APPLICATION OF A NEW HEARING MODEL FOR DETERMINING THE SOUND QUALITY OF SOUND EVENTS

Klaus Genuit, Roland Sottek

HEAD acoustics GmbH
Kaiserstr. 100, 52134 Herzogenrath, Germany

1. INTRODUCTION

Regulations for determining noise levels are based on A-weighted SPL measurements performed with only one microphone. This method of measurement is usually specified when determining whether the ear could be physically damaged. Such a simple measurement procedure is not able to determine annoyance of sound events or sound quality in general.

A loudness measurement which takes account of the spectral and temporal structure of a sound event, better than the A-weighted SPL, has advantages. Nevertheless, it is not universally accepted because the loudness measurement does not allow a complete judgment of complex sound events comparable to human hearing. In addition to loudness, there are other psychoacoustic parameters such as sharpness, roughness and fluctuation strength.

For some years investigations with binaural measurement and analysis techniques have shown new possibilities for the objective determination of sound quality. By using Artificial Head technology in conjunction with psychoacoustic evaluation algorithms - and taking into account binaural signal processing of human hearing - considerable progress has been made for determine the sound quality of sound events /1/, /2/.

This technique has been supplemented by a new calculation method for signals which are complex in terms of signal theory:

The analysis of a sound like door slamming is relatively complicated because its time duration is very short and its frequency spectrum very broad. In strong contrast to a normal FFT analyser, human hearing has a high resolution in the time and frequency domains. This is why an objective evaluation of a slamming door, corresponding to subjective classification values such as loudness, fluctuation strength and roughness, is insufficient, being based only on simple, quasi-stationary test signals. An additional calculation, based on three new types of signal processing, in addition to the Wavelet transformation, can be used to improve objective analysis of the sound of door slamming.

2. BINAURAL MEASUREMENT TECHNOLOGY

Human hearing differs in many respects from conventional sound measuring systems. The outer ear is a directional filter which changes the sound pressure level at the ear drum by +15 to -30 dB, depending on frequency and direction of sound incidence. These filter properties of the outer ear are due to diffractions, reflections and resonances caused by the pinna and concha geometry. Human hearing has two paths - the left and right ears. This capability permits binaural signal processing and pattern recognition in conjunction with spatial hearing, selectivity and noise suppression.

The auditory impression is not only determined by the sound pressure level, but also by psychoacoustic properties. Depending on the time sequence of the signals or spectral distribution, there can be different subjective impressions due to pre-, post- and simultaneous-masking properties of the hearing mechanism.

Human hearing involves very complex signal processing but it has very short memory. By faithfully recording, digitally storing and reproducing a sound event, the comparative human ear equivalent judgement of different sounds becomes feasible and can be documented. The psychoacoustical properties of human hearing determine the subjective impression of sound events such as noise annoyance. Also, both sound pressure level and loudness are important. Standard methods of measuring loudness do not consider the binaural signal processing that occurs in human hearing. This explains the poor correlation between measured sound levels and the subjective impressions of noise. Since the evaluation of sound quality by large numbers of test subjects usually produce similar results, the sounds must have physically measurable properties which can be correlated to subjective impressions. This suggests there are properties that have not been identified and accounted for by conventional measurement methods. Objective acoustic measurement methods used up to the present time have not provided sufficiently useful algorithms for evaluation of individual sources in a complex sound field. An artificial head measurement system that emulates human hearing as the "receiver" has been introduced for the judgement and analysis of sounds. The artificial head measurement system in accordance to /3/ has transmission characteristics that can be calibrated. They are comparable to human hearing and can be configured to give results compatible with conventional measurements. The binaural measurement method can be used in all fields where acoustic emissions (acoustic energy radiated by a source) and immissions (acoustic energy incident on a receiver) must be determined or where sound serves as an indicator of comfort, quality an safety.

Fig. 1 illustrates how recorded information from an artificial head can be evaluated with the help of digital signal processing to identify and eliminate annoyance. Binaural digital signal processing within a computer controlled measuring system allows input, storage and processing of an event and listening to selected segments which can be continuously repeated without audible artifacts. The measuring system displays both right and left ear signals in the time and frequency domains. The sound segment sample can then be directly manipulated with the result being displayed and reproduced by headphones in "real time." Euphony is evaluated as part of the analysis; that is calculations of roughness, sharpness, pitch and timbre as well as loudness, that take into account the ear's pre-, post- and simultaneous-masking properties.

