Efficiency of loudness models for the evaluation of airplane cockpit noise comfort

E. Parizet[a], Y. Padayashi[a] and O. Collery[b]

[a]Laboratoire Vibrations Acoustique INSA Lyon, 25 bis avenue Jean Capelle, Villeurbanne,
F-69621 Lyon, France
[b]Airbus Opération SAS, 316 route de Bayonne, 31060 Toulouse, France
etienne.parizet@lva.insa-lyon.fr

Aircraft interior noise is a major design stake with respect to airline requirements for passenger and crew comfort. The focus was put here on cockpit noise. The goal of this study was to evaluate unpleasantness of such sounds and to assess the relevance of existing metrics for its prediction. Six sounds recorded in various airplanes were used; one of them was modified in order to reduce the level of an emerging low-frequency component, leading to a number of seven stimuli. These stimuli were used in a pair-comparison experiment in a sound-proof booth. 31 listeners participated to the experiment. Results could be analysed using a BTL model, which provided an unpleasantness scaling of the stimuli.

Unpleasantness was clearly related to loudness. Predictions from two loudness models were investigated: the ISO-532 one and the ANSI S3.4 one. It appeared that the former gave better results. It is argued that this is due to the evolution of loudness values in the low frequency range, which is different between these two models.

1   Introduction

Aircraft interior noise is a major design stake with respect to airline requirements for passenger and crew comfort. Innovative acoustic techniques are therefore required to come up to customer expectations. The focus is put here on cockpit noise aiming at quantifying the potential of loudness metrics to assess pilots comfort for specific noise samples.

Aircraft interior noise comfort is classically evaluated using sound metrics as A-weighted sound pressure level or Speech Interference Level (SIL). The motivation of this study was to evaluate the accuracy of loudness metrics, which have been proved to be useful in many sound quality applications.

2   Subjective experiment

2.1   Stimuli

Six stimuli were provided by Airbus corresponding to recordings in flight conditions with a microphone in the cockpit of six aircraft (fs = 48 kHz, 16 bits). One of the stimuli exhibited an abnormally high emerging frequency (63 Hz). An additional stimulus was created by filtering out this frequency; the spectra of the original and modified signals are presented in Figure 1.
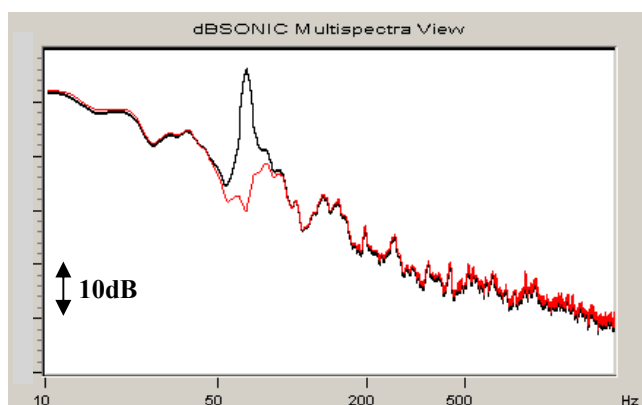

Figure 1: Spectrum of the original signal (black thick curve) and the modified on (red thin curve).

The seven sounds duration was limited to 5 seconds, and a 150 ms initial and final fading was applied to signals.

2.2   Experimental set-up

The experiment was conducted in the sound-proof booth of the laboratory. Sounds were presented by a pair of loudspeakers (Mackie Tapco S8) placed relatively to the listener on a triangular arrangement, at an approximate distance of 2 m from the subject. A two-ways third octave equalizer allowed correcting the frequency response of the set-up, so that this response (as measured at the position of the listener) was flat enough between 50 Hz and 16 kHz. In this frequency range, the variation of one-third-octave band levels was lower than 3 dB.

The overall sound pressure levels of stimuli constitute representative data of sound levels experimented in cockpits.

