Multichannel level alignment, part I: Signals and methods

Pekka Suokuisma
Nokia Research Center, Speech and Audio Systems Laboratory, Tampere, Finland
Pekka.Suokuisma@research.nokia.com

Nick Zacharov
Nokia Research Center, Speech and Audio Systems Laboratory, Tampere, Finland
Nick.Zacharov@research.nokia.com

Søren Bech
Bang and Olufsen, Struer, Denmark
sbe@bang-olufsen.dk

Abstract

This work forms part of the studies of the Eureka 1653 Medusa (Multichannel Enhancement of Domestic User Stereo Applications) project. This is the first paper in a series dealing with level calibration of multichannel sound systems. The paper focuses upon the different methods for evaluating loudness and the development of new signals for multichannel level alignment. The subjective evaluation of these signals is presented and discussed in part II.

1 Introduction

This work forms part of the studies of the Eureka 1653 Medusa (Multichannel Enhancement of Domestic User Stereo Applications) project. The Medusa project is a 3.5 years joint research project with the following partners: British Broadcasting Corporation, The Music Department of the University of Surrey, Nokia Research Centre, Genelec Oy, and Bang & Olufsen A/S.

The purpose of the project is to examine the variables of the domestic multichannel sound system, with and without picture, to carry out the essential optimisation leading to consumer end products. These products will combine the requirements of multichannel reproduction together with the less complex modes of reproduction, such as stereo and mono. These, of necessity, will involve linked studies of programme production and perceptual elements, leading to a single optimised approach to domestic reproduction.

This is the first paper in a series dealing with level calibration of multichannel sound systems. The purpose of the relative level calibration is to ensure that the spatial properties of the programme material are reproduced, as closely as possible, in accordance with the intentions of the programme maker. This is well known for two-channel systems intended for reproduction of stereophonic signals. The main requirements for such systems are 1) that the reproduction level at the listening position is identical for both loudspeakers for the same applied signal and 2) that the listener is positioned on the line of symmetry such that the signals from the two loudspeakers arrive at the listening position simultaneously. This will ensure the optimum reproduction of the spatial properties of the programme material.

Today, the so-called 5.1 multichannel reproduction systems include three channels for reproduction of frontal information and two channels for surround information. The spatial reproduction of a 5.1 system thus includes the additional factor of front-back localisation, ambience or surround sound level, and balance between the two surround channels when compared to the conventional two-channel system. Correct alignment of a multichannel system will thus have a higher degree of importance compared to conventional two-channel systems.

The large number of different reproduction loudspeakers employed further increases the complexity of such 5-channel systems. Often both the directivity and bandwidth of such loudspeakers may differ for the left/right, centre and surround channels. In addition, the practicalities of domestic set-ups may also lead to a compromise in the geometrical set-up proposed in ITU-R BS 775 [8] leading to asymmetries in the room/loudspeaker interaction.

The importance of relative level alignment has been recognised by the film industry for a long time, and today quite elaborate measurement schemes exist to ensure that cinema reproduction systems fulfil the requirements. Various schemes have also been established for domestic systems, however, not in a standardised form. Bech [4] has shown that the calibration signal can have a significant influence on the relative level calibration for a system including three separate front channels and one surround channel (3/1 system). It was further shown that the observed differences in level calibration lead to significant differences in the quality of reproduction of spatial information and overall quality for standard film material.

Bech [4] suggested that a number of factors, in addition to the signal, could have an influence on the relative level calibration: the acoustical and physical characteristics of the reproduction room, the reproduction system, and the number of separate channels.

It has been the aim of one of the subtasks of the Medusa project to investigate these additional factors for a 5.1 reproduction system. Zacharov et al [15] discuss the influence of the calibration signal, the reproduction room, and the physical position of the loudspeakers of the reproduction system and the present paper discusses the generation of the different signals that were investigated by Zacharov et al.

2 Psychoacoustic background of level calibration

The task of the subject in a level calibration experiment is to adjust the level of a channel to be subjectively equal to that of a reference. The reference is traditionally the centre channel and the four other channels are then adjusted to match the centre. The experimental paradigm that traditionally has been used is a so-called method of adjustment (MOA) [5]. This method allows the subject to switch at will between the reference (the centre) and the channel in question and on a continuous basis adjust the level of the variable until it is subjectively equal to the reference. The initial presentation level of the variable channel is either to high or too low to establish a clear difference between the centre and the channel. The initial level is randomly selected to minimise bias effects. The MOA procedure is not optimal seen from a psychophysical view as it includes a number of bias effects that could be avoided if other methods were employed [5]. The influence of the MOA procedure is the topic of a forthcoming paper.

The subjects are asked to match the channels to be equally loud which, in terms of psychoacoustics, corresponds to a match the loudness of the two presentations (channels). This is a standard task in psychoacoustics, however, the usual assumption is that the two signals only differs in terms of loudness[*]. This is often not the case for a multichannel sound system positioned in a domestic room. Due to the different positions of the loudspeakers i.e. the angle of incidence of the direct sound to the listener (it is assumed that the listener is constantly facing the centre channel) quite large difference in timbre can exist between the channels[†]. These additional differences could be expected to influence the loudness matching so it should not, per se, be assumed that the subjects are only using loudness cue for the matching.

Aarts [1, 2] employed a similar such subjective paradigm to test and correlate subjective alignment of signals with objective measures. Seven loudspeakers of different quality and a panel of five naive and expert listeners with normal hearing were employed. Subject adjusted the level of loudspeakers to equal a reference loudspeaker using pink noise. The one-third-octave levels were recorded based on which the metrics were calculated. These metrics were compared to subjective results. The conclusions of this study where that B-weighted sound pressure level (SPL) and loudness metrics according to Zwicker provided the best correlation to the subjective alignment, whilst A-weighted SPL and Stevens method provided an inferior correlation.

Bech [4] tested the assumption that the subjects are using loudness as the main cue by calculating the loudness according to ISO [7] for the centre channel and the adjusted channels. The calculation was based on the average one-third octave spectra of four spectra measured in spatially different positions around the listening position. The results suggested that the subjects had used overall loudness as the main cue for the calibration.

The hypothesis was put forward that the subjective strategy adopted for relative level calibration is based on using the spectral regions of the calibration signal that are most sensitive to level changes. This will be a region where the specific loudness function of the calibration signal is constantly decreasing as a function of frequency as the up-word spread of masking function of the hearing system will results in larger relative changes in specific loudness in frequency ranges characterised by a constant decreasing level of specific loudness. This is compared to frequency ranges with a constant or increasing level.

The use of different frequency ranges for the signals tested by Bech [4] could explain why the signal resulted in different calibration levels: the sound field at the listening position will be different for the different loudspeakers (channels) as a function of frequency due to the different positions of the loudspeakers. The hypothesis also suggests that the investigated, band limited signals might not be the optimum choice for a calibration signal. The programme material will include the whole frequency range so the subjective calibration should also be based on the whole frequency range. This suggests that, if the above hypothesis is correct, the optimal calibration signal should force the subject to use the whole frequency range and that this could happen for a signal with a constant specific loudness spectrum.

To test the hypothesis a number of different signals were generated and the remaining

[*] Note that experiments have been reported where the stimuli differ in other aspects as well.
[†] Here it is assumed that similar loudspeakers are used for all channels, as different types of loudspeakers will only emphasise the problem.

chapters of this paper describes this process.

3 Proposed test signals

In an effort to define the optimum calibration signal in accordance with the above hypothesis, two categories of signal have been considered:

- Uniform excitation noise spectrum
- Constant specific loudness spectrum

In both cases it is necessary to have an auditory model to be able to predict the specific loudness and excitation spectra. Several perceptual loudness models have been developed in recent times and this small section is included only as a brief introduction to the complex topic.