3. COMBINED ANALYSIS OF SPECTRUM AND TIME STRUCTURE

As is well-known, human hearing not only performs spectral analysis, but also the evaluation of the envelope from the signal. The ability to distinguish between simultaneously occurring sounds can be modeled using bandpass division. The frequency resolution obtainable depends essentially on the bandwidth of the hearing-related filters. Modulations are primarily recognized from changes in specific loudness level over time, i.e. from the non-linear, distorted lowpass filtered envelope curve of individual bandpass signals.

The discrete Fourier transform is often applied for short-term spectral analysis of acoustic signals. Since in this technique, the same window is applied for the investigation of all relevant spectral components, the modulation remains constant over the whole frequency domain. Definition of the analysis window simultaneously defines time and frequency modulation. This kind of spectral analysis allows either high frequency modulation of low-frequency signals (long window), or high-resolution analysis of timestructure in the upper frequency range (short window), and this makes it less suitable for a perception-oriented signal description. Thus, selection of window length always represents a compromise between these various demands.

These considerations resulted in the development of the “variable” Fourier transform. The basis of this aurally-equivalent spectral analysis is efficient sub-band division (using filters with hardly any overlap). The spectralcomposition of the bandpass signals is subsequently investigated by applying analytic procedures of various length (adapted to human hearing). A combination of an appropriate filter bank makes it possible to approach the resolution of human hearing in a series of steps.

Other hearing-oriented methods aim at spectral representation as a quasi-continuous function of time (in contrast to block-by-block processing) with hearing-related frequency modulation. These generally more time-consuming techniques can be represented as bandpass division of the signal under investigation, where the impulse responses applied are of different length. The basic differences between these various approaches include which type of filter and which bandwidth as a function of the filter center frequency are selected. The Fourier T transform, as a short-time spectral analysis with an exponentially decreasing weighting factor of proceeding values, has been known for some time (e.g. /4/ S. 141 ff###, /5/, /6/). The essential advantage of a “filter bank principle” is that time constants, analogous to frequency selectivity in human hearing can be individually selected for individual analysis filters.

In the case of continuous Wavelet transform a time signal is also analyzed using bandpass filters of various width. The Wavelet transform of time-discrete signals can be interpreted equally well as a special form of sub-band division. It may also be understood as octave band filtering with perfectly reconstituting filters. The fact that various earlier approaches to signal definition with variable frequency and time resolution can also be represented in terms of WT opens the way to a uniform definition and thus adds to the development of new ideas.

4. AUDITORY MODEL ACCORDING TO SOTTEK

Human hearing perceives slight differences in frequency and rapidly changing time structures simultaneously. Any mathematical analysis applied exclusively in the frequency and time domains is therefore unable to provide aurally-equivalent results. An essential feature of the auditory model developed by SOTTEK /7/ is calculation of excitation distribution over time. This is obtained as a function of two variables (frequency and time) from the curve of the sound pressure function and provides a basis for calculating psychoacoustic values. The model for calculating excitation distribution according to ZWICKER and FELDTKELLER /8/ is limited to steady signals, since only spectral data are evaluated. Furthermore, the phase of individual vibrational components is neglected, which results in an inadequate simulation of the excitation due to several superimposed sound components: The spectrum of a sinusoidal, amplitude-modulated tone is composed of three neighbouring spectral lines. Ignoring the phase angles, the excitation distributions due each of the three individual components would produce a temporal constant running of the total excitation, although an amplitude modulation has, of course, a characteristically pronounced time structure.

