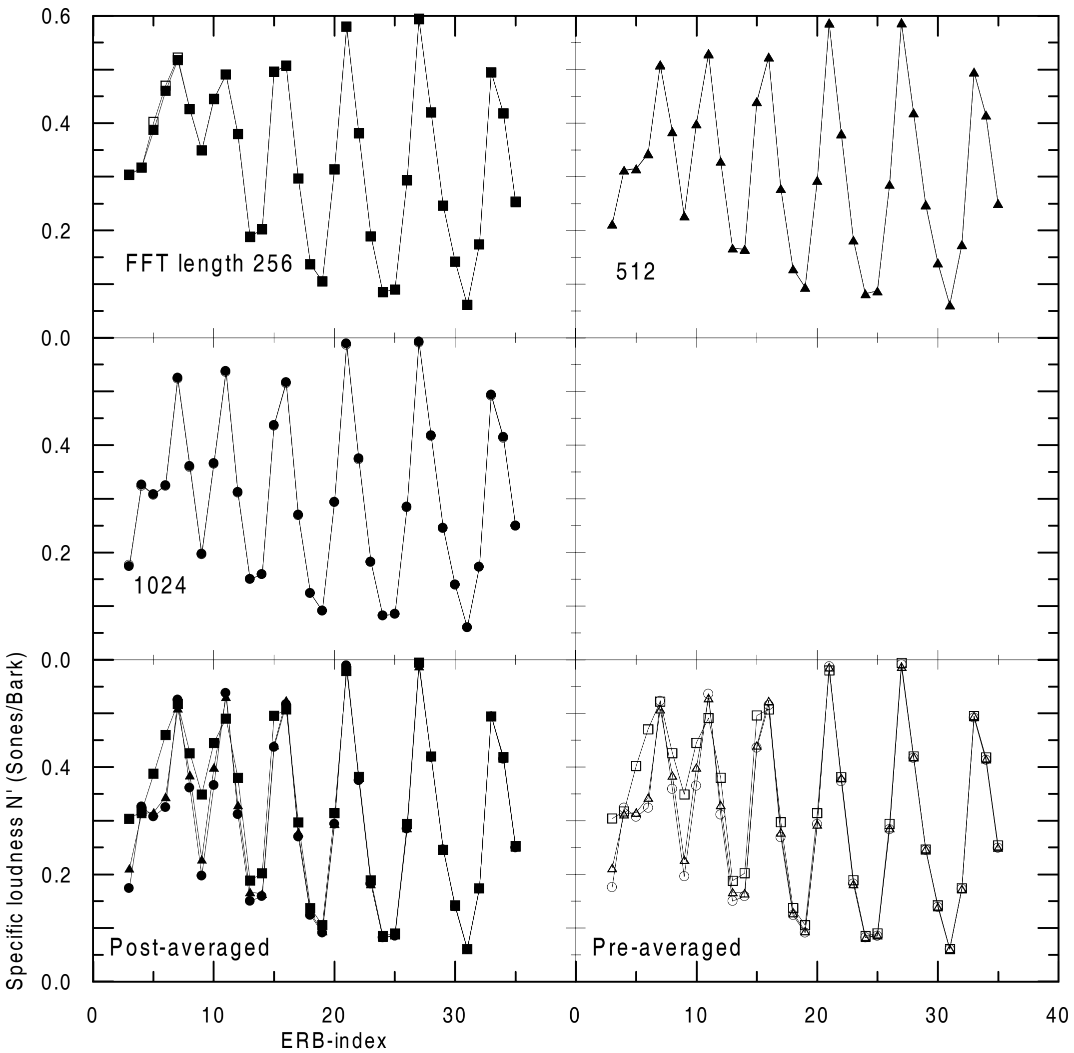


Figure 2.2: *Specific loudness for sinetone complex with varying FFT length and with averaging before and after AUDMOD. Open symbols: pre-averaged, closed symbols: post-averaged.*

Not apparent in the above plots is the effect of missing data in the final FFT frame (i.e. when the final FFT frame extends beyond the end of the signal, as it almost always will). This gives a 'spectral splatter' which can distort the analysis if it is included, and if the total number of frames is not very large. The effect is shown for a single 1000 Hz sine tone in figure 2.3. Note that the total loudness for the final frame is higher, even though the signal energy is lower for this partially empty frame. (NB Although the signal and spectrogram frames appear to end at the same time, they do not really do so. Only the first half of each overlapped FFT frame is shown.)

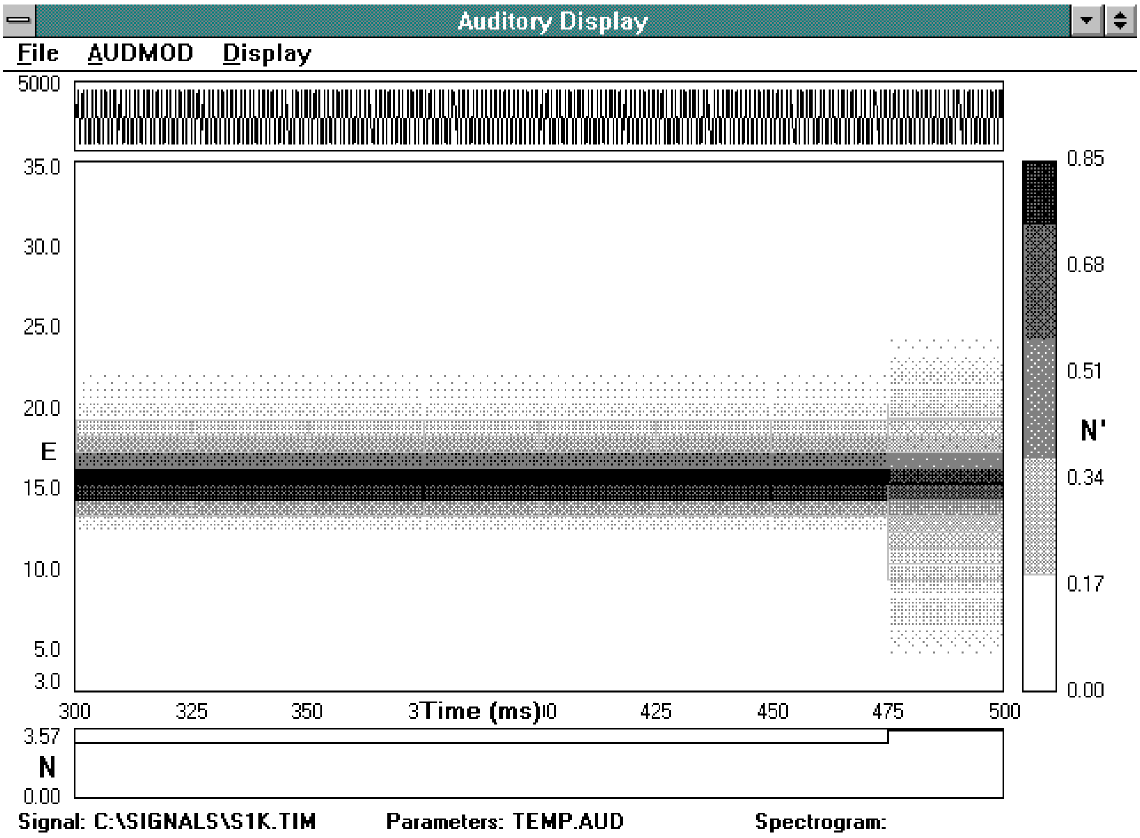


Figure 2.3: *AUDDISP display of analysis of 1 kHz tone. Upper window shows signal, middle window specific loudness 'spectrogram', lower window total loudness.*

Conclusions: Pre-averaging gives much shorter calculation times, and is therefore to be preferred when the FFT length is 1024 or more. Otherwise, post-averaging is probably advisable. Inclusion of the final FFT frame in either kind of average leads to an apparent spread of excitation and consequent errors in specific loudness when the signal is not broadband. Therefore an additional 'All-1' averaging option has been added to AUDDISP and AUDMOD.

2.1.3 FFT length

The estimated signal spectrum is used in two ways: first the energy in each ERB band is used as one of the factors controlling the Roex filters' profiles (=spread of masking as signal level increases), and then the whole spectrum is used as input to the adjusted filter bank, leading to the actual excitation and loudness values appearing at the output.

