forum acusticum 2026
8-12 September · Graz, Austria

INVESTIGATION OF METHODS FOR ASSESSING THE AUDIBILITY OF WARNING SOUNDS IN TRAIN CABS: COMPARATIVE ANALYSIS OF PERCEPTUAL TESTING AND PREDICTIVE INDICATORS

F. Aubin[1], M. Toward[2], M. Lower[2], D. Lurcock [2], T. Heumann[3], M. Ripert[4] and C. Eichenlaub[5]

[1]*SNCF Voyageurs, 4 allée des Gémeaux, 72100 Le Mans, France*

[2]*ISVR Consulting, University of Southampton, Highfield,, Southampton, SO17 1BJ, UK*

[3] *Siemens Mobility, Duisburger Str. 145, 47829 Krefeld, Germany*

[4]*ALSTOM Transport SA, 48 Rue Albert Dhalenne, 93400 Saint-Ouen, France*

[5]*DB Systemtechnik GmbH, EUREF-Campus 17, 10829 Berlin, Germany*

Abstract

This study was carried out within the working group for Railway applications and Noise emission under the European Committee as part of ongoing work towards a new European standard for evaluating audible warnings in train drivers' cabs. The purpose was to investigate and compare different methods of assessing audibility to inform the proposal for a standardised methodology.

Controlled subjective tests were conducted in a mock-up train cab, where ambient noise recordings from several train cabs were replayed at realistic levels and eight representative warning sounds were superimposed. Participants rated the audibility of each sound presented at each of three signal-to-noise ratios. These subjective results were compared to the Technical Specification Interoperability "Locomotives and Passengers" homologation requirement (+6 dB signal-to-noise ratio) and to sound levels computed using 'Detectsound', a tool developed by the University of Ottawa to predict the audibility of sounds in noise from the one-third-octave band spectra of the signal and noise. Detectsound models the auditory filter characteristics of the ear and takes into account the hearing abilities of different listener populations.

Initial findings confirm that the +6 dB signal-to-noise criterion aligns with perception for speech signals but tends to be conservative for tonal warning sounds, resulting in warning sounds that are sometimes louder than desirable. Detectsound calculations show closer agreement with perceived audibility, although discrepancies remain. Detectsound also highlights the substantial influence of hearing ability. Warning sound levels need to be increased by 5 dB to 7 dB for persons just meeting the minimum hearing requirement for train drivers in Europe compared with persons with normal hearing.

Keywords: *audibility, warning sounds, train cabs*

1. Background and motivation

Audible warnings in the driver's cab are intended to ensure an appropriate response from the train driver. Various types of warning signals are used to convey information, attract attention, or indicate urgent situations (e.g., fire detection or the emergency opening of one or more doors). These warnings range from high-priority alarm tones to speech-based announcements communicating specific operational conditions to the driver.

To be effective, audible warnings must be clearly distinguishable from the background noise present in the driver's cab. Insufficient warning levels may result in missed signals, while excessive levels may impair driver performance through startle effects, annoyance, or discomfort. Furthermore, overly loud warnings may contribute to noise exposure levels exceeding regulatory requirements and could potentially affect the driver's hearing. This paper compares several methods for assessing the audibility of audible warnings in the driver's cab.

2. Existing methods of assessing audibility

2.1. Overview

Three methods of assessing the audibility of non-speech warning sounds in train drivers' cab were considered: the TSI LOC&PAS [1], ISO 7731 method c [2], and Detectsound [3,4]. The TSI LOC&PAS requirement, is specific to warning sounds in a driver's cab while ISO 7731 and Detectsound are universal tools.

2.2. The LOC&PAS TSI

The LOC&PAS TSI requires the A-weighted sound levels of audible warnings to be at least 6 dB above the A-weighted background noise level. More precisely, the ERA [5] specifies the equation (1)

$$L_{\mathrm{pAFmax,signal}} - L_{\mathrm{pAeq,T}} \geq 6\ \mathrm{dB} \qquad (1)$$

In electric trains, the background noise usually increases with speed whereas in diesel locomotive cabs the noise depends on the traction power. Highest noise levels in diesel cabs often occur at low speeds when pulling away from rest. The TSI requirement does not take account of the noise spectrum, masking or hearing abilities of drivers.

2.3. Detectsound and ISO 7731 method c

The most accurate method of ISO 7731 is method c. Detectsound, developed by the University of Ottawa, and method c of ISO 7731 are similar in principle. Both assess the sound levels of audible warnings relative to the 'masked threshold' imposed by the background noise in one-third octave bands.

