TEST SIGNALS FOR VALIDATING SOUND QUALITY MEASUREMENT INSTRUMENTATION

D Simmons Acoustics Team, NPL, Hampton Road, Teddington, Middlesex, UK

1 INTRODUCTION

Psychoacoustic metrics, largely developed by Zwicker[1] are designed for the assessment of perceived characteristics of sound. They are increasingly used to complement subjective jury listening tests in the process of evaluation and refinement of sound quality (SQ) of consumer products or machines throughout their development cycle.

Although several of the commonly used metrics are defined and have units, few are standardized. In fact, it is only the metric for loudness that has been standardized. This has resulted in manufacturers of sound quality measurement systems adopting their own implementation for the calculation of these metrics.

The objective of this validation exercise is to assess the variation of calculated values of the SQ metrics, between the systems, for practical signals, typical of those used in their application area. This validation is achieved by applying a suite of audio test signals to commonly used SQ measurement systems and instruments.

This paper reports upon the preliminary results of this validation study. At the time of writing, the performance of three systems has been assessed using eight test signals. The number of test signals and measurement systems will increase as the study progresses.

2 SOUND QUALITY MEASUREMENT SYSTEMS

A SQ measurement system may consist of hardware for signal acquisition and software for analysis and reporting. A head and torso simulator (HATS) may be used to acquire binaural sound, as commonly used in the case of automobile interior noise analysis, or a standard measurement microphone for monaural acquisition. Calibrated audio data is acquired and stored digitally on the system with any synchronous tachometer derived data associated with the machine under evaluation. The audio analysis is normally post-processed using the measurement system's software. All the software assessed for this study has been installed as 'stand-alone' on a personal computer.

3 LOUDNESS AND SOUND QUALITY METRICS

Two methods for the calculation of loudness are specified in ISO 532[2]. The first method, '532(A)', also referred to as Stevens method, uses an approximate calculation and is only applicable to diffuse sound fields. The second method, '532(B)', is based on loudness calculation developed by Zwicker[1], and is applicable to both frontal and diffuse sound fields. This second method has been more widely adopted and has been implemented as a BASIC computer program[3].

Central to the psychoacoustics model of hearing developed by Zwicker[1] is the Bark frequency scale and the effects of masking. Masking is the process by which one sound is masked by the presence of another sound. The Bark scale is a scale that maps the audible frequency range into 24 equally weighted frequency bands, called critical bands that correspond to the frequency resolution of the human auditory system. The loudness within each critical band is calculated taking into consideration the effects of masking. This loudness frequency distribution is referred to as specific loudness. The overall loudness, normally reported in Sones, is calculated as the sum of the specific loudness values.

The calculation procedure specified in 532(B) allows determination of loudness for Free-field (frontal) or diffuse sound field conditions. These are referred to as GF and GD respectively. The difference between the two calculation procedures is the application of a frequency dependant weighting representing the transfer function to the human ear.

Sound that is steady or invariant with time is referred to as stationary. All calculation procedures specified in ISO 532 are intended for the determination of loudness of stationary sound. This calculation of stationary loudness takes into account only spectral masking[1]. Non-stationary sound, although based on stationary loudness calculation, also takes into account the effect of temporal masking[1]. Many of the commonly used metrics, including the metrics investigated in this paper are dependent to some extent upon the calculation of loudness.

3.1 Metric definitions

The four commonly used metrics that are being investigated in this paper are described in the subsections below.

3.1.1 Loudness

A continuous pure tone of frequency 1kHz, at frontal incidence to the listener, with a sound pressure level re 20μPa (SPL) of 40 dB is defined as having the perceived loudness of 1 Sone. A sound perceived to be twice as loud has a loudness of 2 Sones, and for pure tones, will have a corresponding SPL increase of 10 dB. Alternatively, the loudness level may be expressed in decibels in units of Phon, which is numerically equal to the SPL of a pure tone at 1 kHz, which produces the same sensation of loudness as the sound in question.

3.1.2 Sharpness

A narrow band noise centred on 1 kHz with a bandwidth lower than 150 Hz, and with a SPL of 60 dB is defined as having a sharpness of 1 Acum. Sharpness is derived from specific loudness with a frequency weighting to produce prominence at high frequency. Zwicker originally proposed a calculation of sharpness. However, the method subsequently modified by Aures[4] has been more widely adopted.