The loudness models suggested by Paulus and Zwicker [14] and Moore et al [10] follow the basic function illustrated in Figure 1. The Zwicker and Moore models differ in a number of areas of which the most important are

- The characteristics of the transmission through the outer and middle ear
- Calculation of the excitation patterns
- Transformation of the excitation to a specific loudness scale

The details of the models are outside the scope of this paper and the interested reader is referred to the original publications. For ease of understanding the presented data, a comparison between different frequency scales is presented in figure 2 and appendix 2.

Both models have been implemented in accordance with the specifications as Matlab files providing the possibility of fast optimisation of the test signals, which will be discussed later. The Zwicker model allows for estimation of the specific loudness spectrum assuming free or diffuse field conditions at the listening position. The characteristics of the sound field at the listening position, as a function of frequency, depend on the directivity characteristics of the loudspeaker and the acoustical conditions in the room. The sound field will neither be free nor diffuse so to include all possibilities a free field and diffuse field version were calculated for the Zwicker model.

In addition to the designed signals the three signal used by Bech [4] were included plus a new signal that has established itself as a de facto standard for calibration of 5.1 systems. The signals and their characteristics are shown in Table 1.

4 Motivations for and generation of test signals

The following section describes the motivations behind and the generation of the nine test signals used in these studies, which are also summarised in appendix 1. To assist in understanding these signals, details spectral plot have been created in figure 3-11 to illustrate the spectra in different domains. For clarity, the spectrum of each signal is plotted in one-third-octaves, FFT spectrum, Zwicker loudness (diffuse and free field) and Moore loudness[‡] (diffuse field) domains. Additionally, the linear, A-, B-, C- and D- weighted sound pressure levels (SPL) are provided.

In accordance with the needs of the subjective tests [15] all signals have been designed for a reproduction level of 20 Sones (Zwicker diffuse field).

All test signals were obtained using Matlab 4.2c.1 from the same white noise signal of length of five seconds. A 48 kHz sampling rate was used. This source signal was quantized to 16-bit representation as well as all the others. The white noise signal was generated using Matlab-command *randn()* to give 240 000 samples of uniformly distributed noise.
All signals were normalised to give the same RMS-power over whole sequence length by calculating the ratio over the power of original white signal and that of signal to be normalised and by multiplying the latter by that ratio. The total power was calculated according to equation (4.1) [12].

$$Power = \sqrt{\frac{(\sum x(n)^2)}{length(x(n))}} \tag{4.1}$$

All signals were stored in 16-bit wav-format and are presented below in order of generation complexity.

4.1 Pink noise (Signal 8)

Pink noise, often used in auditory research, is shaped to provide a uniform octave band level, i.e. has a spectrum whose level decreases at 3 dB per doubling in frequency (see figure 10). This type of noise signal is very widespread in all areas of perceptual research and has been employed in the studies of level alignment [2, 3]

Pink noise was obtained from white noise using inverse fast Fourier transform (IFFT). Because sampling rated used was 48 kHz, FFT-length was chosen to be 48 000. From linear 48 000-point frequency vector the desired amplitude response vector was calculated by equation (4.2).

$$A_d = 10^{-3 \cdot \log_2(\frac{f}{f(1)})/20} = 10^{-0.5 \cdot \log_{10}(\frac{f}{f(1)})} = \sqrt{\frac{f(1)}{f}} \tag{4.2}$$

Next the 48 000-bin FFT was calculated and the obtained magnitude response was replaced by desired response. By using the original phase response the new magnitude-phase pair was transformed back to time-domain by IFFT. Taking the real part of the IFFT result, the pink noise sequence of 48 000 samples was obtained. The total sequence of five second was the result of five similar block transforms.

‡ The Moore model presented in [10] is employed for this work, whilst the development of an updated Moore model has been presented in [11] which also includes a free field assumption.

4.2 B-weighted pink noise (Signal 9)

The B-weighted pink noise signal is a shaped signal that provides a improved perceptual shaping of the noise, approximately according to the 70 phon equal loudness contour (see figure 4).

Analytical equation to give weighting curve for B-weighting is presented in equation (4.3). Using this as a desired magnitude response, a long FIR-filter with linear phase response was designed by frequency sampling method with least squared error optimisation [13]. The signal was obtained by filtering the previous pink noise by Matlab-routine filter().

$$B(f) = \frac{12200^2 \cdot 1.0125 \cdot f^3}{0.993 \cdot (f^2 + 20.6^2) \cdot (f^2 + 12200^2) \cdot \sqrt{f^2 + 158.5^2}} \qquad (4.3)$$

4.3 Bandpass filtered pink noise

Currently available de facto -standard test signals are band limited pink noise, which can be filtered by Butterworth characteristics. Filters were designed using Matlab-routine Butter(). For 6 dB/Oct. slope first order, for 12 dB/Oct. slope second order, and for 18 dB/Oct. slope third order Butterworth filter coefficients were calculated. Using previously explained pink noise the signals were filtered using corresponding filter coefficients.

4.3.1 Signal 1

The commercially available test signal 1 (see figure 5) is filtered with:

Highpass: second order Butterworth with corner frequency of 700 Hz
Lowpass: first order Butterworth with corner frequency of 700 Hz

4.3.2 Signal 2

Test signal 2 (see figure 6) is filtered with:

Highpass: first order Butterworth with corner frequency of 250 Hz
Lowpass: first order Butterworth with corner frequency of 500 Hz

4.3.3 Signal 3

The commercially available test signal 3 (see figure 7) is filtered with:

Highpass: third order Butterworth with corner frequency of 2000 Hz
Lowpass: third order Butterworth with corner frequency of 500 Hz

This signal was developed to take into account several factors [6], namely:

- Avoid the low frequency variations between rooms occurring below the Schoeder frequency (approximately 500 Hz in domestic rooms)
- Minimise the position dependant effects in the sound field at higher frequencies
- Provide a sufficiently broad frequency range signal to be representative of the

loudspeakers output

4.4 Psychoacoustically shaped test signals

Pink noise and B-weighted pink noise both provide approximations to perceptual shaping caused by the auditory system. However, these are coarse approximations and do not take into account the fine detail of the auditory systems characteristics, nor its level dependence. In an effort to define signals that can provide equal perceptual weight to all frequency bands for a level alignment task, two categories of signal have been considered:

- Uniform excitation noise (UEN) spectra
- Constant specific loudness (CSL) spectra

The uniform excitation noise is employed in auditory research as a means of providing uniform stimulation of the inner ear at an excitation level, by providing constant intensity in each critical band (Signal 7, Figure 8). This noise has been specified by Zwicker [16], but does not consider the effects of the outer and middle ear transforms to the spectrum.

The constant specific loudness signals have been developed by the authors in an attempt to provide uniform, frequency independent stimulation, at a loudness level. Based upon the models of Zwicker and Moore, it is not know which of these models is best suited to loudspeaker level alignment in small rooms. Bearing this in mind, three such signals were developed consisting

- Constant specific loudness according to the Zwicker diffuse field (Signal 4) model (Figure 9), [14].
- Constant specific loudness according to the Zwicker free field (Signal 5) model (Figure 10), [14].
- Constant specific loudness according to the Moore diffuse field (Signal 6) model (Figure 11), [10].

These specifications were employed in conjunction with optimisation routines to generate the signals.

The Optimization Toolbox for Matlab was used to find the third-octave input levels with which the loudness calculation program would give as constant specific loudness as possible around the desired level. The leastsq() function solves non-linear least squares optimisation problems and can be stated in matrix terms as

$$\min_X \sum F(X).*F(X) \qquad (4.4)$$

where the MATLAB notation “.*” indicates element-wise multiplication. The function F(X) was determined as difference of target level and specific loudness spectrum given by applied model.

$$F(X) = W(n) * \left[T - L(n)\right] \qquad (4.5)$$

where T is the target level, L(n) is vector containing specific loudness and W(n) for weighting the error function.

