A NEW SET OF DIRECTIONAL WEIGHTS FOR ITU-R BS.1770 LOUDNESS MEASUREMENT OF MULTICHANNEL AUDIO

Leandro Pires[1], Maurílio Vieira[2], Hani Yehia[2], Tim Brookes[3], Russell Mason[3]

[1]Agência Nacional de Telecomunicações, Belo Horizonte, Brazil
[2]Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[3]Institute of Sound Recording, University of Surrey, Guildford, UK

Abstract – *The ITU-R BS.1770 multichannel loudness algorithm performs a sum of channel energies with weighting coefficients based on azimuth and elevation angles of arrival of the audio signal. In its current version, these coefficients were estimated based on binaural summation gains and not on subjective directional loudness. Also, the algorithm lacks directional weights for wider elevation angles ($|\phi| \geq 30°$). A listening test with broadband stimuli was conducted to collect subjective data on directional effects. The results were used to calculate a new set of directional weights. A modified version of the loudness algorithm with these estimated weights was tested against its benchmark using the collected data, and using program material rendered to reproduction systems with different loudspeaker configurations. The modified algorithm performed better than the benchmark, particularly with reproduction systems with more loudspeakers positioned out of the horizontal plane.*

Keywords – Broadcasting, loudness, signal processing, spatial audio, subjective test.

1. INTRODUCTION

In recent years, audio signal normalization based on loudness became a regular practice and the loudness algorithm described in Recommendation ITU-R BS.1770 is now ubiquitous in broadcasting and content producing workflows. With the advent of object and scene-based sound systems, it is reasonable to wonder if a channel-based loudness measurement, originally designed for stereo and 5.1 content, would still suit the needs of program level management and control. After a hiatus, Rapporteur Group on loudness measurement algorithm (RG-32) was re-established in March 2019 to resume studies on measurements for object and scene-based audio in light of the new Recommendation ITU-R BS.2127, which specifies the reference renderer for these immersive audio formats [1]. Whether the algorithm will evolve to a format-based loudness computation or remain multichannel, and be used in the output of the renderer, is still to be determined.

An attempt to address this issue was made in the latest version of ITU-R loudness model (BS.1770-4), in which the positional weighting coefficients were extended to an unrestricted number of channels [2]. The weighting coefficient G_i for an i-th position is derived from Table 1, which is a generalization of the original 5.1 weighing scheme for the horizontal plane, and no weighting is applied in broader elevation angles ($|\phi| \geq 30°$). Although RG-32 considered differences of directional loudness in wider elevation angles, the Rapporteur Group decided to keep simple positional weights envisioning backward compatible extension of ITU-R BS.1770-3 and, consequently, avoiding implementation issues.

Table 1 – Position-dependent channel weightings in ITU-R BS.1770-4

Elevation (ϕ)	Azimuth (θ)		
	$\lvert\theta\rvert < 60°$	$60° \leq \lvert\theta\rvert \leq 120°$	$120° < \lvert\theta\rvert \leq 180°$
$\lvert\phi\rvert < 30°$	1.00 (± 0 dB)	1.41 (+ 1.50 dB)	1.00 (± 0 dB)
else	1.00 (± 0 dB)		

At that time, RG-32 conducted experiments with two different datasets using 5.1 and 22.1 loudspeaker configurations. The obtained Pearson correlation coefficients, describing the linearity of the relationship between measurements and subjective results, were $r = [0.817, 0.902]$, which were lower than the correlation obtained in the original tests of ITU-R BS.1770 ($r = 0.977$) [2, p.17]. Also, authors in [3] tested this set of weights using 5.1 and 7.1 systems with and without screen loudspeakers. From a common database comprised of object and channel-based content, listening tests were made in three test sites and the observed correlation coefficients were $r = [0.962, 0.820, 0.898]$, being the first coefficient the only one on par with the correlation from the original ITU-R tests. Thus, the question on how to account for directional weighting in the ITU-R loudness model is still valid.