Figure 2.4 shows the relationships between FFT point frequencies and ERB boundaries below 2 kHz, as the FFT length is varied. To be sure of an accurate estimation of signal energy in a given frequency band, there should be sufficient FFT points within that band. Inadequacy will of course show up at the low frequency end. The lowest ERB index of interest is unlikely to be lower than 3. In fact, the variation of Roex filter profile with ERB power is rather slow, so errors are well tolerated, even completely unsampled bands (interpreted by AUDMOD as having an SPL of -100 dB, i.e. ~150 dB below normal). Nevertheless, to obtain at least one FFT point in each ERB band requires an FFT of at least 512 points.

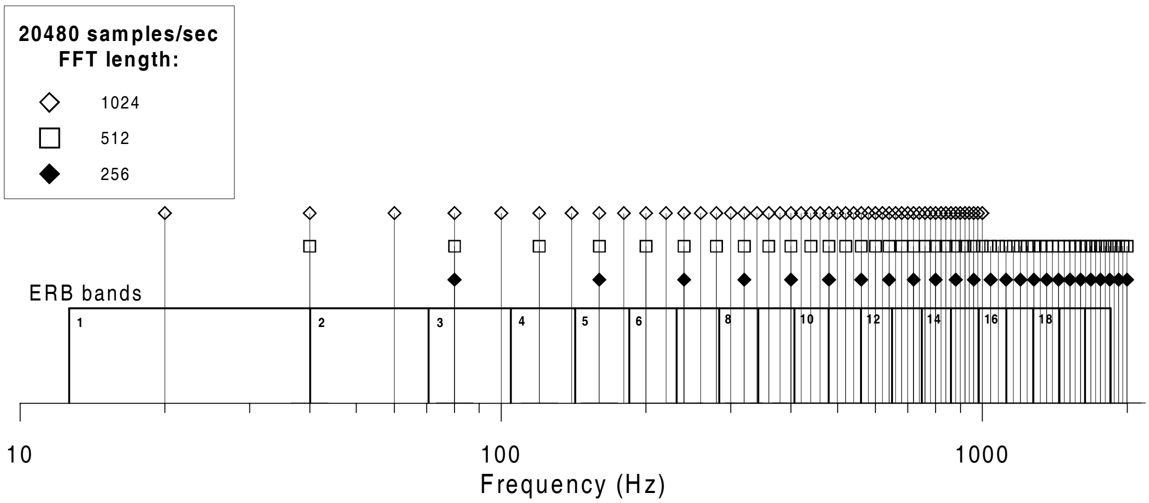


Figure 2.4: *Relationships between FFT point frequencies and ERB boundaries below 2 kHz.*

Roex filter shapes are quite 'peaky' and have steep flanks, so it is desirable to sample the frequency scale finely enough to follow the filter slopes reasonably well. This argues in favour of longer FFTs, without giving any absolute guidelines.

Figure 2.5 shows the output of AUDMOD (mean, standard deviation, maximum and minimum of specific loudness and total loudness values) with pink noise as input, for FFT lengths of 256, 512, 1024 and 8192 points. As expected, the spread of results reduces as FFT length increases. However, the mean values are almost indistinguishable for FFT lengths greater than 256 (see figure 2.1, lower left panel).

The same comparison is shown for the sinetone complex in figure 2.6, and the means are superimposed in the bottom left panel of figure 2.2. With only 256 points in the FFT, some frames yield a very poor estimate of the sine wave amplitudes, even at high frequencies, causing the mean to deviate from that obtained with longer FFTs, which agree closely.

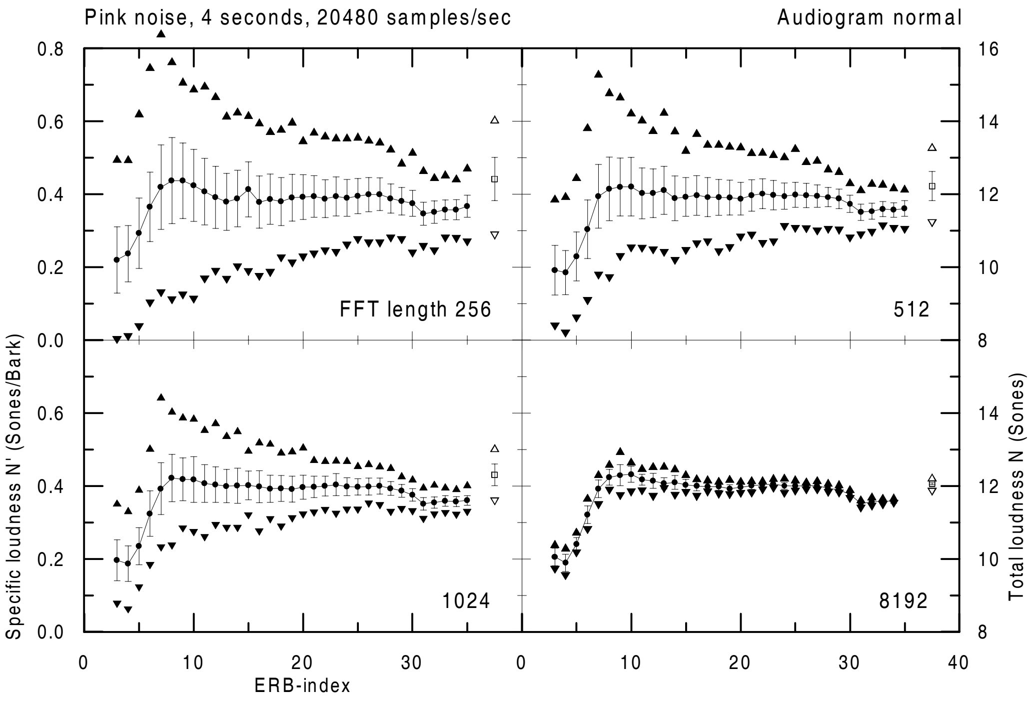


Figure 2.5: M*ean, standard deviation, maximum and minimum of specific loudness and total loudness values of pink noise signal, for various FFT lengths.*