The optimisation was started with initial 1/3-oct levels estimate and terminated with 0.01 tolerance of function result and variable fluctuation.

The small variation visible at low frequencies for these signals is not considered to be significant within the scope of the subjective task [15] as these occur below the cut-off frequency of the reproduction loudspeakers.

Having found the desired third-octave levels, the same technique with interpolation and block-IFFT was used, as for the generation of the pink noise signal.

5 Summary

In this paper the subjective and objective alignment of multichannel sound systems has been considered. Having reviewed earlier work in this field, nine signals have been implemented to represent the current defacto calibration signals. Additionally, as study of loudness models has lead to the development of specific psychoacoustically shaped signals. The detailed motivations and methods for the generation of these signals is considered and presented employing both linear SPL and loudness metrics.
This work forms the basis for subjective tests on multichannel level alignment to be considered in the next paper in this series (part II) [15].

6 Acknowledgements

The authors would like to thank the funding bodies for supporting the Eureka Medusa project, including Tekes (Technology Development Centre of Finland) and Danish DTI. All members of the Medusa project are thanked for their comments and discussion throughout the project so far.

We would like to extend our thanks to Brian Moore for his assistance in interpreting the loudness model and to Tom Holman for his comments on calibration signals.

References

[1] Aarts R. M., Calculation of the Loudness of Loudspeakers during Listening Tests, Journal of the Audio Engineering Society, Vol. 39, pp.27-38, January/February 1991.

[2] Aarts R. M., Comparison of Some Loudness Measures for Loudspeaker Listening Tests, Journal of the Audio Engineering Society, Vol. 40, pp.142-146, March 1992.

[3] Aarts R. M., On the design and psychophysical assessment of loudspeaker system, Proefschrift Technische Universiteit Delft, 1995.

[4] Bech, S., Calibration of relative level differences of a domestic multichannel sound reproduction system, J. Audio Eng. Soc., vol. 46, pp 304 – 313, April 1998.

[5] Cardozo B. L, Adjusting the method of adjustment: SD vs DL, J. Acoustical Society of America, 37(5), May 1965.

[6] Holman T., personal communication, May 1998.

[7] ISO Rec. R. 532, “Method for calculating Loudness Level, Method B”, International Organization for Standardization, Geneva, Switzerland, 1966.

[8] ITU-R Recommendation BS 775-1, Multichannel stereophonic sound system with and without accompanying picture, Geneva, 1994.

[9] Moore B. C. J., An Introduction to the Psychology of Hearing, Academic Press, 1997.

[10] Moore B. C. J., Glasberg B. R., A revision of Zwicker’s Loudness Model, Acustica, Vol.82, pp. 335-345, 1996.

[11] Moore B. C. J., Glasberg B. R., and Baer T., A model for the prediction of thresholds, loudness, and partial loudness, J. Audio Eng. Soc., vol. 45, pp 224 – 239, April 1997.

[12] Oppenheim A. V., Scafer R. W., Discrete-Time Signal Processing, Prentice-Hall International, Inc., 1989.

[13] Parks T. W., Burrus C.S., Digital Filter Design, John Wiley & Sons, 1987.

[14] Paulus E., Zwicker E., Programme zur automatischen Bestimmung der Lautheit aus Terzpegeln oder Frequenzgruppenpegeln, Acustica, Vol. 27. pp. 253-266, 1972.

[15] Zacharov, N., Bech, S., and Suokuisma, P., Multichannel level alignment, part II: The influence of signals and loudspeaker placement, presented at the 105th Convention of the Audio Engineering Society, September 1998.

[16] Zwicker E., Fastl H., Psychoacoustics, Facts and Models, Springer-Verlag, 1990.

Figures

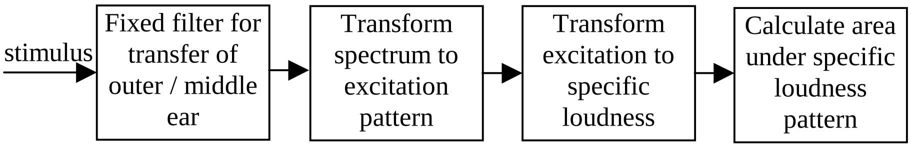

Figure 1 Block diagram of Moore's model for calculating loudness [10]

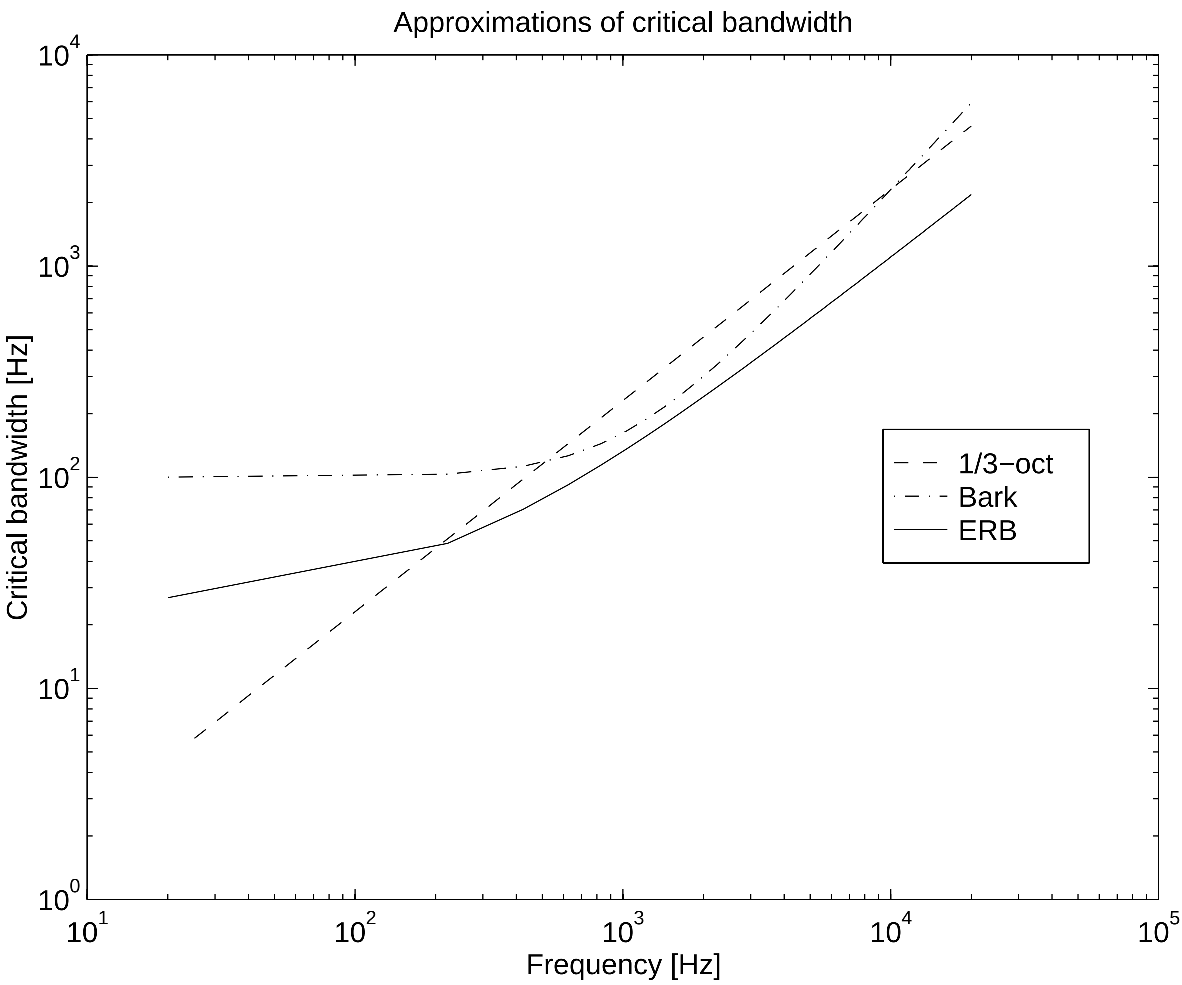

Figure 2 Most common critical bandwidth scales, adapted from Moore [9]