1.1 Directional weighting estimation in ITU-R BS.1770-4

Directional weights in the Recommendation were suggested in contributions to the Rapporteur Group, and later disclosed by Komori *et al.* [4]. Although the documents do not provide further detail on how these calculations were made, they can be traced back to the references below.

Robinson and Wittle first performed a subjective test to investigate loudness as a function of the orientation of the sound source. Through a series of sound pressure level (SPL) measurements at the ears of the listeners, the authors stated a binaural summation law of the form:

$$L = g \times log_2 \left(2^{\frac{L_{left}}{g}} + 2^{\frac{L_{right}}{g}} \right), \tag{1}$$

where g is a 6 dB binaural gain and L is the sound pressure level equivalent to any combination of incident sound pressure levels, being them diotic ($L_{left} = L_{right}$) or dichotic ($L_{left} \neq L_{right}$) [5].

A different binaural summation gain for Equation *(1)* was derived with the method proposed by Sivonen and Ellermeier in an experiment with narrowband, anechoic stimuli. The experimental gain g was estimated by a minimization of the sum-of-squares of the errors (SSE) between the directional loudness sensitivities (DLS) of listening test subjects and the sensitivities computed by Equation *(1)*. Squares were summed across I azimuth angles and J repetitions [6]. The minimum SSE is calculated as:

$$\min_g \left[\sum_{i=1}^{I} \sum_{j=1}^{J} \left\{ \mathrm{DLS}_{i,j} - \left[L_{comp_i}(g) - L_{ref}(g) \right] \right\}^2 \right], \tag{2}$$

where L_{comp_i} and L_{ref} are levels computed with Equation *(1)*, corresponding to the compared incidence, and to the frontal incidence of reference, respectively. The study obtained L_{ref} and $L_{comp_i}, \forall i$ from individual Head-Related Transfer Functions (HRTFs) of the expert subjects in their listening test. A value of $g \approx 3$ dB was then estimated by averaging Equation *(2)* computations per participant.

Authors in [4] computed the channel weighting values, or binaural loudness summation gains, in Table 2 by computing Equation *(1)* with $g = 3$ dB and $L_{left}(\theta)$ and $L_{right}(\theta)$ obtained from HRTFs of each azimuth angle $\theta = \vartheta$ in the table. Based on the verification that the effect of incidence angle on loudness is attenuated for wideband and reverberant sounds [7], the authors chose to normalize results to 1.5 dB and approximate them in 1.5 dB steps to ensure backward compatibility, leading to the directional weighting gains of the ITU model summarized in Table 1.

However, a further study by the authors in [6] computed Equation *(2)* with L_{ref} and $L_{comp_i}, \forall i$ obtained through SPL measurements taken with a Head and Torso Simulator (HATS), and with DLS subjective scores taken from naive participants. Minimization of the objective function in Equation *(2)* yielded $g \approx 6$ dB, closer to the binaural summation law in [5]. The authors observed that the effect of the contralateral incidence in the response variable was larger in the listening test with naive participants, although the difference in binaural gains in both studies might be due to chance, according to their statistical analysis [8].

Table 2 – Binaural summation gains computed for $\phi = 0°$ and directional weights proposed by the authors of [4] to RG-32.

Azimuth (θ)	0°	±30°	±60°	±90°	±110°	±135°	180°
Computed levels (dB)	0.00	1.36	4.47	5.22	4.46	0.84	−8.25
Normalised gains (dB)	0.00	0.39	1.29	1.50	1.28	0.24	−2.37
Proposed weights (dB)	0.00	0.00	1.50	1.50	1.50	0.00	−1.50

Additionally, experiments in [5, 6] were conducted in anechoic chambers using single channel narrowband noises as stimuli, while the derived weights for $|\phi| < 30°$ were tested in [3, 4] with broadband content rendered to 5.1, 7.1 and 22.2 loudspeaker settings, resulting in different correlations between objective measurements and subjective scores observed in the test sites. It is possible that these different results were due to elevation effects not accounted for in the weighting scheme of Table 1. Therefore, the question on how to model directional effects on the ITU-R loudness algorithm requires further investigation.