3.1.3 Roughness

The sensation of roughness[1] results from amplitude modulations of the sound at frequencies ranging from around 15 Hz to 300 Hz. At higher frequencies, the modulated signal becomes indistinguishable from a tone. The unit of roughness is Asper. A pure tone of 1 kHz at a SPL of 60 dB, that is modulated at a frequency of 70 Hz with a 100% modulation depth defines a roughness of 1 Asper.

Deleted:

3.1.4 Fluctuation strength

Fluctuation strength[1] is the sensation associated with amplitude modulation at frequencies below around 15 Hz. The sensation reaches a maximum at a frequency of 4 Hz. The unit for fluctuation strength is the Vacil and is determined by a similar method to roughness, but at a modulation frequency of 4 Hz instead of 70 Hz.

4 TEST SIGNALS FOR INSTRUMENT VALIDATION

The suite of test signals used in this initial study consist of a selection of artificial digitally generated sounds and real recordings of products and machines. They represent a range of products and machines that vary in characteristics such as overall level, temporal variability, spectral shape and tonality. At the time of writing of this paper, highly impulsive sounds and one-shot sounds have not yet been incorporated into this study.

Artificially generated signals that are produced using procedures specified below, are included for two reasons. Firstly, to allow the signals to be readily reproduced, and secondly to allow the sound characteristics to be controlled.

The test signals used in this study were digitally stored as 16 Bit, 44.1 Hz sampling rate, mono PCM Wav file format with a duration of approximately 10 seconds. The test signal files are imported into each system, together with their associated calibration factors, in order to convert the signal into sound pressure data. The RMS SPL of a signal will therefore depend upon its assigned calibration factor.

Deleted:

4.1 Artificial basic signals

The signal referred to in the following sections as ‘1k pure tone at 94 dB’, represents a continuous sine wave having a sound pressure level (SPL) of 94 dB. This signal type and level combination is commonly used for system calibration purposes. The signal referred to in the following sections as ‘1k pure tone at 40 dB’, represents the sine wave at an SPL of 40 dB. This reference signal should produce a loudness of exactly 1 Sone (GF). White noise and pink noise signals are also included to represent ideal broadband noise sources.

4.2 Artificial machine signals

Two simulated machine signals are considered for this paper. Each signal may be produced and specified with ease and each signal is representative of a category of machine sound.

The first signal, ‘Buzz’, is constructed using the following procedure. An amplitude envelope, having initial amplitude of 100% and linear decay to final amplitude of 0%, is applied to a 20 ms duration white noise signal. A 6^{th} order Bessel band pass filter with low cut-off 1 kHz and high cut-off 4 kHz is then applied. The signal is then looped producing a pulse frequency of 50 Hz. Subjectively, the sound produced is perceived to be similar to the noise of small motor powered products such as an electric razor. The signal has comparatively little low frequency content and exhibits no temporal variability below 50 Hz.

The second artificial machine, ‘Squeak’, is a representation of a sound component rather than a complete machine. This artificial sound is produced by applying a 6^{th} order Bessel band pass filter, with a low cut-off of 5 kHz, and high cut-off of 5.5 kHz, to the ‘Buzz’ signal specified. A sinusoidal amplitude modulation of depth 50% and frequency 10 Hz is then applied to the signal. The subjective impression of the sound is that of a periodic squeak of a rotating component of a large machine. This signal has little low frequency content and exhibits a prominent periodic temporal variability.

4.3 Real machine signals

The majority of the signals that will eventually be used in this study will be of recordings of real machines. For the purposes of this paper only two such recordings are considered. These two machines, described below, are considered to be representative of two categories of product sound.

The ‘Lawn mower’ signal is an in-situ recording of a petrol-driven lawn mower, situated on grass, operating under normal conditions but in a static position. It is typical of many machines characterised by a dominant periodic pulse resulting from engine 'firing', and to a lesser extent, noise and temporal variation resulting from stochastic processes. This signal type is subjectively typical of internal combustion engine powered outdoor machines and products such as portable power generators and petrol powered mowers. Subjectively, the sound of the machine may be described as non-stationary due to its low engine revs.