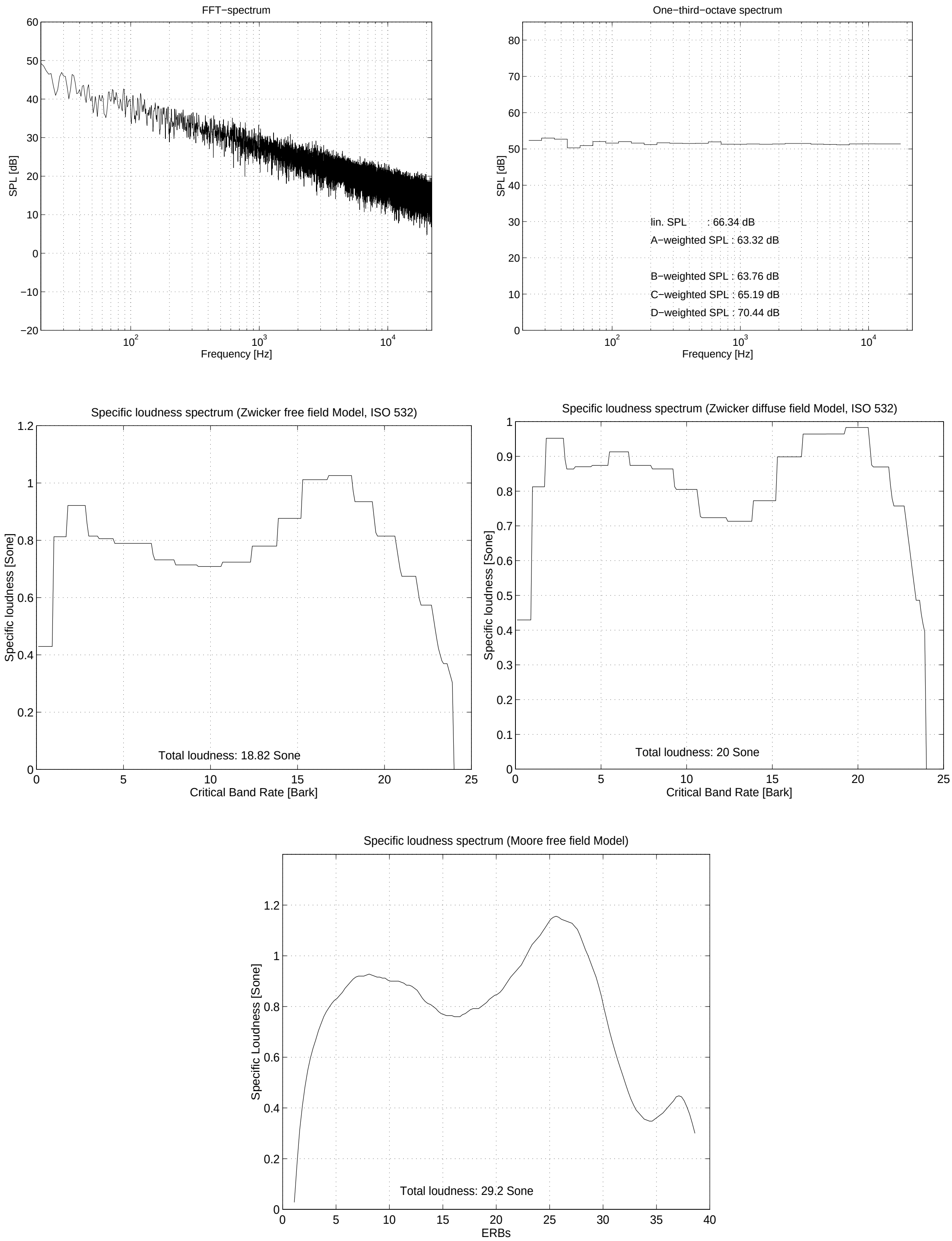

Figure 3 Objective spectra of test signal 8.

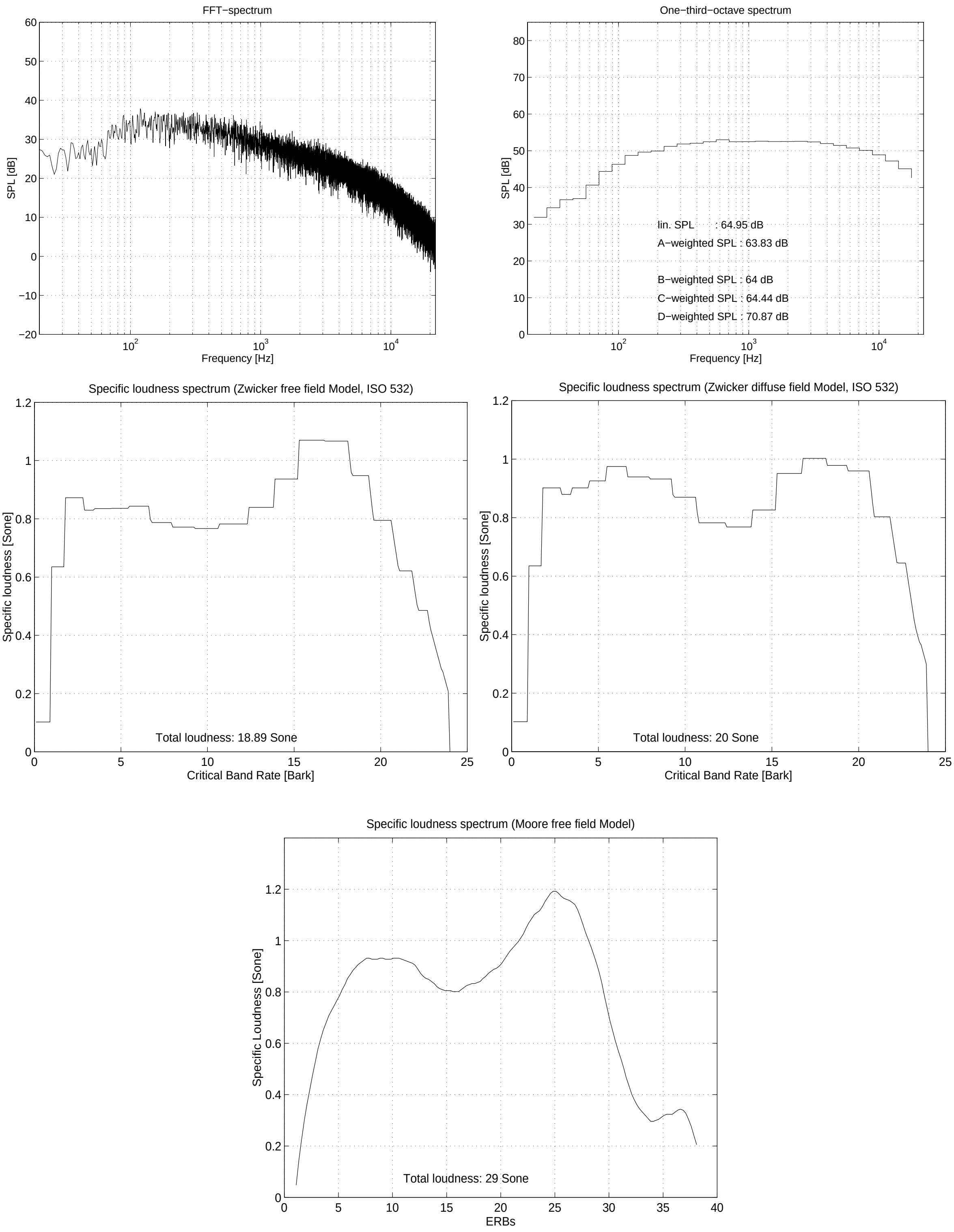

Figure 4 Objective spectra of test signal 9.