Modulated tones or noise signals result in an impression of roughness, depending on their frequency and the modulation frequency. In determining this psychoacoustic value the time structure of the excitation is of particular significance and must be taken into account. When various sound components are contained within a given critical band, their phase has a particularly visible effect, because in this case separate processing of the various components in different channels is not possible. If the phase of the sidebands of an amplitude modulation is rotated through 90°, this is only perceivable if the physical spectrum is contained within a single critical band. The time structure of a (longer) signal segment is determined by phase data. Any change to the phase relationship affects the amplitude response of the time signal. If the amplitude spectrum is determined at sufficiently short time intervals, "phase" becomes less significant. This is taken account of in short-time spectrum analysis, by dividing the signal into weighted blocks of defined length and then Fourier transforming the segments. This results in a meaningful representation of the amplitude distribution and of the time as a function of frequency. Admittedly, the equidistant frequency modulation resulting from this technique does not take into account the resolution of the human ear. The procedure developed by SOTTEK does take into account the time and frequency resolution of human hearing, i.e. enables appropriate processing of non-steady signals.

5. APPLICATION IN DOOR SLAM NOISE

It is hard to describe the door slam noise in terms of signal theory, because this signal is very short and is based on broad-band excitation. However, if the signal analysis does not make possible a sound evaluation according to human signal processing, an objective determination of the subjectively perceived sound quality of the door slam noise becomes more difficult. Investigations of four different door slam noises will be represented in the following. The investigations were made in an anechoic chamber with an Artificial Head Measurement System. The system was positioned 1,5 m off the left front door with its face towards the door and recorded the sound event faithfully to allow both a subjective and objective evaluation. Four different vehicles of different classes producing widely differing door slam noises were evaluated. Due to the great differences in sound quality the listeners determined a clear priority order in the subjective listening tests. The four doors (Door 1, Door 2, Door 3 and Door 4) are represented below in the order of improved sound quality.

Previous considerations showed that a simple FFT analysis is not suitable for this kind of analysis. Fig. 2 shows the relationship of specific loudness in dependence on time for the four doors. This representation and analysis are not suited completely to show the subjective impression adequately. In case of Door 1 the increased loudness at higher frequencies becomes evident which is - among other things - responsible for the negative judgment of the sound. Further classifications of the individual door noises are, however, not possible. It does not become apparent, particularly in the case of Door 3, that a clear echo by slamming the interior ventilating flap could be heard. The subjective evaluations showed that Door 4 was preferred due to strong levels in the low-frequency spectral range. This does not become obvious in Fig. 2. Here, it is disadvantageous that the curves of the same loudness according to ZWICKER /8/ are based on listening tests with stationary signals. A statement for short-time impulse-like wide-band sounds cannot be made easily.

Fig. 3 shows the result for the variable frequency resolution VFR (comparable to Wavelet). The temporal structures of the door slam noise are striking: the long decay in the frequency range of 800 Hz up to 3500 Hz in case of Door 1, the time structure due to the non-uniform contact of individual components in case of Door 2 as well as the discrete echo in case of Door 3. At the same time a high frequency resolution of up to 20 Hz is guaranteed, the strong boominess of Door 4 as compared to Door 3 becomes obvious. An adequate analysis and representation according to human signal processing is shown in Fig. 4 and is based on the hearing model for monaural signal processing developed by SOTTEK /7/. At the left hand-side the length is scaled according to the basilar membrane and at the right-hand side the corresponding frequency ranges.

Fig. 4 shows the relationship of excitation distribution and time. For higher frequencies the representation and clarity regarding spectral distribution and the respective temporal structures is similar to the VFR analysis represented in Fig. 3. In the frequency range below 500 Hz with the same good frequency resolution a significantly better reproduction of the temporal structures becomes apparent. Door 4 has the highest energy in the low-frequency spectral range when the door is slammed. This energy is produced at the same time as the entire door slam noise. In case of Door 3 the maximum in the low-frequency range exhibits a pronounced narrow-band characteristic and has a temporarily shifted structure.

6. SUMMARY

In many sound situations it is necessary to make faithful recordings using artificial head technology in order to consider selectivity of human hearing on the one hand, and to take into consideration psychoacoustic parameters for judging sound quality in case of phase and level differences on the other hand. The improved method to mathematically describe psychoacoustic parameters according to SOTTEK allows a more accurate objective determination of sound quality. While questions as to the physical damage of hearing can be answered by the A-weighted SPL, the evaluation of sound quality is possible only in conjunction with psychoacoustic parameters. Using the door slam noise as an example it was possible to present an optimum analysis comparable to subjective evaluation.