The goal of the present study was to obtain subjective data on directional effects in order to estimate a new set of binaural summation gains. The next sections contain a description of the listening test, followed by an attempt to reproduce the estimation that led to ITU-R BS.1770-4 and by a new approach to the problem. The modified algorithm was then tested against a different set of subjective data on multichannel audio.

2. LISTENING TEST

A loudness matching test was undertaken to obtain DLS responses through SPL adjustments required for equal loudness of sounds coming from different azimuths and elevations. For this listening test, a 22-channel electroacoustic system was used to reproduce broadband pink noise test signals in a ITU-R BS.1116 critical listening room [10] .

2.1 Design

Broadband pink noise stimuli, bandlimited from 200 Hz to 15 kHz, were reproduced by a 22 loudspeaker setup specified as layout 'H' in Recommendation ITU-R BS.2051 for advanced sound systems [9]. The layout is described in Table 3 where labels indicate bottom, middle, upper and top loudspeakers; and their correspondent azimuths. The time-aligned and level-equalized system was mounted in an ITU-R BS.1116 standard listening room with dimensions 7.35m length, 5.7m width, and 2.5m height [11]. Mean reverberation time between 500 Hz and 1 kHz octave bands is $RT_{60} = 0.22$s.

Table 3 – Azimuths and elevations of the 22.2 reproduction layout 'H' in Recommendation ITU-R BS.2051 for advanced sound systems [9] .

Azimuths θ (°)	Elevations ϕ (°)	ITU-R BS.2051 labels
-45	-30	B-045
0	-30	B+000
$+45$	-30	B+045
-135	0	M-135
-90	0	M-090
-60	0	M-060
-30	0	M-030
0	0	M+000
$+30$	0	M+030
$+60$	0	M+060
$+90$	0	M+090
$+135$	0	M+135
$+180$	0	M+180
-135	$+30$	U-135
-90	$+30$	U-090
-45	$+30$	U-045
0	$+30$	U+000
$+45$	$+30$	U+045
$+90$	$+30$	U+090
$+135$	$+30$	U+135
$+180$	$+30$	U+180
0	$+90$	T+000

All loudspeakers but the sub-woofers were *Genelec 8330A*. Sub-woofers were not used since the ITU-R loudness algorithm does not include Low-Frequency Effects (LFE) channels in its power sum. *Genelec Loudspeaker Manager* software was used for level alignment with respect to the central listening position and for automatic calibration of the frequency response equalization. A HATS placed at the listener position was used for calibration. The system was calibrated so that a -23 LKFS (Loudness, K-weighted, relative to nominal full scale) pink noise signal reproduced from the frontal loudspeaker ($\theta = 0°$, $\phi = 0°$) measured 65 dBA$_{(slow)}$ (Slow, A-weighted Sound Level) at the ears of the dummy head. HATS internal levels were also adjusted so that the binaural capture of the calibration signal also measured a loudness level of -23 LKFS.

Subject response format is given by Directional Loudness Sensitivity (DLS), which is the level difference between the frontal incident sound of reference and the non-frontal incident test sound after the loudness of both sounds are matched. The experiment was performed by twelve expert listeners in two fifty-minute sessions with a one-day break in between. The group was composed of postgraduate students and staff from the Institute of Sound Recording and the Centre for Vision, Speech, and Signal Processing at the University of Surrey; and undergraduate students from the *Tonmeister* course at University of Surrey. All subjects had prior experience of critical listening tests.

2.2 Methodology

Loudness matching tasks were performed with a method-of-adjustment procedure, in which subjects were required to adjust an acoustic attribute of a sound event (level) until the auditory event (loudness) corresponded to the auditory event of a reference stimulus. Participants were presented with a graphical user interface shown in Fig. 1 containing the instructions for the test. The interface was free from sliders, faders, Volume Unit (VU) meters or any indication of current levels and ticks for visual anchoring. This was made to avoid subject bias caused by intuitive notions of scaling and to ensure that DLSs were collected based solely on acoustic information.