The masked threshold of a sound is usually defined as the minimum sound level of that sound which is just audible in the background noise. However, a sound just below the masked threshold may sometimes be audible, and a sound just above may sometimes be missed. The masked threshold is the sound level at which the warning will be detected in the noise on 50% of occasions.

An effective warning sound must not only be audible, but its level must exceed the masked threshold by a suitable margin to attract attention and be recognised; Patterson [6] reported that a sound which is 10 dB above its masked threshold is "easy to hear", and a sound 15 dB above is "difficult to miss". ISO 7731 method c and Detectsound specify margins of 13 dB and 12 dB respectively above the masked threshold. The term 'clearly audible' is often used for sounds which meet these criteria rather than simply 'audible'.

ISO 7731 requires that a warning should be audible within 1 s and the sound should remain audible for at least 2 s if noise causes temporary masking. This is not compatible with short warnings requiring a rapid response.
ISO 7731 also mandates that time-weighting 'S' be used for measuring both the ambient noise and signal. This time weighting is too slow to respond to short duration or transient warning sounds. Time-weighting 'F' is more suitable and better reflects the averaging time of the human ear.

Detectsound has some advantages over ISO 7731. For simplicity, ISO 7731 method c generally assumes that the one-third octave sound levels will approximate the masked threshold in each band. Detectsound calculates the masked threshold more precisely in each one-third octave by modelling the ear and hearing system as a bank of auditory filters, taking into account the effective rectangular bandwidth of an ear and the sensitivity in each frequency band. Validation tests [4] showed that Detectsound estimated masked thresholds to within ±1 dB with a standard deviation of 2.5 dB for listeners with normal hearing. By varying the filter parameters, which can be estimated from an audiogram, Detectsound allows the hearing characteristics of different sample populations to be taken into account.
For example, audibility can be assessed for drivers whose hearing is only just sufficient to obtain a train driving licence.

ISO 7731's generic approach provides conservative estimates of audibility for a large portion of the population with a wide range of hearing abilities, which are unspecified.
Sample calculations with Detectsound indicate that a 'worst case' driver with a moderate hearing loss but able to meet the requirements for driving a train would need louder warning sounds than a driver with normal hearing, by 3.1 dB on average in an electric train, and by 3.5 dB in a diesel cab.

Whereas ISO 7731, Detectsound and the TSI specify the minimum levels for a warning sound, Detectsound alone also specifies the maximum levels, usually 25 dB above the masked threshold, to ensure that warnings are not too loud or aversive, and do not disrupt the driver's thought or interfere with communication.

2.4. Using Detectsound

The auditory filter parameters in Detectsound were updated in 2014 to take account of more recent research, but the earlier 2004 parameters still can be selected.

Detectsound requires three inputs:

- The one-third-octave spectrum of the warning sound, if known
- The one-third-octave spectrum of the background noise (usually the L_{Zeq} in each band
- The representative audiogram of the listening population, ideally corresponding to the minimum hearing requirements for train drivers [7]

Detectsound does not prescribe a single metric for warning-signal levels. The L_{Zeq} in each band could be used if the sound is steady in level and longer than, say 250 ms, but for short duration sounds or those that vary in level or fade quickly with no definite duration, such as a 'ping' or a 'chime', the L_{ZFmax} is more logical and appropriate, and can also be used for longer sounds.
The L_{ZFmax} was previously used to assess cab warnings for the Rail Safety and Standards Board in Britain [8]. However, a study in Dutch cabs [9] used the average level over the first 200 ms for transient decaying signals. The L_{ZFmax} is currently being considered for the proposed CEN standard for drivers' cabs.

Detectsound calculates the masked threshold in one-third octave bands from the noise spectrum, and outputs the minimum and maximum levels for warning sound components in each band. The band levels of an existing warning can be assessed to ensure that they fall between the minimum and maximum values. The minimum and maximum levels can also be used as a design window for new sounds.
As an example, Fig 1 shows the minimum and maximum band levels for clearly audible warning sound components calculated by Detectsound from the noise spectrum in the cab of an EMU travelling at 160 km/h for a driver with the minimum acceptable hearing capability.

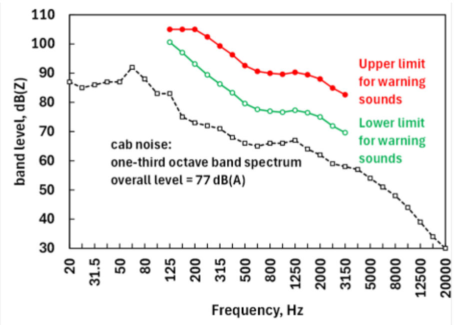

Figure 1: The maximum and minimum warning sound levels recommended by Detectsound (2014 version) in an EMU cab.