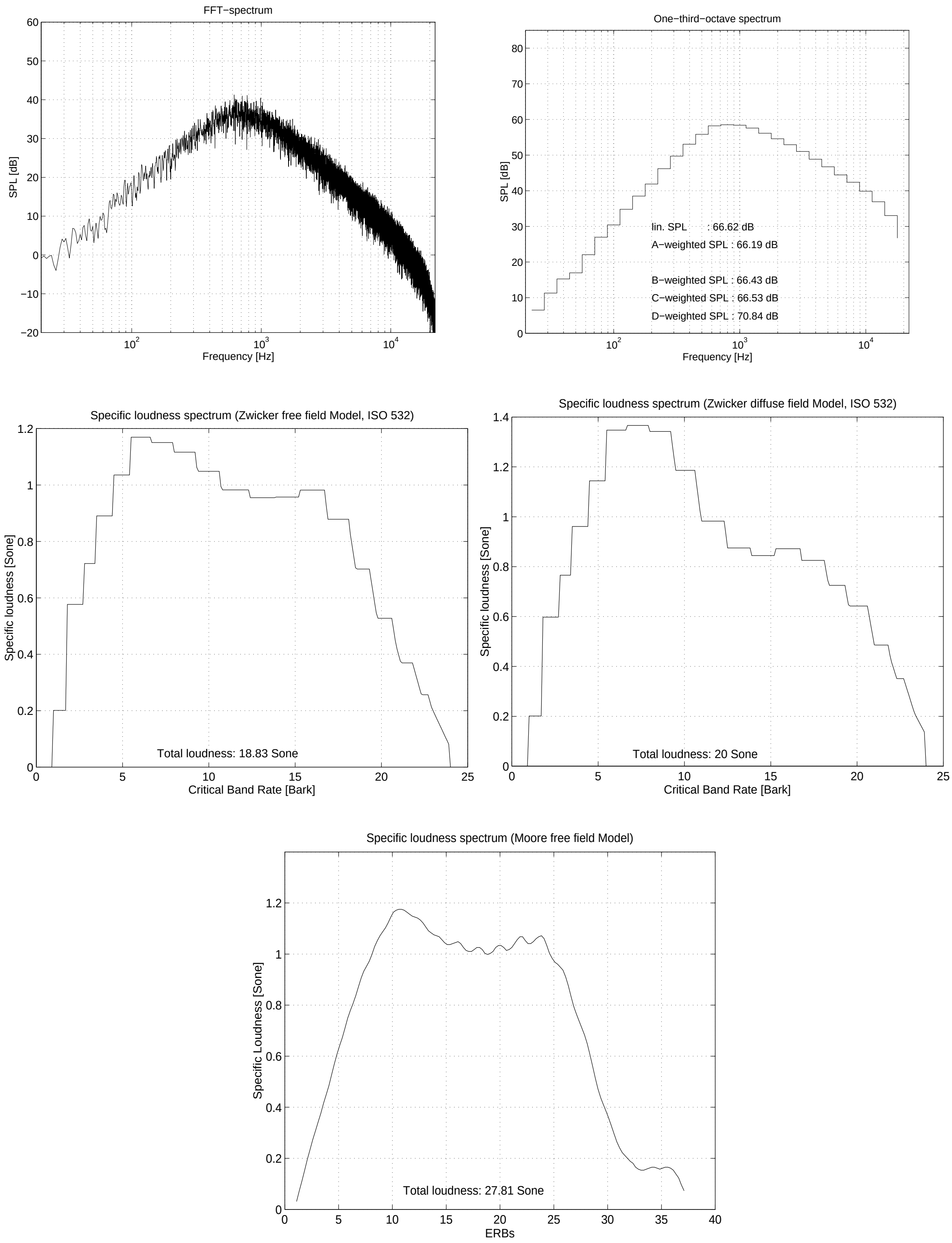

Figure 5 Objective spectra of Signal 1.

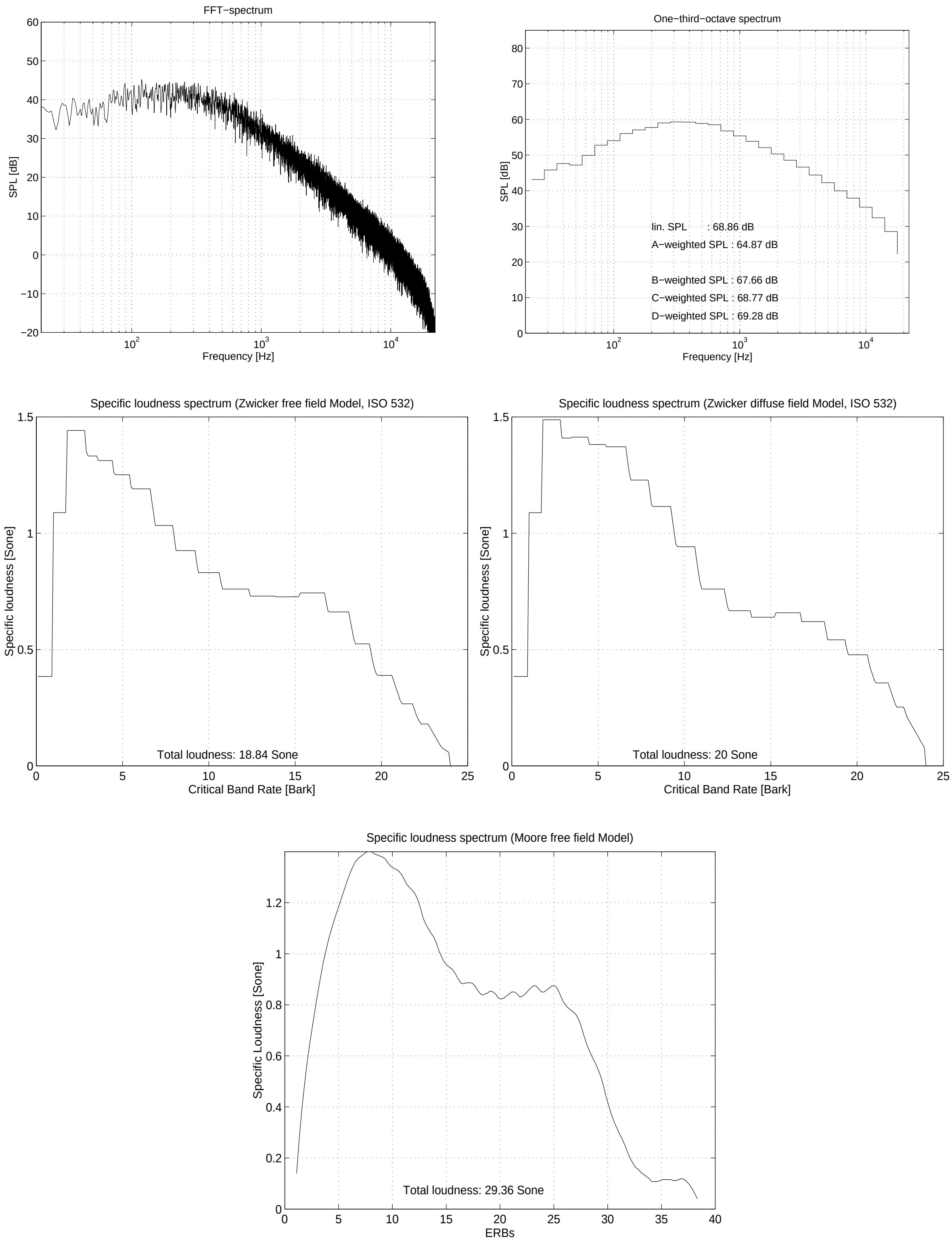

Figure 6 Objective spectra of Signal 2.