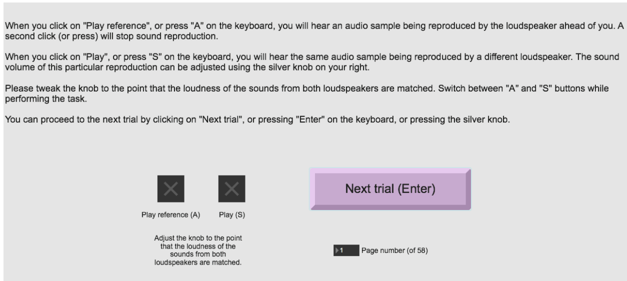

Fig. 1 – Graphical User Interface (GUI) of the listening test.

Broadband stimuli specified in Section 2.1 were reproduced by the electroacoustic system in a series of trials. In each trial, a test sound was randomly presented by one of the sources in Table 3, along with the reference sound presented by the frontal source M+000 ($\theta = 0°$, $\phi = 0°$). Sounds could be seamlessly interchanged by pressing specific keyboard buttons. The test sound was initially presented 10 dB above or below the reference level in a random fashion, and the listener could adjust its level by tweaking an infinite and unlabeled physical knob with ± 0.1 dB steps. When loudness matching has been done, the participant could then proceed to the next trial by pressing the physical knob or a keyboard button.

2.3 Results

Even though there was no missing data in the response variable, some DLS values were closer (or equal) to full-scale values of ± 10 dB and considered outliers. Five scores greater than two and a half times the standard deviation were considered extreme and deleted from the set. Scores were then broken into levels of the experimental factor and two scores greater than one and a half times their correspondent interquartile ranges were replaced by the highest non-outlier scores.

Subjects performed the listening test reasonably well. Mean sensitivities with 95% confidence intervals per participant are shown in Fig. 2. Note that $\frac{3}{4}$ of the confidence intervals fell within the ± 0.5 dB range, which is the just noticeable difference (JND) for loudness of broadband noise [12, p. 144]. Also, the remaining $\frac{1}{4}$ did not stand out so much, falling within the ± 1.0 dB range. Even though these results denote diligence in task performance, they might also indicate that direction effects are not large in size.

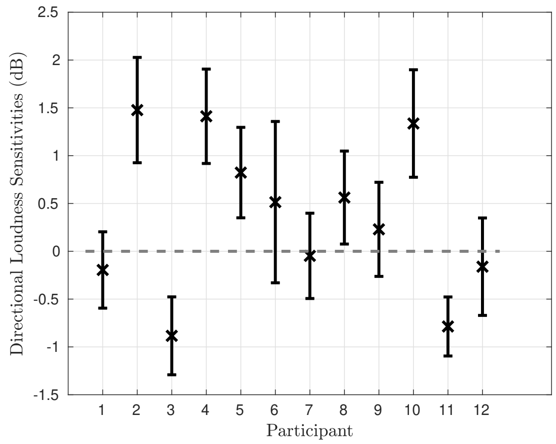

Fig. 2 – Means and 95% confidence intervals of subject scores (DLS).

The boxplot of subject responses displayed in Fig. 3 reinforces this notion. Sensitivities were higher on azimuths closer to $\pm 90°$ and lower at back incidences. This behavior is consistent through all horizontal planes. On the other hand, all interquartile ranges crossed the 0 dB line, which suggests that scores from non-discriminated directions are within the middle 50% of observations.