3. Perceptual test campaigns

To investigate how the objective audibility criteria relate to subjective perception, a campaign of controlled listening tests was carried out. The tests were designed to reproduce the acoustic conditions of a train driver's cab under controlled laboratory conditions and to collect audibility ratings for a range of representative warning sounds presented at different preselected signal-to-noise ratios. The subjective ratings obtained in this way provided the reference against which the +6 dB overall-level emergence criterion and the Detectsound predictions were compared.

3.1. Test procedure

Eight representative warning sounds were selected for the campaign, covering the main categories of audible warnings used in drivers' cabs, namely speech-based announcements and tonal signals. Ambient noise recordings captured in the cabs of two different trains, one electric (EMU) and one diesel (DMU) unit, were used to represent the background noise. The warnings had to be detected in the presence of these background noises. Each recording was reproduced at a realistic sound level corresponding to the operational conditions in which it had been measured.

During the tests, each warning sound was superimposed on a background noise recording and presented to the participants at a subset of the prepared signal-to-noise ratios. The range of prepared signal-to-noise ratios was chosen to span levels below and above the +6 dB overall-level emergence required by the TSI LOC&PAS, with particular emphasis on the borderline configurations in which the warning is close to being either too quiet or too loud.

The following average levels relative to the background noise were prepared: -3 dB, 0 dB, +3 dB, +6 dB, +9 dB, +12 dB and +15 dB. However, not all seven levels were tested for each signal. Depending on the signal, either the four levels -3 dB, +3 dB, +9 dB and +15 dB, or the three levels 0 dB, +6 dB and +12 dB were used. The signals were adjusted for each background noise. This allowed the transition between clearly inaudible and clearly audible conditions to be resolved for each type of signal.
In combination with the two background noises, the different warning (speech or tone signals) and the superimposed warning sounds corresponded to the selected loudness subsets, a total of 52 listening tests were performed.

The listening tests were performed with five participants at Siemens Mobility in Krefeld and three participants at ISVR Consulting in Southampton. For each signal-to-noise configuration, participants rated the audibility of the warning sound on a discrete numerical scale from 1 to 5 and the meaning of each rating is defined by 1 : Hardly audible, 2 : Easy to miss, 3 : Easy to hear, 4 : Difficult to miss, 5 : Too loud.

The playback chain was calibrated prior to testing so that both the reproduced background noise and the superimposed warning sounds corresponded to the intended sound levels at the listening position.

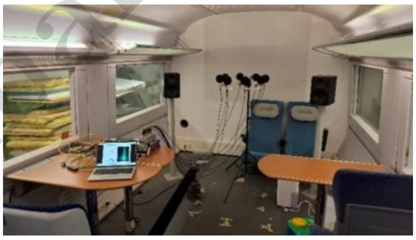

Figure 2: Mock-up used for the perceptual test campaign at Siemens Mobility Krefeld.

3.2. Test environment

For logistical reasons, the listening tests were conducted at two sites: (i) the Siemens Mobility site in Krefeld, Germany, using a mock-up of a passenger compartment (Fig 2 and Fig. 3); and (ii) the ISVR at the University of Southampton, where a purpose-built acoustically treated listening room was used (Fig. 4).

Commenté [TH1]: Please check if correct:
2 background noises
8 signals with DMU -> 28 tests
13 signals with EMU -> 36 tests

-->64 tests.

Commenté [MR2R1]: We tested 7 signals with EMU and and 8 signals with DMU. 24 cases tested with EMU and 28 cases with DMU is 52 cases in total, I updated the values

Commenté [CE3]: n clause 4.1 this is translated into numerical scores without explaining the mapping of judgements and scores.

Commenté [TH4R3]: Maybe add analysis? Here or in 4.1?
@Maxime

Commenté [MR5R3]: «The individual responses were subsequently averaged across participants and classified as too low, acceptable or too loud, according to the criteria detailed in the following section" I think this should be deleted as it is described in 4.1
However, I think we should describe here the 1-5 scale that was used for the evaluation and list the associated semantic for each level

Commenté [MR6R3]: 1 : Hardly audible
2 : Easy to miss
3 : Easy to hear
4 : Difficult to miss
5 : Too loud

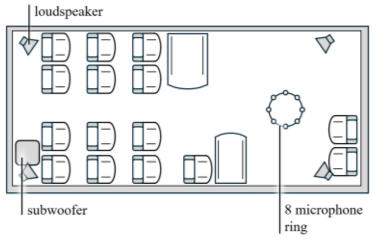

Figure 3: Schematic overview of the Siemens mock-up used for the perceptual test campaign

The mock-up at Siemens Mobility consists of a short section of an ICE 2-carbody. While not being identical to a cab, it provides a controlled and reproducible acoustic environment in which the recorded cab background noise can be replayed and the warning sounds superimposed, while the participants were seated in a listening position representative of that of a train driver. The test arrangement is illustrated in Figs. 2 and 3.