LITERATURE

/1/ K. Genuit;
Artificial Head Measurement Systems for subjective evaluation of sound quality
Sound & Vibration, March 1992, S. 18

/2/ K. Genuit;
Binaural Technique for the objective measurement of subjectively perceived sound quality
Noise Con 91, 14.-16. July 1991, Tarrytown, New York

/3/ K. Genuit;
A special calibratable Artificial-Head-Measurement-System for subjective and objective Classification of Noise
INTER-NOISE 21.-23.07.86, Cambridge Massachusetts, S. 1313-1318

/4/ Flanagan, J.L.;
Speech Analysis, Synthesis and Perception
Springer Verlag, Berlin, Heidelberg, New York 1972, 2. Auflage

/5/ Terhardt, E.;
Fourier Transformation of Time Signals
Conceptual Revision Acustica, Bd. 1, S. 155 - 182, 1979

/6/ Terhardt, E.;
Verfahren zur gehörbezogenen Frequenzanalyse
Fortschritte der Akustik, DAGA´85, DPG GmbH, Bad Honnef, S. 811 - 814, 1985

/7/ R. Sottek;
Modelle zur Signalverarbeitung im menschlichen Gehör
Dissertation RWTH-Aachen 1992, Verlag M. Wehle

/8/ Zwicker, E; Feldtkeller, R.;
Das Ohr als Nachrichtenempfänger
S. Hirzel Verlag, Stuttgart, 2. Auflage 1967

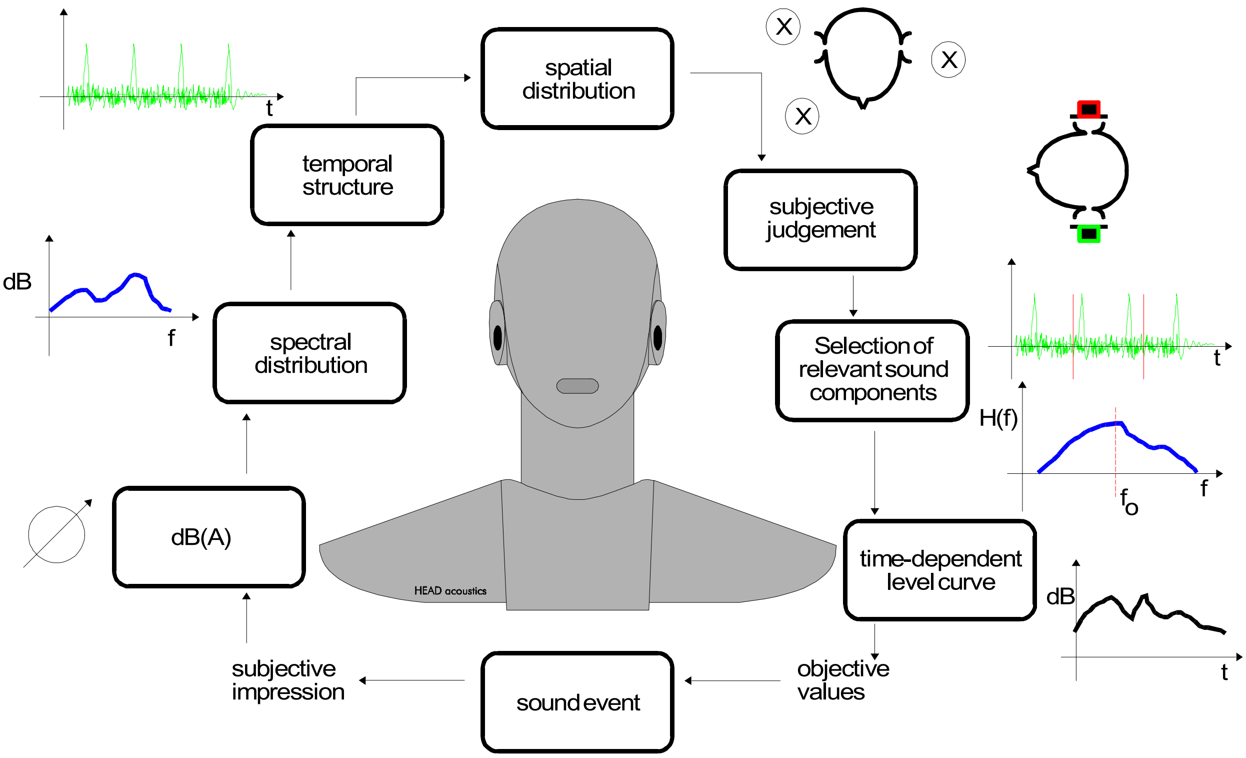

Fig. 1: Subjective and objective diagnosis, analysis and classification of sound.