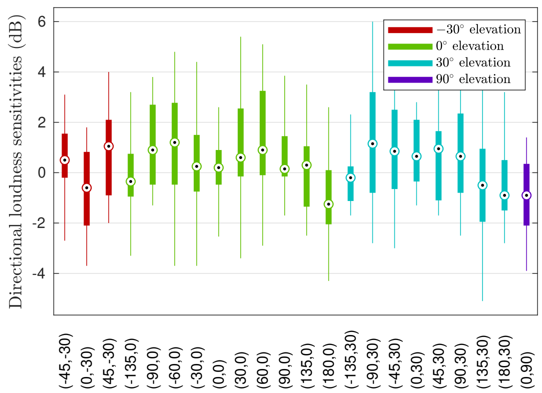

Fig. 3 – Boxplot of DLS per loudspeaker position

Variances were heterogeneous among directions $[F_{(21,501)} = 1.82, p = 0.015]$ and a non-parametric statistical test, Kruskal-Wallis, was needed to assess directional effects. Subjects' sensitivities were very significantly affected by positional changes of the sound source $[H_{(21)} = 69.93, p < 0.001]$. Pairwise comparisons using *t*-tests with non-pooled standard deviations spotted statistically significant differences in 58 out of 231 combinations of azimuth/elevation pairs. A summary of the significant differences is shown in Table 4. Total differences in bold stood out from the rest and corresponded to directions in the median sagittal plane, where the sound source is equidistant from the listener's ears. Effect sizes were large for 27 comparisons ($r > 0.5$), medium to large in 58 comparisons ($0.3 < r \leq 0.5$), and small to medium in 94 comparisons ($0.1 < r \leq 0.3$). The largest effect size observed ($r = 0.63$) corresponds to the difference between the U−090 / M+180 direction pair.

Table 4 – Summary of significant differences spotted on pairwise comparisons of directional loudness sensitivities among source directions.

Directions (θ, ϕ)	Very significant $(p < 0.01)$	Significant $(p < 0.05)$	Total differences
$(-45°, -30°)$	2	1	3
$(0°, -30°)$	3	8	**11**
$(45°, -30°)$	2	2	4
$(-135°, 0°)$	0	3	3
$(-90°, 0°)$	3	3	6
$(-60°, 0°)$	2	2	4
$(-30°, 0°)$	0	2	2
$(0°, 0°)$	0	2	2
$(30°, 0°)$	2	2	4
$(60°, 0°)$	3	4	7
$(90°, 0°)$	1	2	3
$(135°, 0°)$	0	1	1
$(180°, 0°)$	11	4	**15**
$(-135°, 30°)$	0	5	5
$(-90°, 30°)$	3	4	7
$(-45°, 30°)$	2	3	5
$(0°, 30°)$	2	3	5
$(45°, 30°)$	1	2	3
$(90°, 30°)$	0	2	2
$(135°, 30°)$	0	3	3
$(180°, 30°)$	0	7	7
$(0°, 90°)$	9	5	**14**

3. GAIN ESTIMATION

The multichannel loudness algorithm weights loudness values according to the angle of arrival of the signals and performs a linear sum of the results to provide a composite loudness measure [2], thus making adequate gain estimation an important component to address multi-directional sources. Since only $\frac{1}{4}$ of differences among the 22 levels of the experimental factor "direction" were significant, it is now understandable that obtaining directional gains with a straightforward procedure, like deriving a gain curve from subject means, would result in poor estimation. Possible approaches to gain estimation are presented in the following subsections.

3.1 Optimization problem

Directional weights in ITU-R BS.1770-4 were estimated assuming a binaural gain $g = 3$ dB from [6] and computing Equation *(1)* for a set of azimuths whose HRTFs were known. Instead of assuming an overall gain g, an alternate procedure is to compute Equation *(2)* with summations across participants and repetitions, to obtain a vector $\vec{g}$ whose elements correspond to a 22 loudspeaker layout of Section 2. This was done with the collected loudness sensitivities and the sound pressure levels measured at the ears of the HATS during the calibration stage.

The problem was to find a vector $\vec{g}$ that takes a scalar objective function $f(\vec{g})$ to a minimum, subject to constraints:

$$\min_{\vec{g}} f(\vec{g}) \text{ such that } \begin{cases} c(g_i) \leq 0, \forall i \\ lb \leq g_i \leq ub, \forall i \end{cases} \quad (3)$$

where $f(\vec{g})$ is the SSE between responses and predictions of Equation *(2)*, upper bound is perfect loudness summation ($ub = 10$ dB), lower bound is no summation at all ($lb = 0$ dB), and the non-linear constraint $c(g_i)$ is defined such that, in every i-th direction, equivalent monaural SPLs computed by Equation *(1)* cannot exceed the maximum sound pressure level reproduced in the listening test:

$$c(g_i) = g_i \times \log_2\left(2^{\left(\frac{L_{left,i}}{g_i}\right)} + 2^{\left(\frac{L_{right,i}}{g_i}\right)}\right) - 75 \leq 0, \ \forall i. \quad (4)$$