2.3   Procedure

A paired comparison procedure was used, as this gives the most accurate results [1]. 24 pairs of sounds were presented to each subject, including 2 training ones and a pair of twice the same stimulus. First of all, the set of sounds was presented to the listener, who was informed that the stimuli had been recorded in airplanes cockpits. Then the pairs were arranged according to a Ross series [2], after a random permutation of the stimuli.

After listening to a pair, the subject had to compare the unpleasantness of the two sounds and gave his answer by moving a cursor on a continuous scale. In order to help him, five labels were figured on the scale, from "A much more unpleasant than B" to vice-versa. The listener could hear each pair as many times as he felt necessary to answer. The whole procedure was managed by the *Jury Testing* software from 01dB-Metravib.

31 subjects participated to the experiment: 15 students and 16 older searchers of the laboratory, 12 women and 19 men. They did not relate any hearing impairment, but their hearing ability was not checked.

2.4   Results

First of all, the number of circular triads was computed according to the method explained in [3]. The averaged percentage of circular triads was 2%, which indicated that listeners could achieve the task.

The agreement of the panel was evaluated thanks to a hierarchical cluster analysis (using Ward's method). This agreement was good enough to average the results over the whole jury.

Finally, the preference probabilities were transformed in scores using the BTL (Bradley-Terry-Luce) method [4] which proved to give better results than Thurstone III or V methods. These scores are presented in Figure 2, together with their confidence interval (p = 0.95), computed using a boot-strap method (500 trials, 20 subjects, randomly selected with replacement). As it could be expected, the filtering of the 63 Hz frequency strongly reduced unpleasantness.
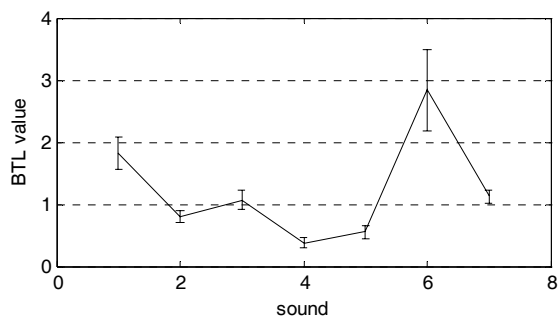
Figure 2: unpleasantness BTL scores of the sounds

3   Metrics

Sound metrics were computed and their values were compared to subjective evaluations (BTL values). It appears that predictions from loudness models are well correlated to subjective results for these seven specific stimuli.

On the one hand, Zwicker's loudness (computed according to ISO 532-B standard) gave better prediction. The relation between loudness values and BTL ones is presented in Figure 3. The correlation coefficient is $R = 0.86$ ($p = 0.013$).
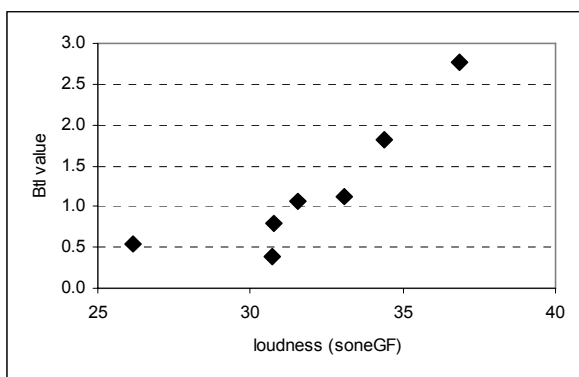

Figure 3: Loudness computed according to ISO 532-B standard and BTL values

The correlation between preference probabilities and loudness ratios was also strong (Figure 4): $R = -0.82$, $p < 0.001$.
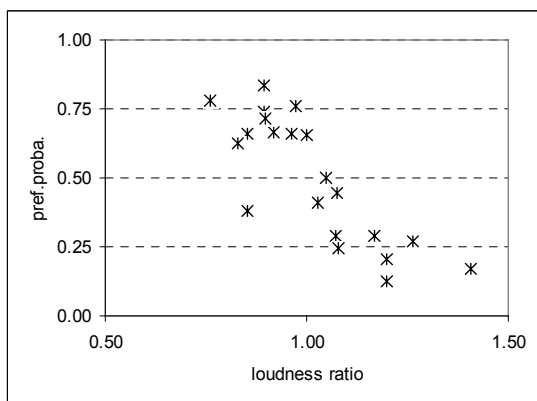