The background noise was played by all four loudspeakers and the subwoofer while the signals were reproduced only by the two loudspeakers close to the setup of eight microphones distributed on a circle of 0.5 m diameter. The participants were seated in the first two rows of seats.

The ISVR tests were conducted in an acoustically treated listening room measuring 4.8 m (L) × 4.0 m (W) × 2.5 m (H) (Fig 4). Participants were seated in a row of three near the centre of the room. The background and source sound was reproduced through two full-range loudspeakers, which were driven with identical signals. The sound pressure level at the central listening position was measured using a Class 1 sound level meter and adjusted to the required playback level. The relative positions of the loudspeakers were similar to those used in the tests at Siemens Mobility, thereby providing a comparable sound field and level at the listening positions.

Figure 4: Interior of ISVR acoustically treated listening room

4. Comparison of evaluation methods

4.1. Method

The responses were averaged across participants and classified according to the following criteria:

- Average score < 3: too low
- Average score between 3 and 4: acceptable
- Average score > 4: too loud

The resulting classifications were compared with the audibility indicators derived from the overall-level emergence criterion and from Detectsound.

For the overall-level evaluation, signal emergences were grouped into three categories:

- < 6 dB: below the minimum TSI LOC&PAS requirement (non-compliant)
- = 6 dB: at the minimum TSI LOC&PAS requirement (compliant)
- > 6 dB: above the minimum TSI LOC&PAS requirement (compliant)

For the Detectsound evaluation, the results were grouped into three categories:

- Too low: all spectral components are below the criterion of 12 dB above the masked threshold
- Acceptable: at least one spectral component lies between 12 dB and 25 dB above the masked threshold, with no component exceeding 25 dB above the masked threshold
- Too loud: at least one spectral component exceeds 25 dB above the masked threshold

4.2. Perceptual results vs +6 dB emergence criterion

The subjective evaluations showed different behaviours for speech and tonal signals. For speech signals, the current TSI LOC&PAS requirement of an overall level emergence of at least +6 dB generally corresponded to ratings indicating satisfactory audibility.

As shown in Table 1, no speech signal meeting the +6 dB criterion was classified as too low. Most compliant configurations were rated either acceptable or too loud, indicating that the criterion provides a reliable indication of audibility for speech signals.

Table 1: Speech signals evaluation (rows) in function of overall level emergence (columns) – values in % of answers
0% 50%

Evaluation	< 6 dB	= 6 dB	> 6 dB
Too low	39%	0%	0%
Acceptable	6%	17%	6%
Too loud	0%	0%	33%

For tonal signals, a different trend was observed (Table 2). A significant proportion of tonal signals were considered acceptable despite not meeting the +6 dB emergence criterion. Specifically, 29% of the evaluations were judged acceptable while remaining below the regulatory threshold. Conversely, 48% of the configurations meeting the criterion (+6 dB or above) were judged too loud. These results suggest that the overall-level emergence criterion may be rather conservative for tonal warning sounds and can lead to signal levels higher than necessary to ensure audibility.

Table 2: Tone signals evaluation (rows) in function of overall level emergence (columns) – values in % of answers
0% 50%

Evaluation	< 6 dB	= 6 dB	> 6 dB
Too low	13%	0%	0%
Acceptable	29%	6%	3%
Too loud	0%	13%	35%

4.3. Perceptual results vs Detectsound predictions

Detectsound predictions were compared with the subjective ratings obtained during the listening tests. The results for tonal signals are presented in **Table 3**. Compared with the overall-level criterion, the lower audibility limit predicted by Detectsound showed closer agreement with perceived audibility. In particular, no configuration rated as too low was simultaneously classified as acceptable by Detectsound, whereas only 3% of the evaluations rated too low corresponded to a Detectsound prediction classified as acceptable.