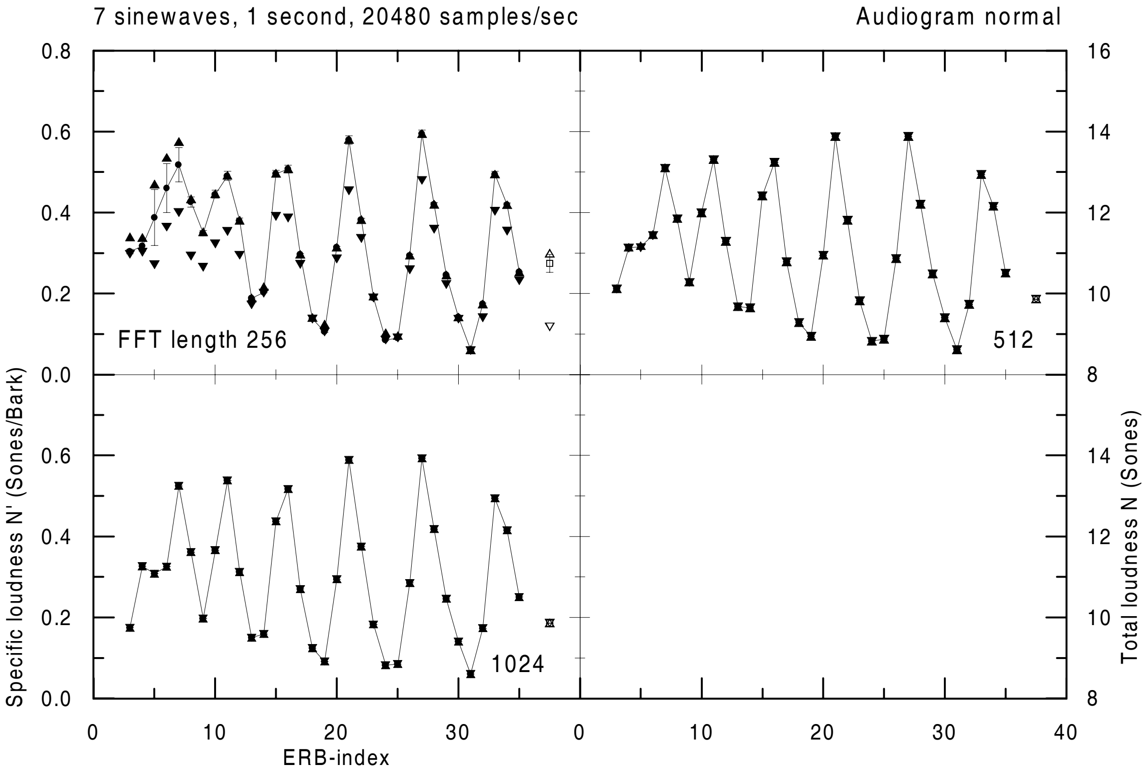


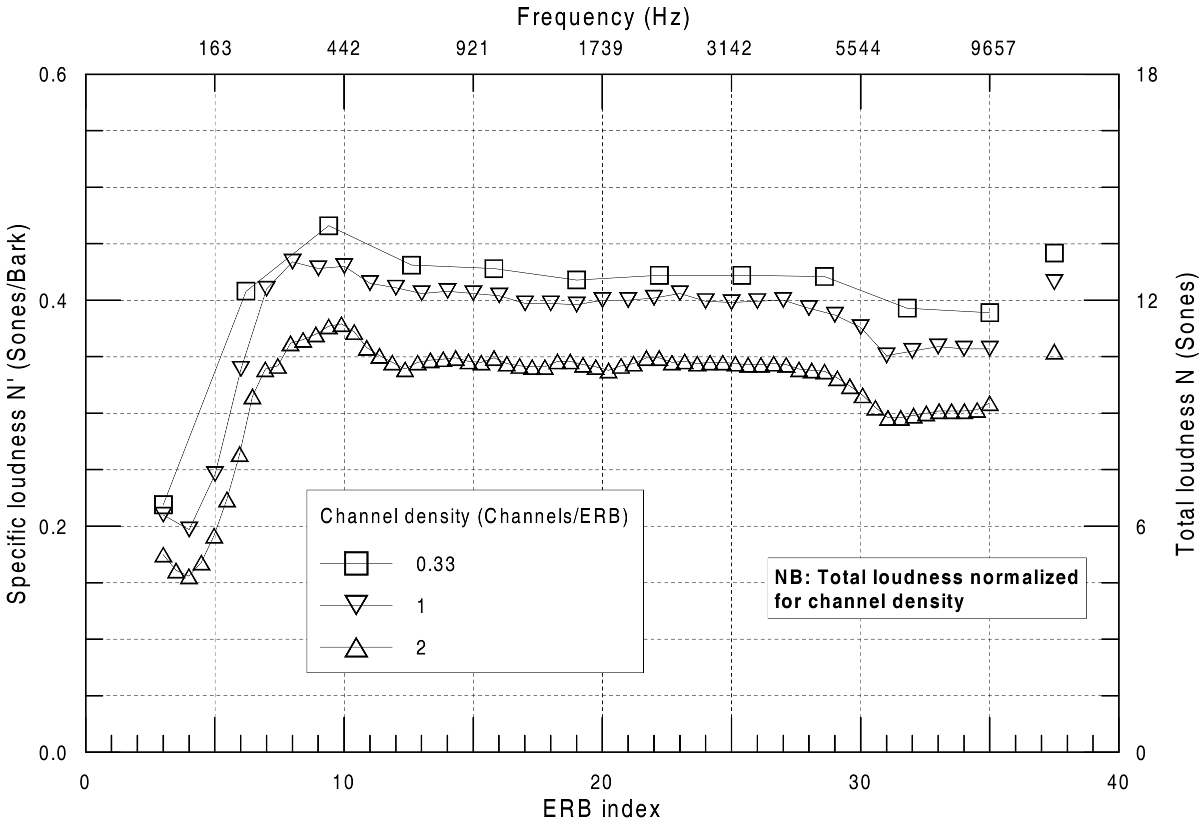
Figure 2.6: M*ean, standard deviation, maximum and minimum of specific loudness and total loudness values of sinetone complex, for various FFT lengths.*

Conclusions: For both noise and tonal signals, FFTs of at least 512 points should be used. The longer the FFT, the better.

2.1.4 Channel density and initialization procedure

The default channel density is 1 per ERB. Although this density arises naturally out of the way the auditory filter bank is conceived, there is no theoretical reason to prevent an analysis being carried out with a different density, so long as the filters are treated as independent. The result should simply be to give an altered resolution of the excitation pattern, without implying anything about the actual placement which the filters have in a real cochlea responding to real sounds. The true cochlear filters are indeed presumed to centre themselves so as to yield maximum signal-to-noise ratio, so using more channels simply corresponds to sampling a greater number of possible filter placements.

With the initialization procedure used in LBN's AUDMOD implementation, intermediate outputs of AUDMOD (ERB energy and Roex filter output or 'Excitation') agree closely regardless of the channel density, but specific loudnesses and total loudness fall as the density increases. This is illustrated in figure 2.7.



Signal: Pink noise, duration 4 seconds.
Analysis: Fs=20480/sec, FFT length 1024, 50% overlap, 161 frames

Figure 2.7: *Specific and total loudness of the pink noise signal as a function of channel density.*

The explanation for the incorrect specific loudness pattern lies with the calibration procedure carried out in AUDMOD before each analysis (see section 3.5.1 of Nielsen (1993) for further details). A complex of sine tones is sent into the model, one tone per filter channel, each at an SPL corresponding to the (interpolated) hearing threshold at the relevant frequency. The excitation thus produced in the given channel is noted and used later for normalization when the Roex filter outputs are converted to specific loudnesses (Eqn. 17 in Nielsen (1993)). This procedure is repeated with tones at 0 dB SPL referred to the free field. In order to save computing time during the calibration, all the tones in the given complex (threshold or 0 dB) are sent into the model simultaneously, instead of one at a time. Since the Roex filters overlap each other, this means that the excitation in a given channel is increased due to contributions from tones intended for neighbouring channels. The later normalization of the test signal's specific loudness pattern will thus be forced downwards. The Roex filter shapes do not get narrower as the number of channels is increased. Hence the greater the density of channels, the lower the apparent specific loudnesses will be. From figure 2.7 it is clear that this error plays a role even with the standard 1 channel/ERB, although its magnitude is probably acceptable.

Since the source of the error lies in the calibration procedure, it is not dependent on the type of signal being analysed. Trials with the seven-tone complex confirm this.

AUDMOD and AUDDISP have been revised to allow the user to choose the fast initialization, in which all the calibration tones are sent through the model simultaneously, or a more strictly correct but <u>much</u> slower procedure whereby the tones are generated one by one. With the revised procedure, the specific and total loudnesses remain stable with varying channel densities.

The question arises as to when it is acceptable to apply the quick initialization procedure. The analyses illustrated in figures 2.8(a)-(c) attempt to answer this question. The figures show the calibration excitations, excitation ratios and specific loudnesses as a function of channel density, for the seven-tone complex and normal hearing. Apparently, no great problems should result from using the quick procedure, when the channel density is at or below 1/ERB. For higher channel densities, the shape of all the curves is well maintained despite increasing displacement. With hearing loss the displacement between the curves does not increase, since the widening of the Roex filters is disabled during calibration.

Thus the quick procedure should be acceptable for studies concerned with relative changes, as long as any displacement is insufficient to force the loudness below threshold mistakenly. When accuracy of absolute loudness values is of primary importance, and channel densities greater than 1/ERB are also required for resolution in the E-domain, the slow procedure must be used.