Figure 4: Loudness ratio (ISO 532-B standard) and averaged preference probabilities

On the other hand, loudness values computed according to ANSI S3-4 (2007) standard were not correlated to subjective evaluations ($R = 0.49$, $p > 0.27$). The relation between these two sets of values can be seen in Figure 5.
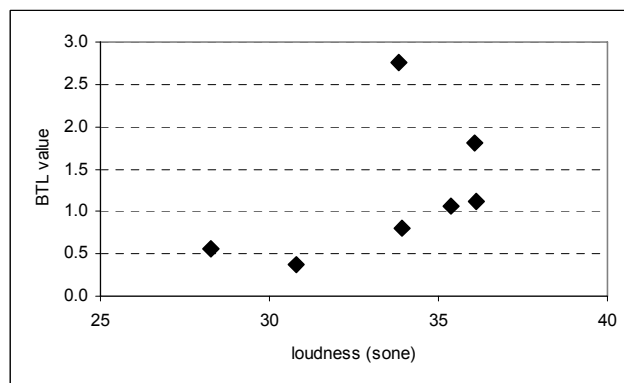

Figure 5: Loudness computed according ANSI S3-4 standard and BTL values

4   Discussion

Figure 4 shows that the low correlation between ANSI computed loudness values and subjective evaluations are mainly due to one of the sample. This sample exhibits the abnormal emerging frequency at 63 Hz (see Figure 1). It can be hypothesized that this frequency increased the unpleasantness of the sound but did not contribute to loudness. But it is a usual experience that loudness is a key factor for unpleasantness. Thus it can be argued that ANSI model underestimates the loudness of this very low frequency component.

This difference between the two models can be seen in Figure 6, which shows the variation of loudness with the sound pressure level for a 63 Hz pure tone, as predicted by the two models.
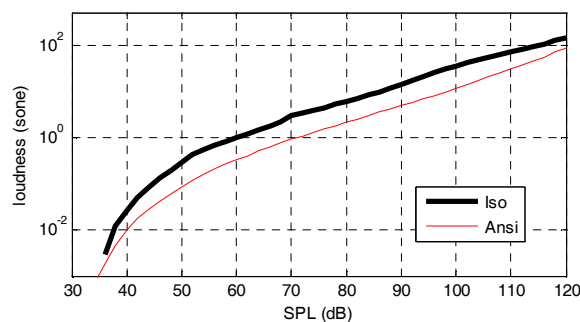

Figure 6: Loudness curve for a 63 Hz pure tone, as predicted by the two standards

In the whole SPL range, ANSI-computed loudness is lower than ISO-computed one. Each model was adjusted in order to reproduce iso-loudness curves, as defined by different versions of the ISO-226 standard (1987 and 2003). Such a difference exists between these two versions of the standard. For example, the iso-loudness value of a 80 dB at 63 Hz pure tone is 79.5 phon according to the oldest standard and 66.5 phon according to the newest one.

5 Conclusion

The results of this study show that the newest standardized loudness model may give less accurate results for very low-frequency sounds. Experiments giving direct loudness estimations of such sounds are now scheduled in order to check the inaccuracy of newest iso-loudness curves.

References

[1] E. Parizet, N. Hamzaoui, G. Sabatié "Comparison of some listening test methods : a case study", Acta Acustica united with Acustica 91, 356-364 (2005)

[2] R. Ross, "Optimum orders for the presentation of pairs in the method of paired comparisons", *J. of Educ. Psych*. 25, 375-382 (1934)

[3] E. Parizet, "Paired comparison listening tests and circular error rates", *Acta Acustica united with Acustica* 88, 594-598 (2002)

[4] H. David, *The method of paired comparison*, Oxford University Press, New-York (1988)