Several tonal signals judged as acceptable by the participants did not satisfy the +6 dB overall emergence requirement but were correctly identified as audible by Detectsound. Nevertheless, discrepancies remained, as some signals perceived as acceptable were classified by Detectsound as either too low or too loud. Overall, the results indicate that Detectsound provides a more frequency-sensitive assessment of signal detectability than the overall-level emergence criterion.

Table 3: Tone signals evaluation (rows) as a function of Detectsound classification (columns) – values in % of answers
0% 50%

	Detectsound prediction		
Evaluation	Too low	Acceptable	Too loud
Too low	6%	3%	3%
Acceptable	3%	26%	10%
Too loud	0%	16%	32%

4.4. Sensitivity to Detectsound version

A comparison between Detectsound 2004 and Detectsound 2014 revealed noticeable differences in the calculated audibility values.
The largest variations were observed for signals whose dominant spectral components were located above approximately 2.5 kHz. For these sounds, the predicted audibility margins differed by up to approximately 2 dB between the two versions. By contrast, signals dominated by lower-frequency components exhibited differences below 1 dB. These variations are mainly attributable to the frequency-dependent correction factors incorporated in the 2014 version of the model.

4.5. Influence of hearing thresholds assumptions

The assumed hearing profile was found to have a significant influence on the calculated audibility. Using the minimum hearing requirements applicable to train drivers in Europe instead of the standard hearing profile (-10 dB HL for all frequencies) resulted in substantial changes to the predicted audibility margins. For signals exceeding the upper Detectsound limit, differences typically ranged from 5 dB to 7 dB. Near the lower audibility limit, smaller but still significant differences ranging from approximately 2.5 dB to 4.5 dB were observed.

5. Conclusion and discussion

This study compared different approaches for assessing the audibility of warning sounds in train drivers' cabs as part of the ongoing development of a European standard within CEN TC256 WG3.

For spoken messages the controlled listenings tests confirmed the current TSI LOC&PAS requirement of $L_{pAFmax,signal} - L_{pAeq,T} \geq 6$ dB as appropriate. This criterion delivers reasonable indications of audibility for speech-based warnings.

Tonal warning signals were often perceived as clearly audible at lower emergence levels. The comparison with Detectsound predictions showed that a psychoacoustic approach based on frequency-dependent masking and hearing characteristics provides a closer representation of subjective perception than a simple overall-level criterion.

The study also highlighted the significant influence of hearing ability on warning-signal perception. Listeners with not more than the required minimum hearing capability need higher warning signal levels.

The study results recommend the use of perceptually based assessment methods for future requirement definitions and assessment methods. Regulation and standardisation can be improved by the use of psychoacoustic tools such as Detectsound. This would enable a more robust framework for balancing audibility, comfort and operational safety in train drivers' cabs.

Acknowledgments

Detectsound was produced by Christian Giguère, Chantal Laroche and Yun Zheng of the University of Ottawa.

References

[1] TSI LOC&PAS, European Commission Regulation 1302/2014. Official Journal of the European Union L356/228 12.12.2014

[2] ISO 7731; 2003, "Ergonomics — Danger signals for public and work areas — Auditory danger signals." International Organization for Standardization. Geneva.

[3] Y Zheng, C Giguère, C Laroche, C Sabourin "Detectsound Version 2: A software tool for adjusting the level and spectrum of acoustic warning signals." Canadian Acoustics / Acoustique canadienne. 31(3) 76-77.2003.

[4] Y Zheng, C Giguère, C Laroche, C Sabourin, A Gagne, and M Elyea, "A psychoacoustical model for specifying the level and spectrum of acoustic warning signals in the workplace. Journal of Occupational and Environmental Hygiene, 4(2), 87–98. 2007

[5] ERA "Guide for the application of the LOC&PAS TSI" Application guide GUI/LOC&PAS TSI/2023, European Agency for Railways, 2023

[6] R.D. Patterson, "Guidelines for auditory warning systems on civil aircraft." CAA Paper 82017, Civil Aviation Authority, London, 1982.

[7] Annex II in Directive 2007/59/EC (Train Drivers Directive) Official Journal of the European Union L 315, 3.12.2007, p. 51–78, 2007.

[8] M.C. Lower and M.G.R. Toward, "Relationship Between Train Horn Test Measurements and Perceived Sound Levels on the Track: Review of the Literature and Industry Practice" Project T1205, Rail Safety and Standards Board, London, 2021.

[9] H.E.M.van der Hoek-Snieders, R. Houben and W.A. Dreschler, "Detectability of auditory warning signals in the ambient noise of Dutch train cabins." Ergonomics 64, 474-484, Taylor & Francis. Published online: 24 Oct 2020.