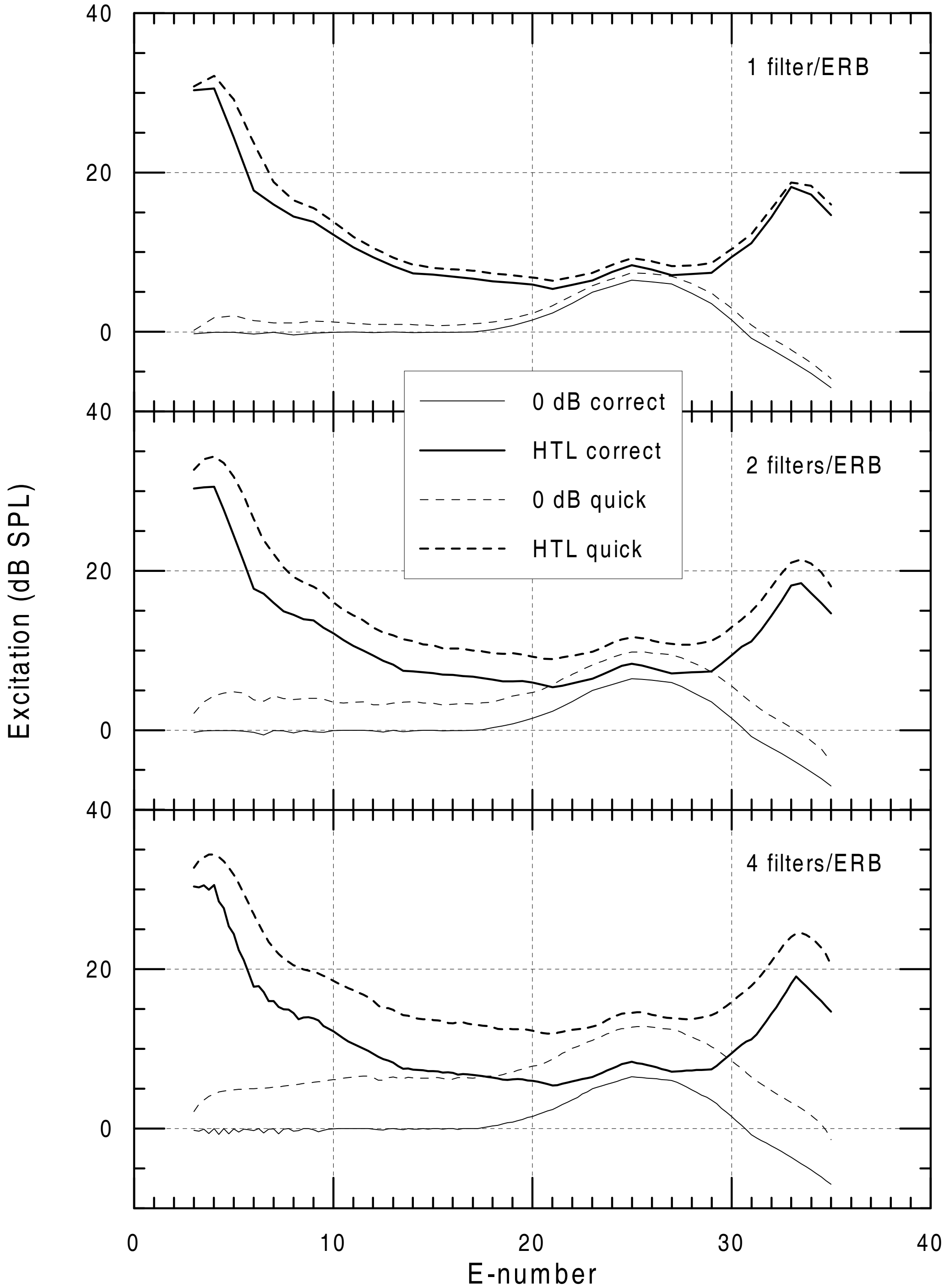


Figure 2.8(a): *Calibration excitations for normal hearing as a function of channel density. Quick (dashed) and correct (solid) initialization.*

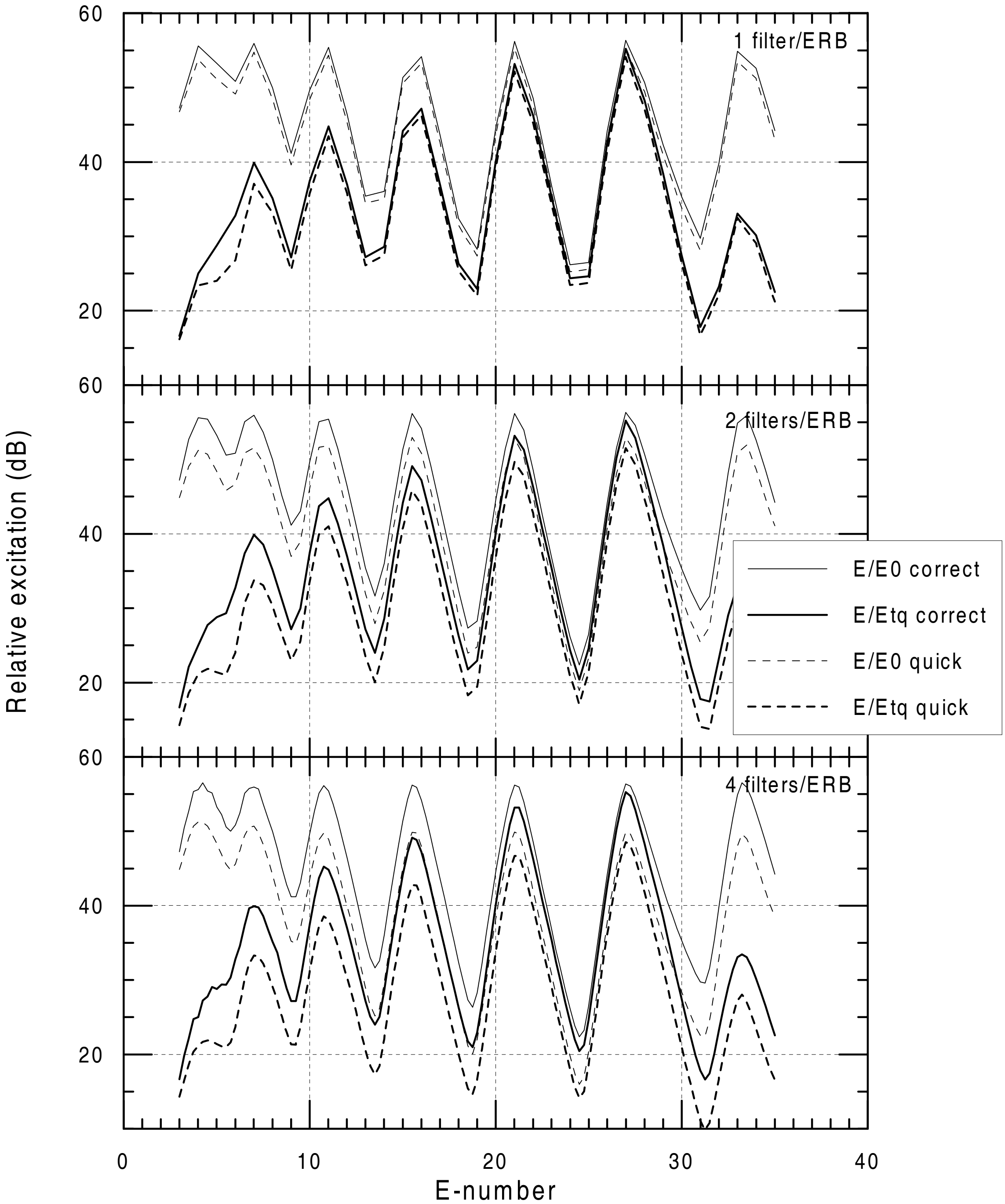


Figure 2.8(b): *Excitation ratios for the sinetone complex with normal hearing as a function of channel density. Quick (dashed) and correct (solid) initialization.*