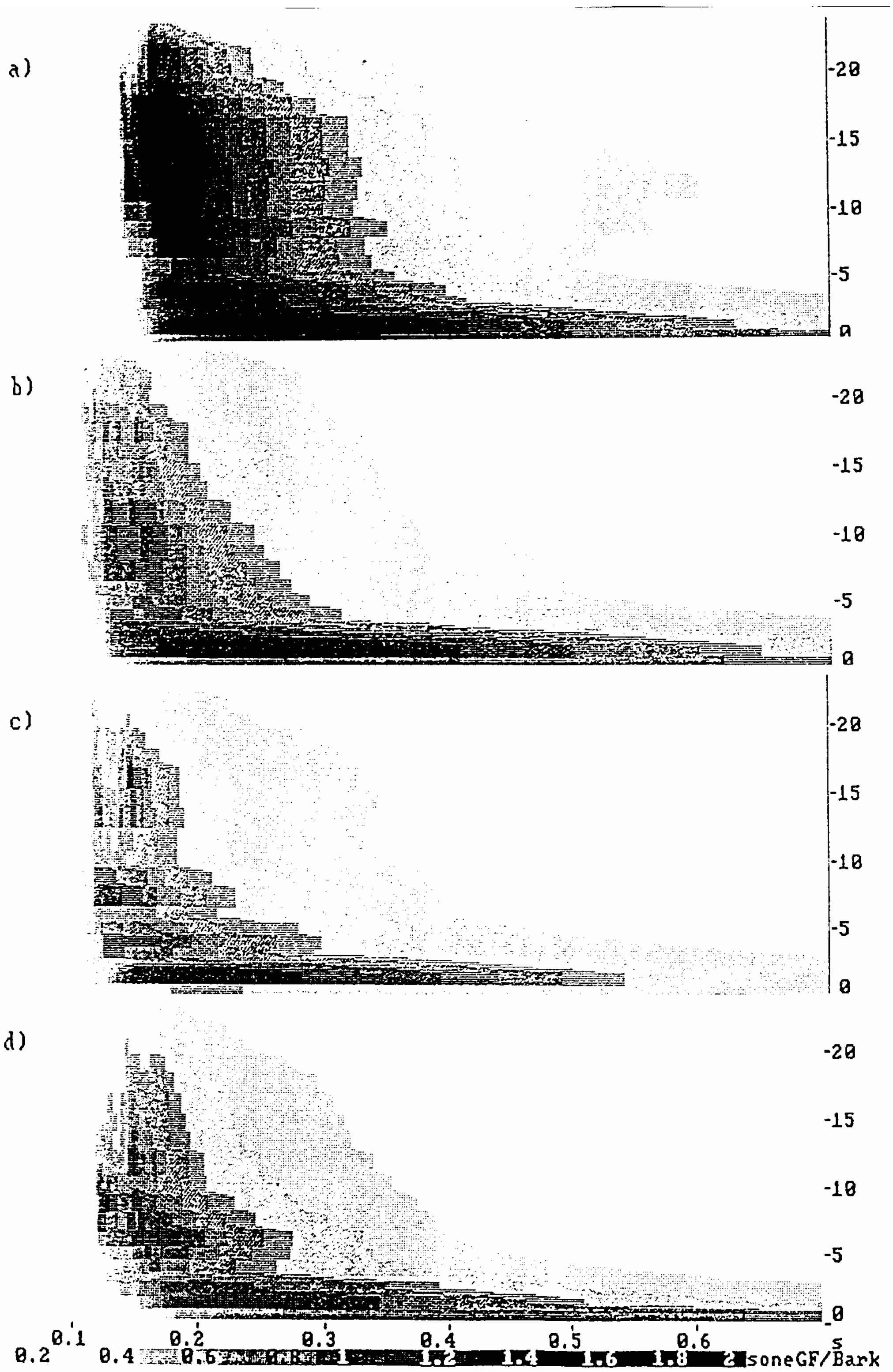

Fig. 2: Specific loudness analysis
a) Door 1 b) Door 2 c) Door 3 d) Door 4