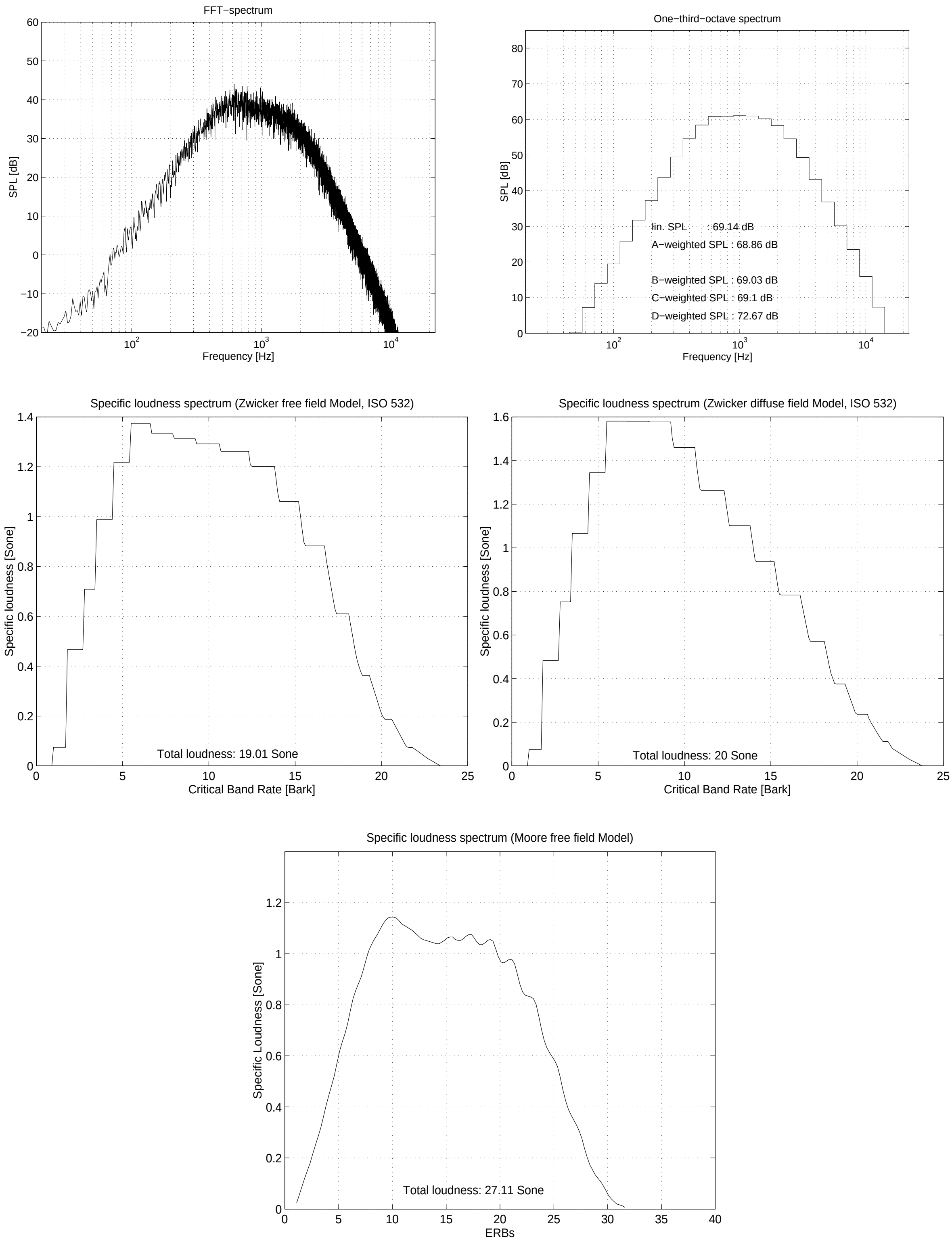

Figure 7 Objective spectra of Signal 3.

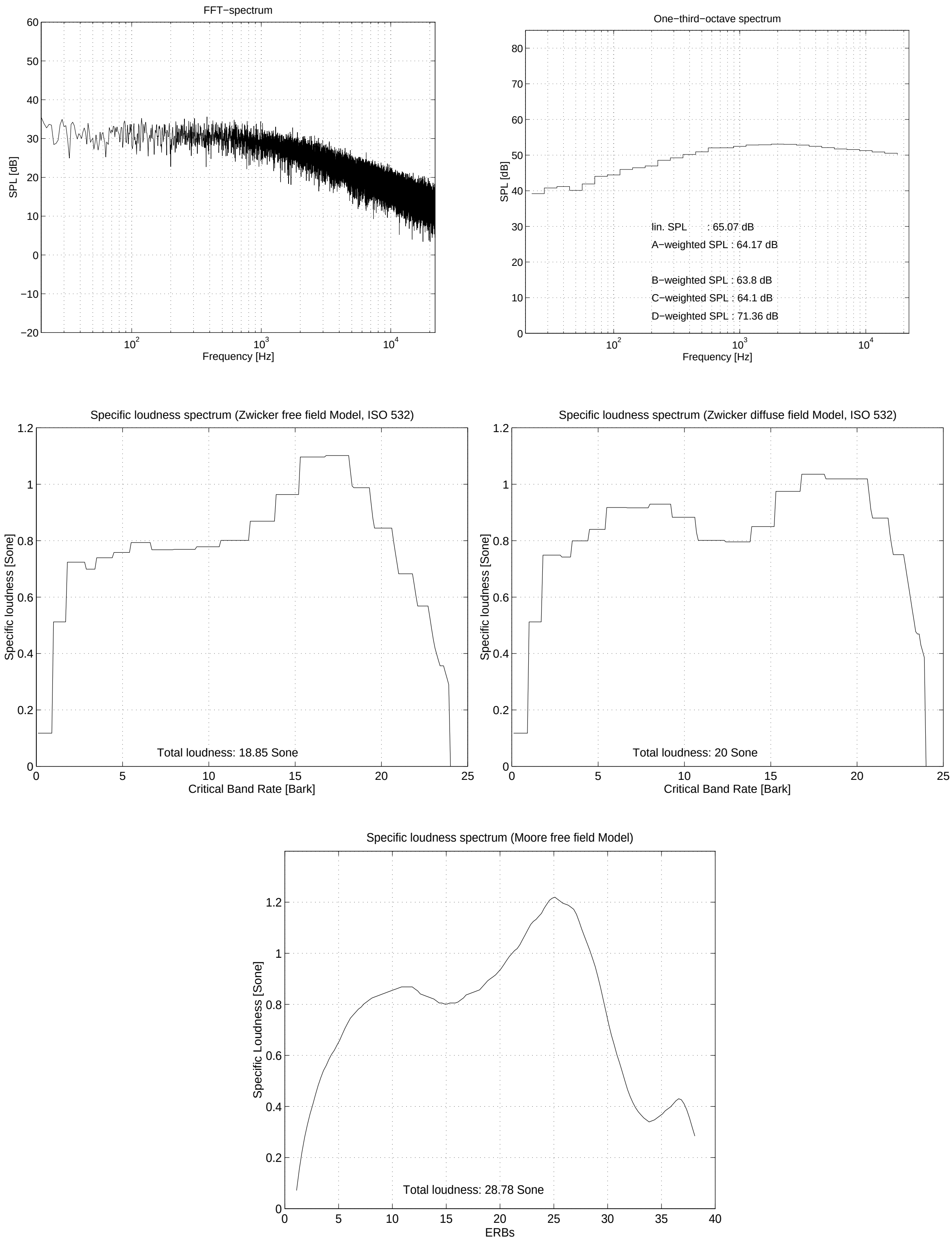

Figure 8 Objective spectra of test signal 7.