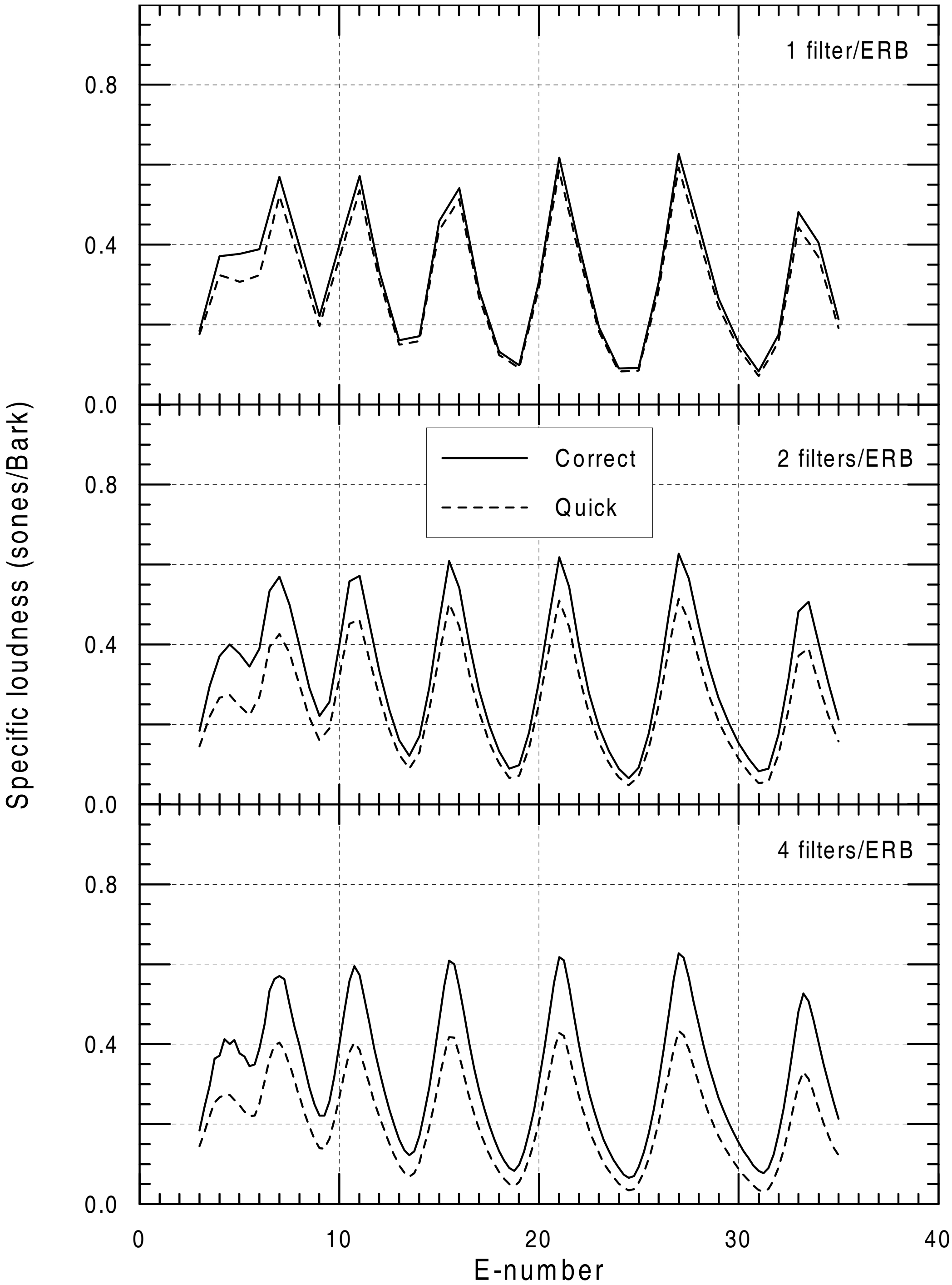


Figure 2.8(c): *Specific loudnesses for the sinetone complex with normal hearing as a function of channel density. Quick (dashed) and correct (solid) initialization.*

Conclusions: Absolute values of specific and total loudness are not invariant under changes of channel density, when the quick calibration procedure is used (they fall as the density increases). The quick procedure is acceptable:

(i) When only relative trends are of interest, and threshold is avoided, or

(ii) When the channel density is 1/ERB or less.

Otherwise the slow procedure should be used.

Filter widening during initialization

On page 48 of Nielsen (1993), it is stated that the widening of the filters with hearing loss is disabled during initialization to avoid elevated threshold excitations arising from the simultaneous excitation. In principle, widening should be enabled if the initialization scheme is correct (i.e. non-simultaneous), but the result for large hearing losses may be an excessively increased excitation at threshold, leading to a lowered specific loudness. This is indeed the case, as shown in figure 2.9.

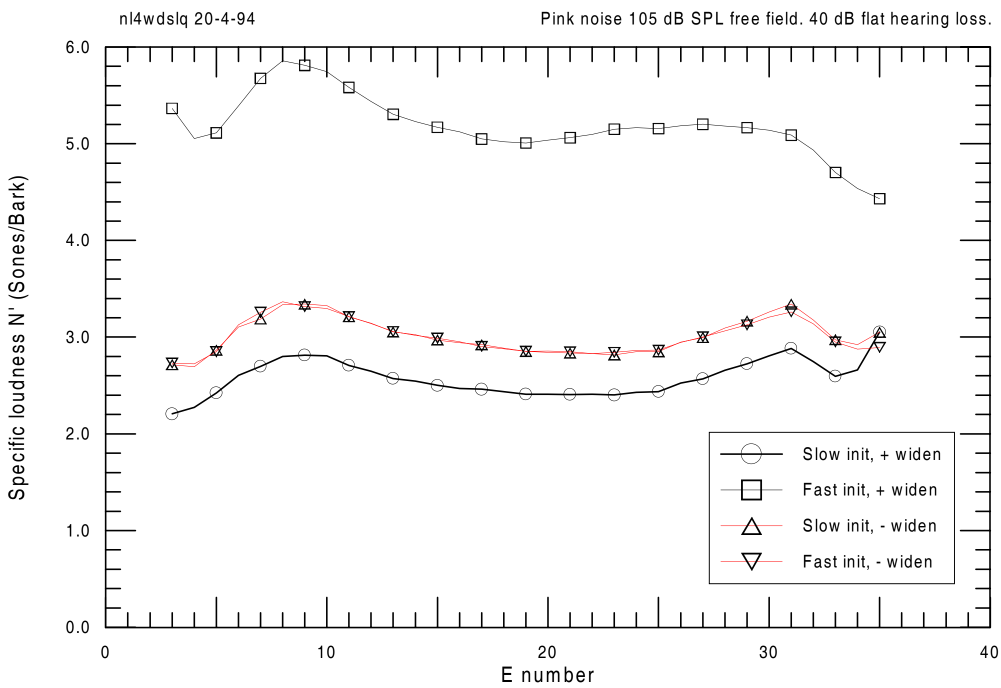


Figure 2.9: *Specific loudness of pink noise signal with flat 40 dB hearing loss for different initializaton conditions.*

Also shown in figure 2.9 is the wildly wrong result of widening in combination with the fast initialization procedure.

Since the inclusion of filter widening causes the apparent threshold to be further elevated above that defined by the audiogram, it has been chosen not to activate it.

2.1.5 Transmission factor

The three alternative transmission factors are shown in figure 2 of Nielsen (1993), and their origins are also described there. Figure 2.11 here shows the two principal ones. Since the same transmission factor occurs during both the calibration phase and the analysis phase, we might expect its effects to cancel, yielding the same loudness pattern regardless of the transmission factor used. Indeed, for the normal hearing case there is very little difference. However, the calibration process is non-linear: not only are calibration constants derived, but also increased input levels or hearing losses cause the Roex filters to be widened prior to the analysis phase. A much more serious effect is though due to a shortcoming in Zwicker's loudness formula. This is further described in section 3.1. The conclusion is that for the time being a0 should always be used.

2.1.6 Recording coupler

Obviously, this should be chosen to correspond to the recording situation actually being modelled. The '711 coupler' options represent approximations to corrections for 'at eardrum' recordings, which will deviate to a greater or lesser extent from the true value required in for example a KEMAR or real-ear recording. The 'free field' option represents a null condition, where no correction is necessary. This is thus the safest alternative, also because the effects mentioned above in connection with choice of transmission factor also occur with variations in coupler correction. A 'diffuse field' coupler correction is strictly speaking inadmissible, but is useful in the context of BINFIT because it allows AUDMOD to generate absolute values of directional effects, rather than effects normalized with respect to the frontal direction (which result from using a 711 coupler correction). At the same time, the general 'open ear' gain characteristic is corrected for. However, as with the 711 coupler, there will be differences between this correction as provided and the true value for the individual ear in which the recording was made. The form of the diffuse field correction is plotted in figure 4.1.

2.2 Behaviour at high frequencies

For applications involving localization, it is essential that the behaviour of AUDMOD is reliable right up to E=35 (9.6 kHz). There are three potential sources of problems at the high frequency end:

1. Breakdowns in the theoretical basis of the model
2. Errors due to missing higher frequency contributions to auditory filter energies
3. Measurement uncertainties in the correction factors used in the model (coupler-to-free-field, free-field-to-eardrum, coupler equivalent threshold SPL)

Any problems which arise should be considered in the light of the application to which the model is to be put. In the present case (BINFIT), it is essentially only relative changes which are of interest. We can thus be tolerant of divergent behaviour, as long as it is consistent and still plausible.

2.2.1 Theoretical breakdown

Breakdowns in the underlying theory are not considered further here. Any breakdown will presumably be a gradual process and (again, presumably) lead to changes in the absolute values of model outputs rather than changes in the mode of behaviour.

2.2.2 Missing high frequency contributions

The specific loudness 'at' a given frequency in reality results from excitation over a wide frequency range around that frequency (see e.g. figure 12 of Nielsen (1993)). In Nielsen (1993) it is suggested that band-limiting the signal (by having a sampling rate just sufficient for the highest ERB band) will cause excitation in the upper flanks of the highest filters to be missed, the result being underestimation of the specific loudness for these bands. In fact there is little downward spread of masking for the highest bands in the model, and additional signal contributions lying above a given ERB band have negligible influence on that band's resulting specific loudness. Experimentation with varying sampling rates confirms this.

2.2.3 Measurement uncertainties

It is important that all contributions, even those coming from the extremes of the frequency range, are at least reasonable, and not subject to extrapolations or correction factors that cause them to vary wildly. The range of frequencies giving significant contributions varies with filter centre frequency, signal level, hearing loss and of course sampling frequency.

The ranges of the various correction factors used were all extended to 12.5 kHz in order adequately to cover the high-frequency region. The basis for making these extensions (all beyond the range found in standards) varied, as detailed below.