Although the estimated gain for rear incidence hit the upper bound, solutions from all azimuths on the bottom plane, frontal incidence, azimuths 0° and 180° on the upper plane, and from the top loudspeaker $(0°, 90°)$; converged to a global minimum of 0.69 dB. After small gain adjustments in order to make them symmetric with respect to the sagittal plane, and normalization of the largest gains in the set, corresponding to $(\pm 90°, 0°)$ directions, to 1.5 dB, the resulting weights are listed in Table 5.

Table 5 – Directional weights estimated by solving a constraint minimization problem.

Azimuths θ (°)	Elevations ϕ (°)	Gain $\vec{g}$ (dB)
−45	−30	0.00
0	−30	0.00
+45	−30	0.00
−135	0	0.44
−90	0	1.50
−60	0	1.10
−30	0	0.65
+0	0	0.00
+30	0	0.65
+60	0	1.10
+90	0	1.50
+135	0	0.44
+180	0	0.00
−135	+30	0.28
−90	+30	0.93
−45	+30	1.06
0	+30	0.00
+45	+30	1.06
+90	+30	0.93
+135	+30	0.28
+180	+30	0.00
0	+90	0.00

Note that all incidences with 0 dB weighting come from the bottom and median sagittal planes. Even though the estimated values seemed reasonable when compared to ITU-R BS.1770 directional weights, no insights could be gained from effects related to elevation and, consequently, no advancements were made on this front. Reproduction of the procedure in [6, 8], which consists on estimating gains per subject and taking the average, yielded an overall summation gain of $g = 3.54$ dB.

3.2 Regression problem

Treating gain estimation as a regression problem means training regression models to predict the sensitivity response variable. Predictors are localization cues chosen according to their correlation with subject responses. Then follows training and cross-validation of a regression model, until it is considered adequate under some performance criteria.

Interaural Level Difference (ILD) and Interaural Time Difference (ITD) are major localization cues in spatial hearing. While the former is accounted for in Equation *(1)*, the latter is considered when computing the Interaural Cross-Correlation Coefficient (IACC), a measure of similarity between ear signals given by the Interaural Cross-Correlation Function (IACF):

$$\mathrm{IACF}(\tau) = \frac{\int_{0 \text{ ms}}^{80 \text{ ms}} s_{left}(t) s_{right}(t+\tau) dt}{\sqrt{\left[\int_{0 \text{ ms}}^{80 \text{ ms}} s_{left}^2(t) dt\right]\left[\int_{0 \text{ ms}}^{80 \text{ ms}} s_{right}^2(t) dt\right]}}, \quad (5)$$

where t is time, τ is the interaural delay and s_{left} and s_{right} are the signals from the left and right ears, respectively. The IACC is defined as the maximum absolute value within $\tau \pm 1$ ms:

$$\mathrm{IACC} = \max_{\forall \tau \in [-1ms,\ 1ms]} |\mathrm{IACF}(\tau)|. \quad (6)$$

The interaural delay τ is an estimate of ITD when $\mathrm{IACF}(\tau)$ is maximum. The quantity $1 - \mathrm{IACC}$ is associated with the magnitude of spatial impression of a sound [13].

Moreover, the effect of contralateral incidence observed in [8] was taken into account in a recent update of Glasberg and Moore's loudness model, which incorporated binaural inhibition [14]. Being $\mathrm{IF}_{ipsilateral}$ the inhibition factor by which the short-term loudness of the ipsilateral signal is reduced by the effect of the contralateral signal, the inhibition model can be written in the form:

$$\mathrm{IF}_{ipsilateral} = \frac{2}{\left[1 + \left\{sech\left(\frac{\mathrm{STL}_{contralateral}}{\mathrm{STL}_{ipsilateral}}\right)\right\}^{\gamma}\right]}, \quad (7)$$

where $\mathrm{STL}_{contralateral}$ and $\mathrm{STL}_{ipsilateral}$ are vectors of short-term loudness values for contralateral and ipsilateral ears, and $\gamma = 1.598$. In the updated model, short-term loudness STL_{left} and STL_{right} are then divided by IF_{left} and IF_{right}, respectively. The value of γ was defined such that for diotic sounds the term in braces yields $[sech(1)]^{1.598} = 0.5$, and a diotic sound is predicted 1.5 times louder than its monaural equivalent.