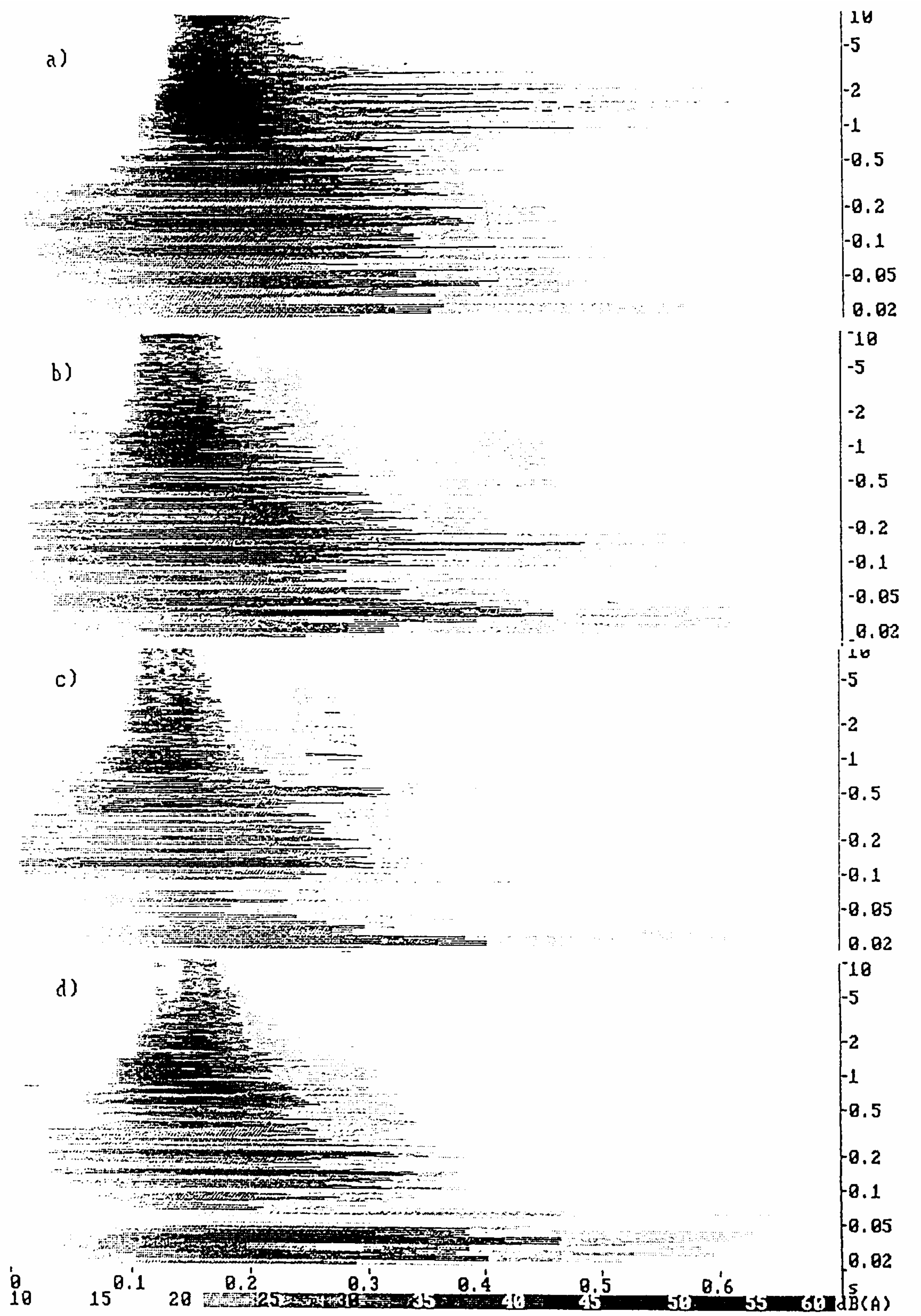

Fig. 3: Variable Frequency Resolution (VFR)
a) Door 1 b) Door 2 c) Door 3 d) Door 4