The 'Vacuum cleaner' signal was also recorded in-situ. This signal is typical of many products and machines whose primary function is to move air such as vacuum cleaners, hair driers and also may include others with a less tonal component such as fan heaters and personal computer fan noise. The noise is typically characterized by a component of broadband noise and a tonal component relating to the fan rotation speed. This recording is of a vacuum cleaner with a particularly prominent tone. Subjectively, this machine sound may be described as stationary, with a prominent tone.

Deleted: ,

For both recordings, a directional recording microphone was used. A description of these signals, along with their assigned RMS SPL and signal category, are summarized in Table 1 below. All the in-situ recordings of real machines are essentially absent of background noise and reverberation.

Table 1 Test signals description

Signal level		
Name	Type	RMS SPL (dB)
1 kHz pure tone at 94	Artificial basic	94.0
1 kHz pure tone at 40	Artificial basic	40.0
White noise	Artificial basic	82.8
Pink noise	Artificial basic	81.5
Buzz	Artificial machine	38.9
Squeak	Artificial machine	50.0
Vacuum cleaner	Real machine	82.3
Lawn mower	Real machine	75.6

Deleted: ¶

5 RESULTS

All three SQ measurement systems provided capability for loudness calculation according to ISO 532(B), (Zwicker) loudness method. For the first system, the loudness calculation had no user specified parameters and the accompanying documentation does not specify whether filters or FFT frequency analysis method is used to calculate the Bark band levels. For the second system the choice of FFT or filter banks is provided. Sixth order filters are selected for all the loudness calculations using this system. The associated documentation for this system states that both methods of implementation fulfill the requirements of 532(B). This system also provides an alternative FFT-based loudness calculation that addresses limitations of the 532(B) procedure. The third system does not offer a choice of FFT or filter based calculation, however, it does offer a choice of Bark band resolution. For all the loudness-based calculations using this system, a frequency resolution of one-eighth Bark was selected.

For all three systems, for all the loudness and sharpness calculations, the sound field was set to 'free-field' (GF loudness) and all calculations of sharpness, roughness and fluctuation strength use the settings specified in this section.

5.1 Stationary loudness

The results for variation in stationary loudness between systems are shown below in Table 2. The column header labeled 'Average' contains values of the arithmetic mean of the loudness values from each system. It can be observed in Table 2, that for the pure tone signals, the range between systems at 94 Hz is greater than that at 40 Hz. This suggests that any difference in loudness calculation may be proportional to loudness or level. For this reason the variation has been normalized to the ratio of the highest loudness value to the lowest loudness and shown in column 4 as a percentage.

Table 2 Stationary loudness (GF) variation between systems

Loudness (GF) (Sones)			
Signal	Average	Range	Variation %
1 kHz pure tone at 94	44.6	1.0	2%
1 kHz pure tone at 40	1.0	0.1	7%
White noise	45.4	2.1	5%
Pink noise	45.6	2.6	6%
Buzz	2.5	0.3	11%
Squeak	2.6	0.0	2%
Vacuum cleaner	53.1	1.7	3%
Lawn mower	31.7	1.7	5%

A preliminary investigation, involving applying test signals to the system that offered both FFT and filter based analysis, revealed that the calculated loudness value was dependent upon method. In the case of the white noise signal this difference is of the same order as the difference between the systems.

These results are broadly consistent with the results of an earlier comparison conducted by Fastl and Schmid[5] that demonstrated a variation of calculated loudness of up to 5% between different manufacturers systems.

5.2 Non-Stationary Loudness

Table 3, below, shows the results for non-stationary loudness. It should be noted that only two of the three systems provide calculation of non-stationary loudness. The first system calculates ISO 532(B) loudness multispectra, but the associated documentation does not specify if this analysis is applicable for non-stationary signals and is therefore omitted from this analysis of non-stationary loudness. For system three, an additional parameter of time resolution, was set to 10 ms. The arithmetic mean of the two results are reported in the 'Average' column, and the variation is calculated as in section 5.1 above.