Audiogram

The audiogram is defined up to 8 kHz. Above 8 kHz the hearing threshold is assumed to equal that at 8 kHz.

Threshold conversion

During the calibration phase of the model, HTL values are converted to free field SPL by adding the RETSPL in an ear simulator (ISO 389/ADD 1, 1983) and the 6cc-to-free field correction of Bentler & Pavlovic (1992). Both of these had to be extended.

The 6cc-to-free field conversion was extended as follows:
8 kHz: = column L - column N in Bentler and Pavlovic (1989).
10 kHz: Free field-to-eardrum from Shaw and Vaillancourt (1985), ISO 226 to give MAF, guessed MAPC.
12.5 kHz: Guessed.

RETSPL was extended by applying the requirement that the combination of the two corrections should yield approximately the right effective monaural MAF.

Ideally, AUDMOD ought to convert HTL values directly to free field SPL's, by simply adding them to the monaural MAF. Upon inspection of the source code, the consequences of such an apparently minor change appear in fact to be rather global, so this has not been attempted.

Transmission factor

The 'a0' transmission factor is already defined up to 12.5 kHz. 'ELC100' is only defined up to 8 kHz. Its trend was continued with a marginal adjustment to ensure agreement with a0 at 12.5 kHz.

The extension of the frequency ranges has only marginal effect in the normal-hearing case, but is rather more significant when hearing loss is present. Figure 2.10 illustrates this for a flat 40 dB hearing loss.

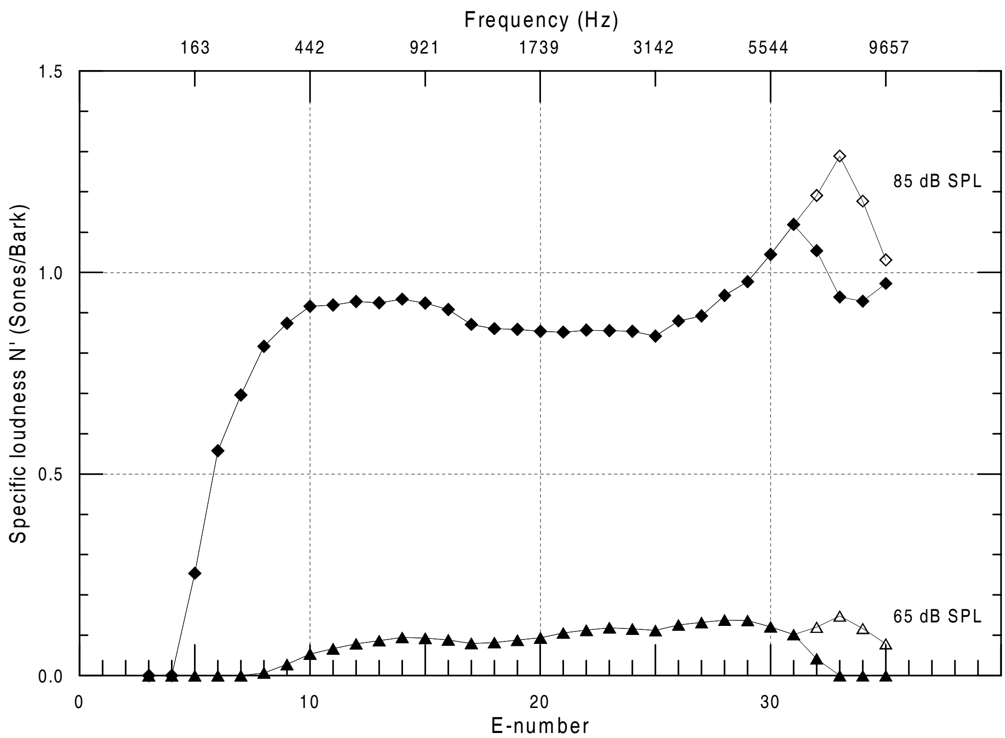


Figure 2.10: *Specific loudness of pink noise signal at 65 and 85 dB for 40 dB flat hearing loss, when using original (open symbols) and extended (filled symbols) ELC100 transmission factor.*

2.2.4 Interpolation and extrapolation of correction curves

The various correction factors used (coupler corrections, transmission factors, audiogram) are only defined at point frequencies, and must be interpolated and extrapolated to provide values at all the spectral lines of the FFT. In LBN's implementation a fifth-order polynomial interpolation is used. Examination of the resulting curves indicates that implausible deviations can occur. Figure 2.11 compares linear and polynomial interpolations for the transmission factors a0 and ELC100 (after extension of ELC100 to 12.5 kHz). The use of a linear interpolation creates only minor discontinuities in the curve, and is free from uncontrolled deviations.

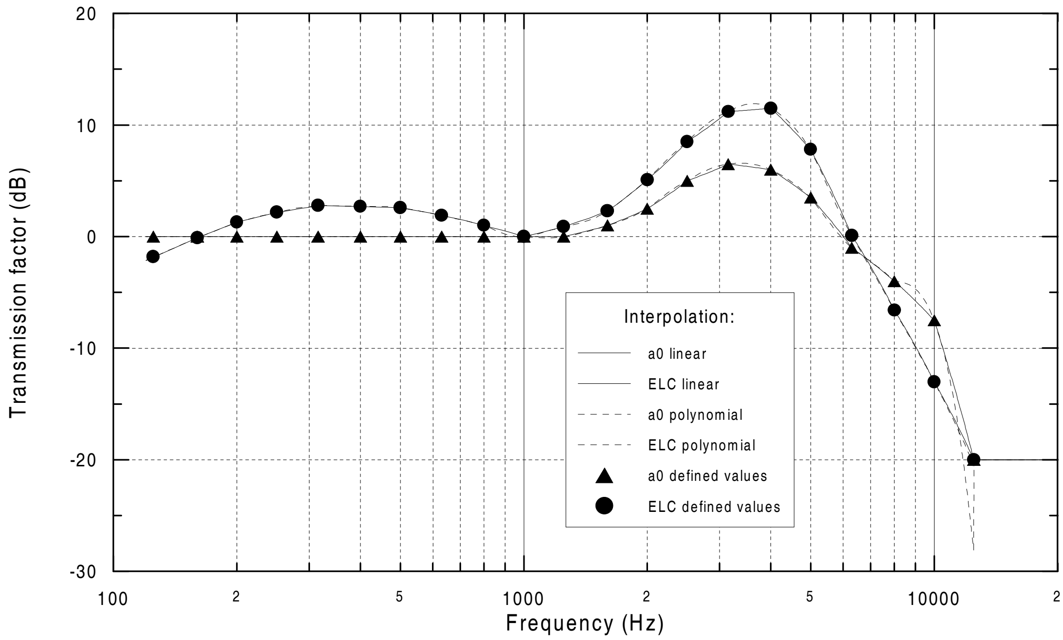


Figure 2.11: *a0 and ELC100 transmission factors: defined point values and linear and polynomial interpolations.*

Extrapolation is carried out simply by assigning to frequencies beyond the defined range the last defined value (i.e. the curve is flat outside the defined frequency range). This is the only sensible method to adopt. In any case, the defined frequency range covers from 125 Hz to 12.5 kHz (see section 2.2.3 re. extensions of corrections to 12.5 kHz).

Conclusion: AUDMOD has been altered to apply linear rather than polynomial interpolations.

3 Transformation from excitation to specific loudness

There are two major problems with the transformation from excitation to specific loudness, which reduce the trustworthiness of the specific loudness outputs. The first of these concerns an inconsistency in Zwicker's loudness formula which causes it to be overly sensitive to changes in the conditioning of the input signal. The second concerns the modelling of recruitment, which also takes place in the same formula. At high sound levels and significant hearing losses, implausible loudness values are generated. Possible solutions to these problems are discussed.

3.1 Effect of changing transmission factor

Figure 3.1 shows the loudness patterns generated with the pink noise signal and 40 dB flat hearing loss, for the 'a0' and 'ELC100' transmission factors. The signal level varies from 65 to 105 dB SPL.