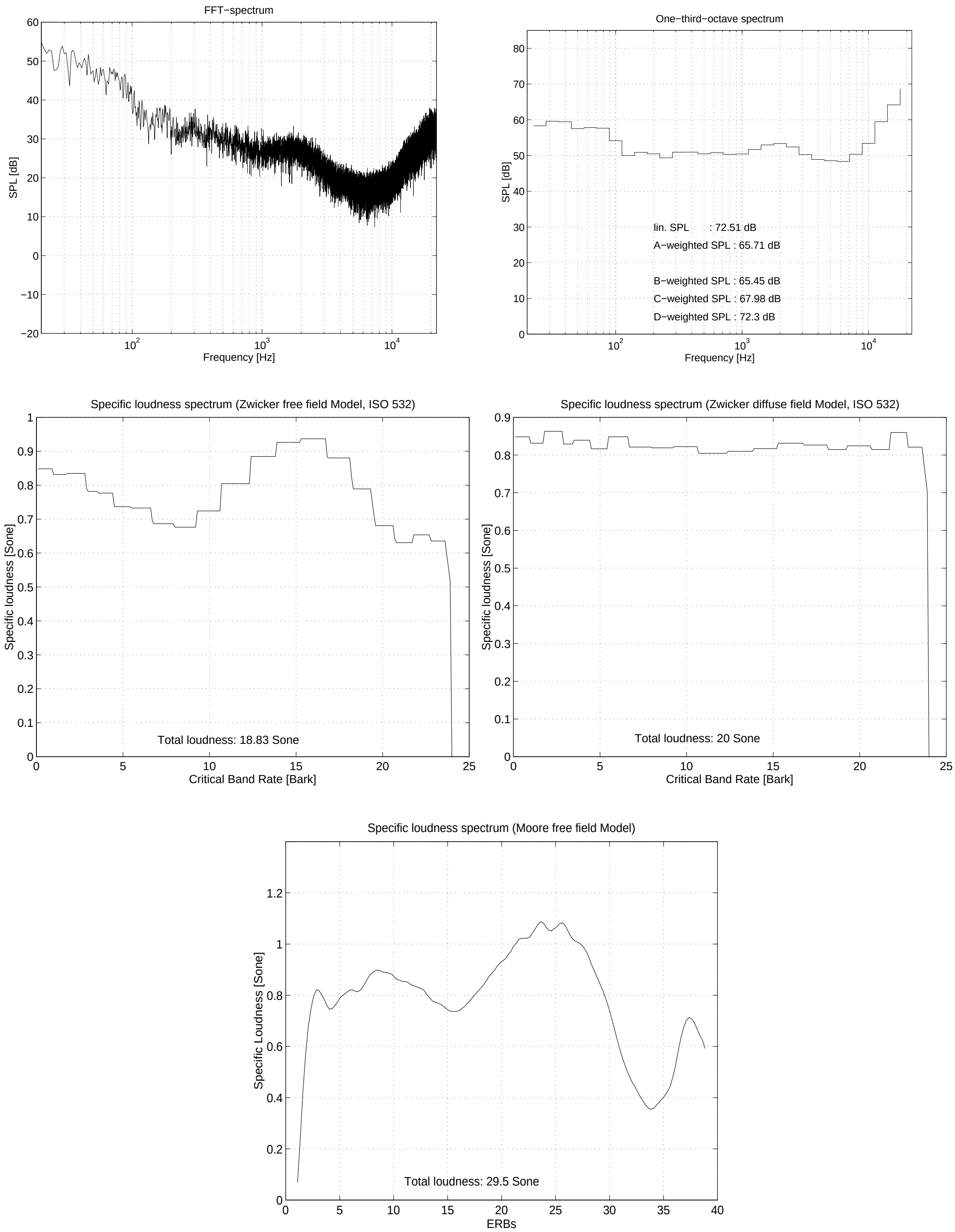

Figure 9 Objective spectra of test signal 4.

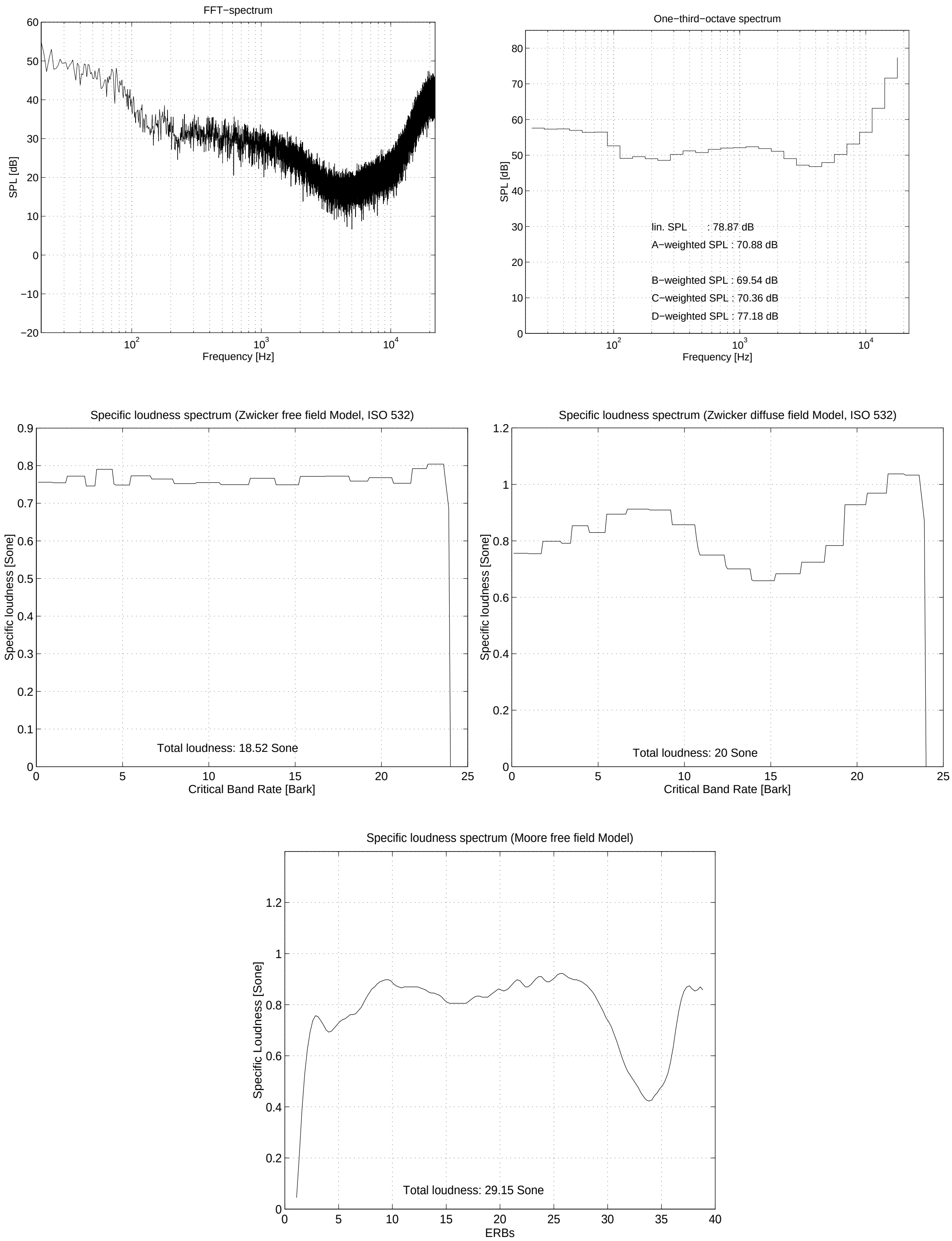

Figure 10 Objective spectra of test signal 5.