Table 3 Non-stationary loudness variations between systems

Loudness Non-Stationary (Sones)			
Signal	Average	Range	Variation%
1 kHz pure tone at 94	42.5	0.2	0%
1 kHz pure tone at 40	1.0	0.0	1%
White noise	44.0	0.4	1%
Pink noise	44.1	0.6	1%
Buzz	2.2	0.0	0%
Squeak	2.1	0.1	6%
Vacuum cleaner	52.1	0.2	0%
Lawn Mower	29.3	0.3	1%

5.3 Sharpness

All three systems provide calculation of the sharpness metric however system number one did not specify the calculation method. Systems two and three provide the Aures method and a variation. The calculation for sharpness is dependent upon the settings for the underlying loudness calculation. The Aures method was selected for systems two and three.

Table 4 Sharpness variation between systems

Sharpness (Aures method) (Acum)		
Signal	Average	Range
1 kHz pure tone at 94	1.7	1.1
1 kHz pure tone at 40	0.9	0.1
White noise	4.6	3.4
Pink noise	3.5	2.6
Buzz	2.7	1.0
Squeak	4.7	1.6
Vacuum cleaner	3.7	2.9
Lawn Mower	2.6	1.8

5.4 Roughness & Fluctuation strength

The results for the roughness and fluctuation strength metrics are reported below in Tables 5 and 6.

Table 5 Roughness variation between systems

Roughness (Asper)		
Signal	Average	Range
1 kHz pure tone at 94	0.09	0.2
1 kHz pure tone at 40	0.85	2.5
White noise	3.96	5.9
Pink noise	3.76	7.0
Buzz	1.02	1.1
Squeak	2.25	5.4
Vacuum cleaner	2.08	2.7
Lawn Mower	4.06	3.7

Table 6 Fluctuation variation between systems

Fluctuation Strength (Vacil)		
Signal	Average	Range
1 kHz pure tone at 94	0.08	0.2
1 kHz pure tone at 40	0.00	0.0
White noise	0.34	1.0
Pink noise	0.38	0.8
Buzz	0.25	0.7
Squeak	0.24	0.7
Vacuum cleaner	0.32	0.7
Lawn Mower	0.61	0.9

There are no applicable user settings for system one. There are no user settings applicable for system number two with the exception of the frequency resolution for fluctuation strength, which is is set to 'half Bark'. For system three, the roughness and fluctuation strength are dependent upon the loudness settings and the multispectra time interval.

6 DISCUSSION AND CONCLUSION

All three systems produced different values for all the metrics. The degree of variation between the systems depended upon the characteristics of the input signal. A preliminary investigation involving application of the white noise signal to one of the systems at several RMS SPL settings indicated that the values of sharpness, roughness and fluctuation strength also depend upon its RMS SPL. Since this is the case, assessment of the effects of temporal and spectral characteristics upon the variation between systems may require these reported values to be normalized for signal loudness or RMS SPL.

All three systems calculated loudness according to ISO 532(B) however it is evident that each system implements the calculation in a different way. A system may provide a variety of method variations for calculation of a metric and provide user options for the calculation parameters. The user, with guidance of the associated system documentation and experience, selects the most appropriate settings for that signal and application. The metrics are often used in combination as input parameters to application specific compound metrics. The significance of any variation between systems for a given metric will depend upon how it is used in practice.

7 FURTHER WORK FOR THIS STUDY

The next stage of the project will be to extend the range of systems and to increase the range of test signals to include signal types encountered within the automotive industry as well as signals that are representative of more impulsive sound sources and signals of shorter duration. Additionally, tonality metrics may be included in the analysis.

8 ACKNOWLEDGEMENTS

The author gratefully acknowledges the support from Acsoft Ltd and Brüel & Kjær. NPL gratefully acknowledges the financial support of the UK Department of Trade and Industry (National Measurement System Directorate).

9 REFERENCES

1. E. Zwicker, H. Fastl, 'Psychoacoustics: Facts and Models' (1990).
2. ISO 532, 'Acoustics- Method for Calculating Loudness Level' (1975).
3. E. Zwicker et al, 'Program for calculating loudness according to DIN 45631 (ISO 532B)', Journal of the Acoustical Society of Japan Vol 12, part 1 (P39-42) (1991).
4. W. Aures, A model for Calculating the Sensory Euphony of Various Sounds Acustica 59, pp 130-141 (1985).
5. H. Fastl, W. Schmid, 'Comparison of loudness analysis systems', Proceedings of inter-noise 97, p981 (1997).

Deleted: (Delete close bracket??)

Inserted: (Delete close bracket??)