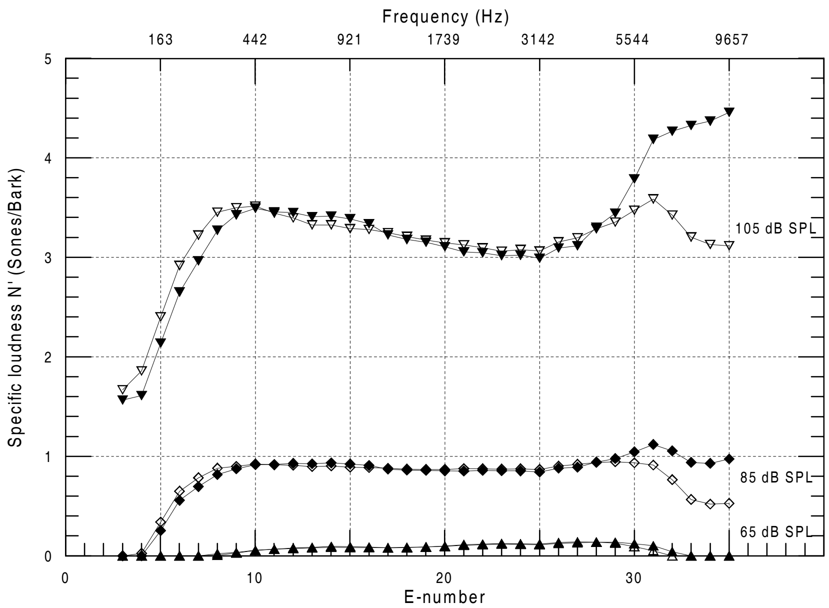


Figure 3.1: *Specific loudness of the pink noise signal at 65, 85 and 105 dB with a 40 dB flat hearing loss. a0 (open symbols) and ELC100 (filled symbols) transmission factors.*

There is a large deviation for E=30 and above, which occurs for all types of signal and requires explanation. The explanation is to be found (after much searching) in the way in which Zwicker's loudness formula combines the excitation values obtained during the calibration phase of the model. The loudness formula (Eqn. 17 in Nielsen (1993)) is shown below, and relates the specific loudness N' in a given frequency channel to the 'excitation' (= Roex filter output, E) generated in that channel by the signal. Also involved are the excitations E_{tq} and E_0 arising from sine tones at the channel's centre frequency. E_{tq} results from a sine tone with a level equal to the supposed hearing threshold (at that frequency) and E_0 results from a level of 0 dB SPL free field. The other quantities are constants which are irrelevant for this discussion.

$$N' = N'_0 \left(\frac{E_{tq}}{sE_0} \right)^{0.23} \left\{ \left[(1-s) + \frac{sE}{E_{tq}} \right]^{0.23} - 1 \right\}$$

Figure 3.2 shows E, E_{tq} and E_0 as a function of E-number when analyzing a 4 kHz tone with a flat 40 dB hearing loss and via the the a0 and ELC100 transmission factors (TF's).

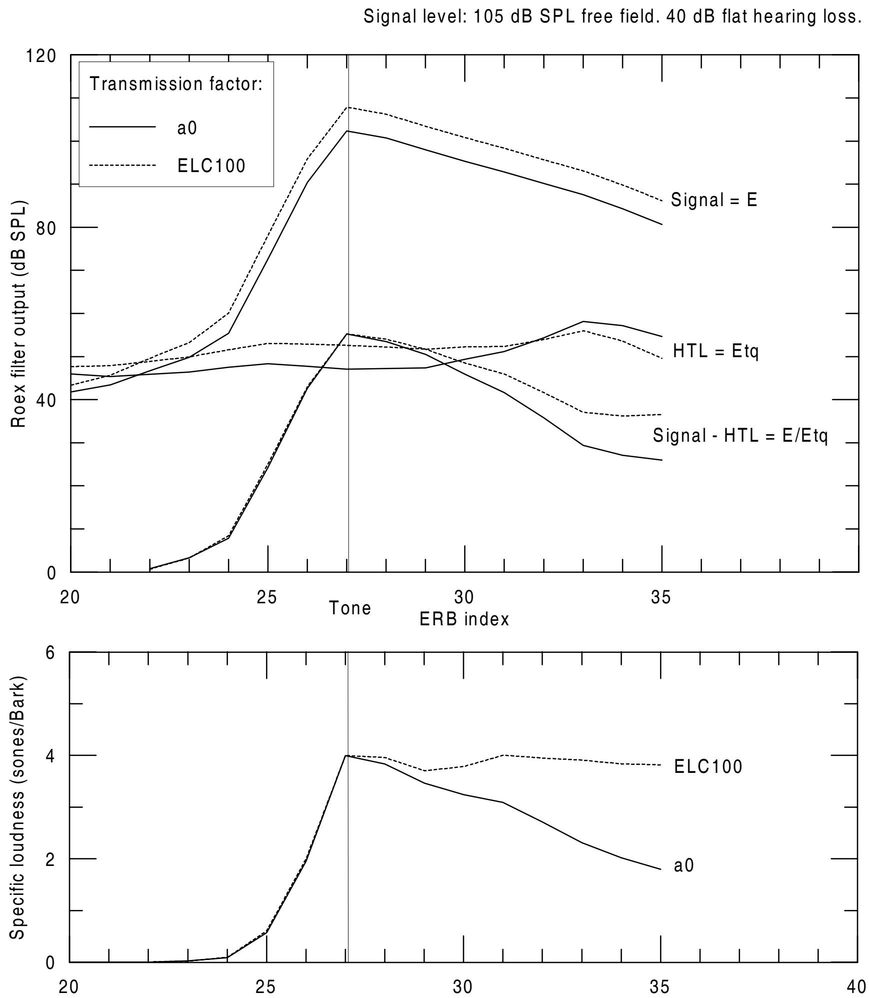


Figure 3.2: *Excitations and specific loudness of a 4 kHz tone at 105 dB SPL, via a 40 dB flat hearing loss and both a0 (solid) and ELC100 (dashed) transmission factors.*

For the two TF's, the difference in *E* <u>at all frequencies</u> equals the difference in TF at the tone frequency (reducing at distant lower frequencies). The difference between the two E_{tq} curves equals the <u>frequency-dependent</u> difference between the TF's. Thus E/E_{tq} is not the same for both TF's, and shows the same tendency as the specific loudness curves. In a given analysis channel, *E* is affected by the TF at all frequencies where a signal is present, whereas E_{tq} is only affected by the TF at that frequency. The strength of the effect depends also on hearing loss, since greater hearing loss leads to greater spread of masking and thereby spread of excitation in *E*. For normal hearing, there is almost no effect of TF. The generality and applicability of Zwicker's formula is thus undermined. The specific loudness it generates will vary as a result of intractable interactions between variations in signal, TF and hearing loss.

Since the a0 transmission factor and the loudness formula were to some extent developed in tandem, the wisest course is always to use the a0 factor in AUDMOD analyses. This is further supported by the results of analyses in Nielsen (1993), in which *E* values calculated using a0 agree well with experimental data.

3.2 Recruitment

Figure 3.3 shows the growth of loudness predicted by AUDMOD for a 1 kHz tone and for four levels of hearing loss. Also shown is the 'classical' theoretical curve showing a doubling of loudness for every 10 dB increase in SPL.

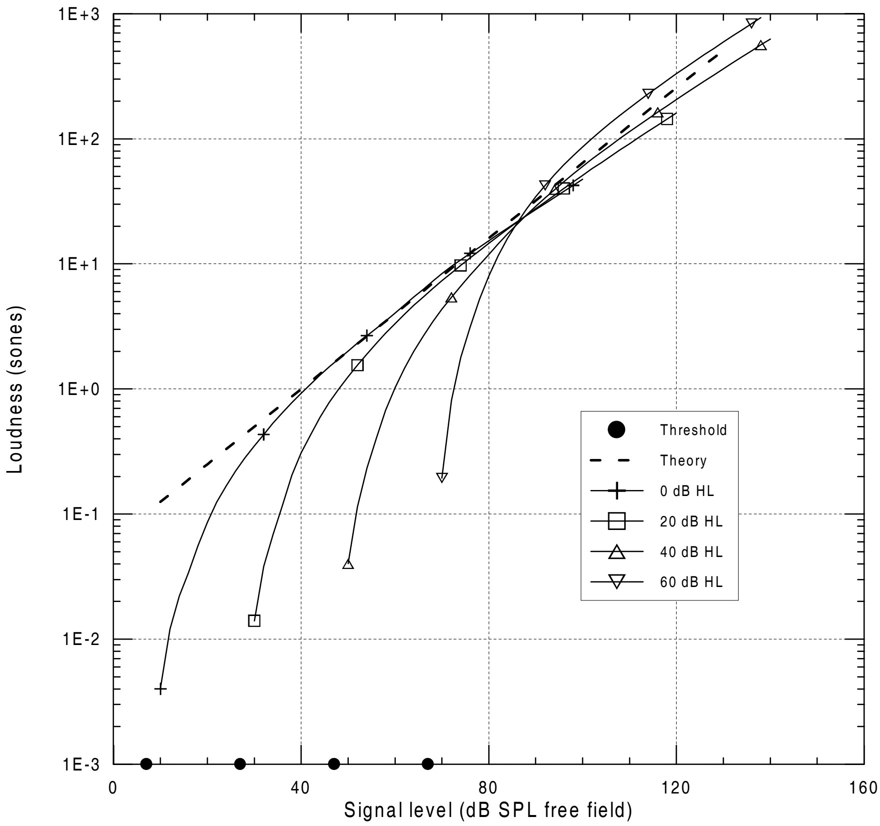


Figure 3.3: *Predicted loudness growth curves for a 1 kHz tone. Hearing loss 0, 20, 40, 60 dB.*

Assuming the theoretical curve to be a valid indicator of the asymptotic behaviour, the following observations can be made:

(i) At first sight, the behaviour of the predictions is fairly good. However, the plot is made on the usual log(sones) axis. By definition, the sone is a fundamentally linear unit; that is, 1 sone at 100 sones is 'as large as' 1 sone at 5 sones. If the data were replotted on a linear sone scale, the deviations from theory at high levels would take on great importance.