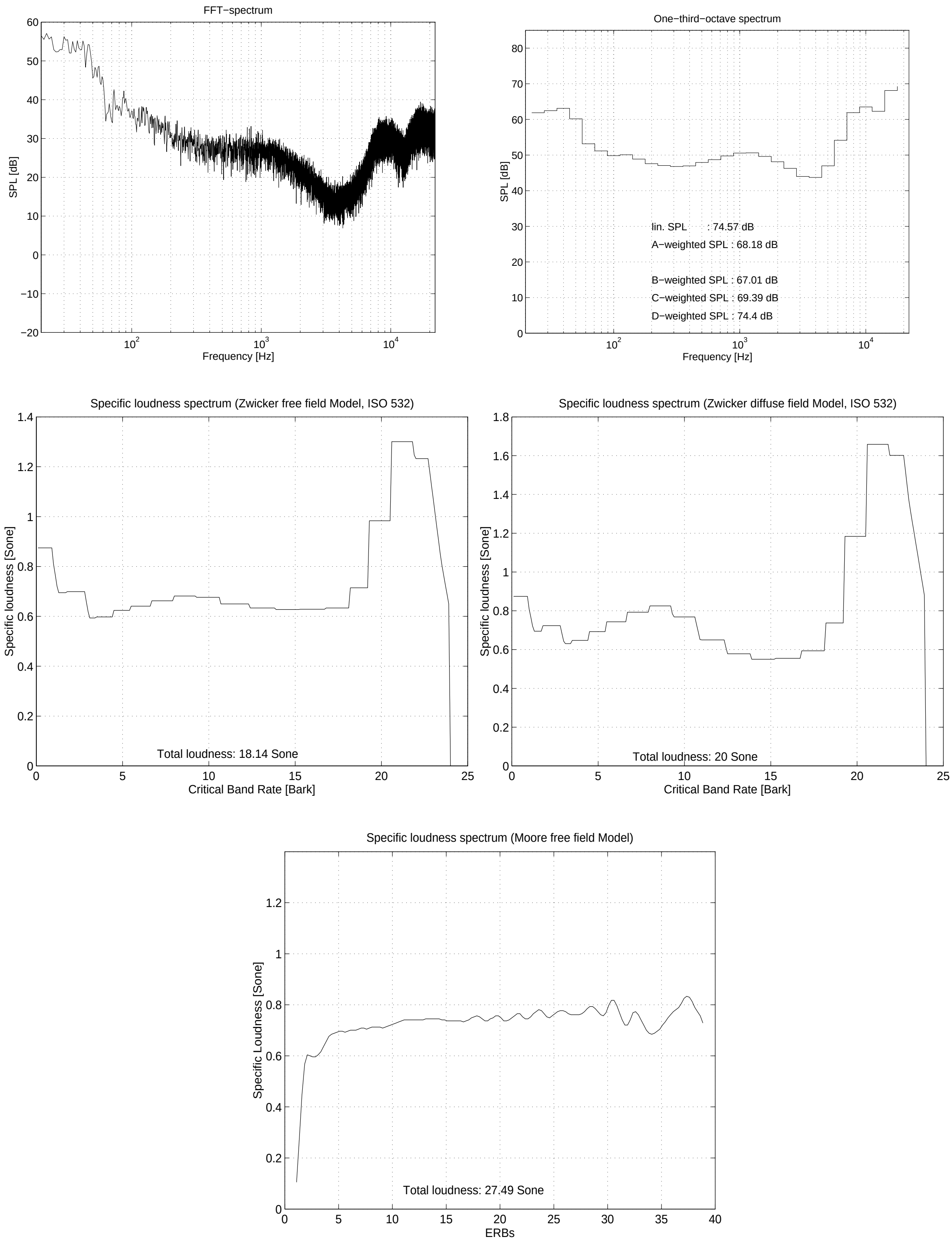

Figure 11 Objective spectra of test signal 6.

APPENDIX 1

Table 1 Characteristics and numbering of the generated signals

Signal name	**High pass filter characteristics**	**Low pass filter characteristics**	**Comments**
	Hz , dB/Oct.	**Hz , dB/Oct.**	
1.	700, 12	700, 6	Commercially available signal
2.	250, 6	500, 6	A signal
3.	500, 18	2k, 18	Commercially available signal
4.			Zwicker constant specific loudness according to ISO 532 (diffuse field)
5.			Zwicker constant specific loudness according to ISO 532 (free field)
6.			Constant specific loudness according to Moore
7.			Uniform excitation noise according to Zwicker
8.			Pink noise
9.			B-weighted pink noise

APPENDIX 2

Table 2 Bark-scale, corner and centre frequencies and bandwidth in Hz.

Bark	fl,fu	fc	Δf	Bark	fl,fu	fc	Δf
0	0			**12**	1720		
0.5		50	100	**12.5**		1850	280
1	100			**13**	2000		
1.5		150	100	**13.5**		2150	320
2	200			**14**	2320		
2.5		250	100	**14.5**		2500	380
3	300			**15**	2700		
3.5		350	100	**15.5**		2900	450
4	400			**16**	3150		
4.5		450	110	**16.5**		3400	550
5	510			**17**	3700		
5.5		570	120	**17.5**		4000	700
6	630			**18**	4400		
6.5		700	140	**18.5**		4800	900
7	770			**19**	5300		
7.5		840	150	**19.5**		5800	1100
8	920			**20**	6400		
8.5		1000	160	**20.5**		7000	1300
9	1080			**21**	7700		
9.5		1170	190	**21.5**		8500	1800
10	1270			**22**	9500		
10.5		1370	210	**22.5**		10500	2500
11	1480			**23**	12000		
11.5		1600	240	**23.5**		13500	3500
12	1720			**24**	15500		
12.5		1850	280				

Table 3 ERB-scale, lower, centre and upper corner frequencies and bandwidth in Hz.

ERB	fl	fc	fu	Δf	ERB	fl	fc	fu	Δf
1	13	26	40	27	**22**	2094	2222	2358	264
2	40	55	71	31	**23**	2358	2501	2653	294
3	71	87	105	34	**24**	2653	2812	2980	328
4	105	123	143	38	**25**	2980	3158	3346	365
5	143	163	185	42	**26**	3346	3544	3752	407
6	185	208	232	47	**27**	3752	3973	4205	453
7	232	258	285	52	**28**	4205	4451	4710	505
8	285	313	343	58	**29**	4710	4984	5272	562
9	343	375	408	65	**30**	5272	5577	5898	626
10	408	444	481	73	**31**	5898	6237	6595	697
11	481	520	562	81	**32**	6595	6973	7372	777
12	562	605	652	90	**33**	7372	7793	8237	865
13	652	700	752	100	**34**	8237	8705	9200	963
14	752	806	863	112	**35**	9200	9722	10273	1073
15	863	924	988	124	**36**	10273	10854	11468	1195
16	988	1055	1126	138	**37**	11468	12116	12799	1331
17	1126	1201	1280	154	**38**	12799	13520	14282	1482
18	1280	1364	1452	172	**39**	14282	15085	15933	1651
19	1452	1545	1643	191	**40**	15933	16828	17772	1839
20	1643	1747	1857	213	**41**	17772	18769	19820	2048
21	1857	1972	2094	237	**42**	19820	20930	22102	2281