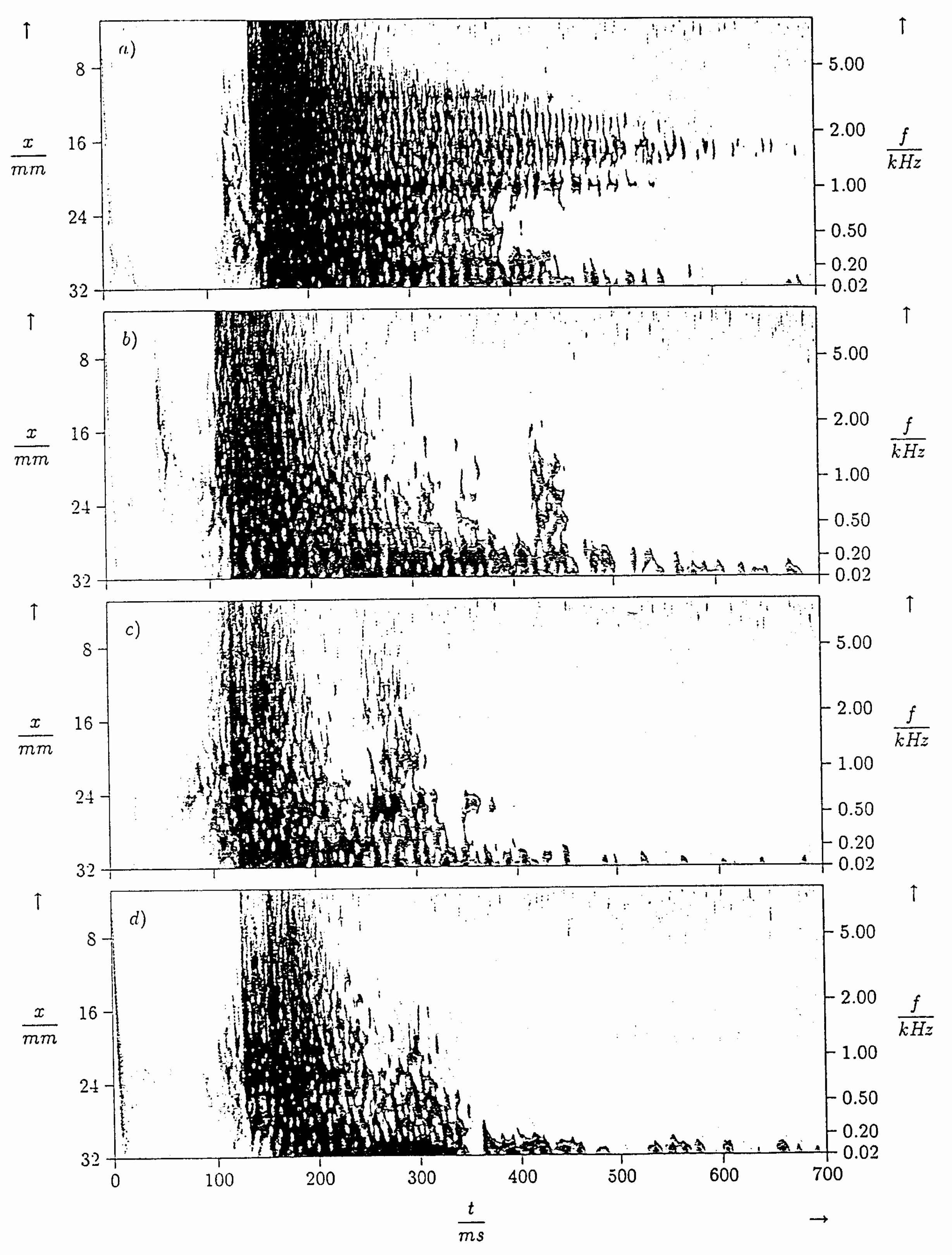

Fig. 4: Analysis with SOTTEK inner ear model
a) Door 1 b) Door 2 c) Door 3 d) Door 4