(ii) The exponent for high levels appears to be slightly lower than the theoretical exponent.

(iii) The predictions for 0 dB and 20 dB HL look very plausible up to 80 dB SPL, and for 40 dB HL up to 100 dB SPL. Thereafter the predictions veer away from the theoretical curve on the low side.

(iv) For 60 dB HL, 'over-recruitment' occurs, followed again by a drooping to below the theoretical curve.

Even if we disregard the theoretical line, the deviations in loudness at high levels seem implausible, both in their magnitude and in the directions they go. There is fair evidence from loudness scaling experiments [Elberling (1994)] that UCL on average corresponds to the same loudness for both normal and impaired subjects. If that is the case, there are several implausible features in figure 3.3:

(i) The average UCL for normals is about the same as that of subjects with a 60 dB loss [Elberling & Nielsen (1993)]. In figure 3.3, these two curves cross around 85 dB SPL, about 30 dB too low.

(ii) At higher levels, UCL with a 60 dB loss is predicted as being typically 20 dB <u>lower</u> than with normal hearing.

If we further drop the assumption of equal loudness for all subjects at UCL, then in the range of mean UCL's for 0-60 dB loss (105-115 dB SPL), predicted loudness is about twice as great with a 60 dB loss as it is with normal hearing. This seems unlikely.

3.3 Possible solutions

Even if the use of the a0 transmission factor generates plausible results in terms of excitation, it is still hard to place complete trust in Zwicker's loudness formula, especially when the recruitment behaviour also is considered. Discussions with Brian Moore and Thomas Baer indicate that they also have reservations, and indeed are trying to develop an alternative (much more complex) transformation. For the time being, excitation appears to be a more reliable quantity than loudness in auditory models, though unfortunately it is neither directly measurable psychophysically nor an easily-defined concept.

Regarding the recruitment problem, it may be possible to adjust the exponents in the loudness formula to yield a 'better' agreement with the theoretical asymptote (see also Nielsen (1993)). An alternative and quite different approach would be to abandon the sone scale, and aim directly for loudness scaling values as they would be produced in e.g. a magnitude estimation task. Comparing predicted excitation values and scaling values obtained in the laboratory for a range of signals, it may be possible to estimate a mapping from excitation to loudness scaling value (for instance with the intervention of a neural network).

4 Revisions to AUDMOD

Choice of recording coupler

For reasons given in section 2.1.6, a 'Diffuse field' 'coupler' has been added to the available options. In the .AUD parameter file, it is specified by a value of 5 on the `recording coupler` line. Figure 4.1 shows the form of this correction. It is derived partly from laboratory estimates made using KEMAR with the small pinna, and partly from Shaw (1980) (also shown).

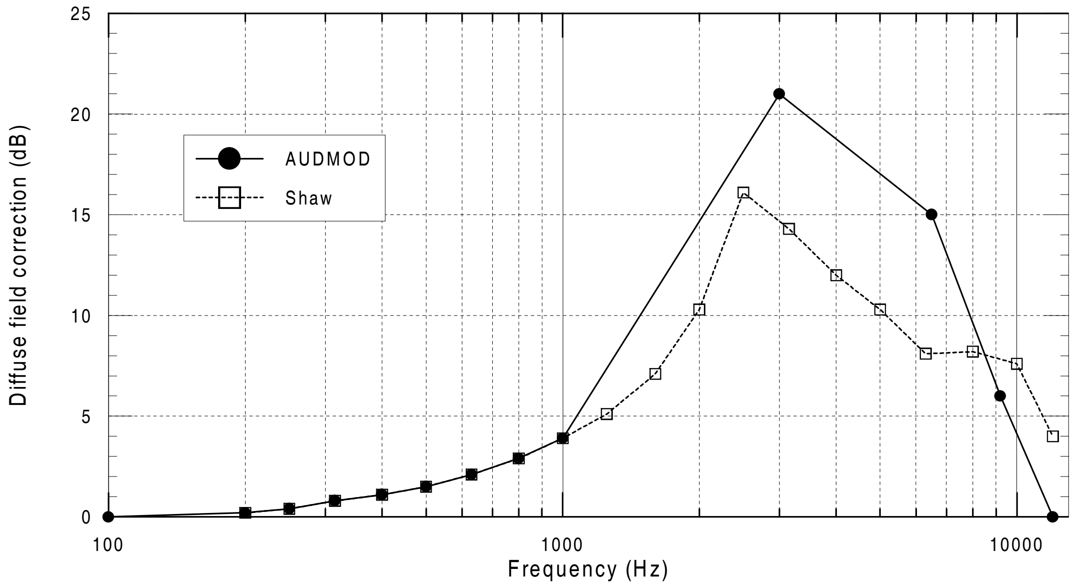


Figure 4.1: *Diffuse field 'coupler' correction implemented in AUDMOD and according to Shaw.*

Choice of fast or slow initialization

For reasons discussed in section 2.1.4, an option has been added to AUDMOD and AUDDISP allowing the choice of a fast or a correct initialization. The default is a fast initialization. The format of the .AUD parameter file is expanded to include a line:

```
Initialization: 0
```

where a 0 indicates slow initialization and a 1 indicates fast initialization.

This extra line follows immediately after the 'Binaural' specification line in the text file.

Spectra to average

For reasons discussed in section 2.1.3, AUDMOD and AUDDISP have been revised to allow the choice of 'All frames except the last one', in addition to the previous 'All frames' or a specified number. In the .AUD parameter file, this corresponds to setting the 'Process' line to

```
Process: -1
```

Summation of specific loudnesses

The summation of specific loudnesses to total loudness has been revised to remove the assumption of 1 channel/ERB. The sum of specific loudnesses is simply normalized by the channel density.

Interpolation of pointwise-defined quantities

Several spectral quantities within AUDMOD are defined only at a few point frequencies (e.g audiogram, coupler correction, transmission factor). These have to be interpolated for application to FFT line spectra. LBN's polynomial interpolation has been replaced with a linear interpolation. For reasons see section 2.2.4.

References

Bentler, R.A. and C.V. Pavlovic 'Transfer functions and correction factors used in hearing aid evaluation and research', Ear and Hearing **10**(1), 58-63, 1989.

Bentler, R.A. and C.V. Pavlovic Addendum to 'Transfer functions and correction factors used in hearing aid evaluation and research', Ear and Hearing **13**(4), 284-286, 1992.

Elberling, C. and C. Nielsen 'The dynamics of speech and the auditory dynamic range in sensori-neural hearing impairment', Report 41-8-2, Oticon Research Unit, 1993.

Elberling, C. 'Loudness scaling versus comfort and discomfort levels in normal and impaired hearing', paper presented at 'Issues in Advanced Hearing Aid Research', Lake Arrowhead, USA, 1994.

ISO 226: 1987 'Acoustics - Normal equal-loudness contours'

ISO 389: 1975/ADD 1 1983 'Acoustics - Standard reference zero for the calibration of pure-tone audiometers'

Nielsen, L.B. 'An auditory model with hearing loss', Report no. 43-8-2, Oticon Research Unit, 1993.

Shaw, E.A.G. 'The acoustics of the external ear' in Studebaker, G.A. and I. Hochberg (Eds.) 'Acoustical factors affecting hearing aid performance', University Park Press, Baltimore, 1980.

Shaw, E.A.G. and M.M. Vaillancourt 'Transformation of sound-pressure level from the free field to the eardrum presented in numerical form', J. Acoust. Soc. Am. **78**(3), 1120-1123, 1985.