To compute Equation *(7)*, ITU-R BS.1770 short-term loudness values, taken with a three-second integration window with one-second overlaps, were measured from the binaural signals recorded by the HATS at the calibration stage. As for Equation *(5)*, signals were integrated in their first 80 ms, when binaural sound pressures were produced by direct and early reflected sounds. Scatterplots of correlations between averaged DLS scores per direction and computed localization cues are shown in Fig. 4. Correlations are strong enough to proceed with these metrics as predictors of a regression model, although 3 dB and 6 dB summations would result in redundant predictors due to their almost identical scattering pattern in Fig. 4. The choice here was to consider only one binaural summation predictor with the same overall gain estimated in Section 3.1.

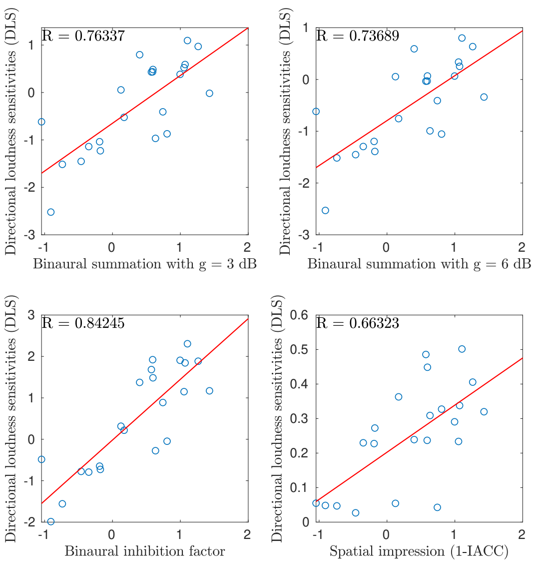

Fig. 4 – Scatterplots of correlations between DLS and localization cues.

A series of known regression model types, with and without Principal Component Analysis (PCA) preprocessing, were trained in a k-folds cross-validation scheme. Data was partitioned into $k = 5$ disjoint set of folds. For each fold, out-of-fold observations were used for training and in-fold observations for validation. Root Mean Square Error (RMSE), the Euclidean distance between a set of predictions and the actual observations, was computed over all folds then averaged. Plain linear regression resulted in the smallest error (RMSE = 1.7872).

For each i-th direction, the resulting model can be written in the form:

$$y_i = \alpha + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} + \varepsilon_i \qquad (8)$$

where y_i are the predictions of the response variable, α is the intercept term, $x_{1,i}$ is the binaural summation formula predictor, $x_{2,i}$ is the binaural inhibition model predictor, $x_{3,i}$ is the spatial impression predictor, and ε_i is the model residual. Intercept ($\alpha = -0.302$) and betas ($\beta = [-0.262\ 0.597\ 0.980]$) were estimated from 523 observations. Predictions for the 22-channel source directions are listed in Table 6. This time a small zero correction was made and no normalization was needed. Also, all gains are within ± 1.5 dB and elevation effects on the estimated gains are now more clearly defined.

Table 6 – Directional weights estimated by solving a linear regression problem

Azimuths θ (°)	Elevations ϕ (°)	Gain $\vec{g}$ (dB)
−45	−30	+0.26
0	−30	−0.68
+45	−30	+0.26
−135	0	+0.12
−90	0	+1.28
−60	0	+1.08
−30	0	+0.60
0	0	0.00
+30	0	+0.60
+60	0	+1.08
+90	0	+1.28
+135	0	+0.12
+180	0	−0.31
−135	+30	−0.09
−90	+30	+0.88
−45	+30	+1.12
0	+30	+0.44
+45	+30	+1.12
+90	+30	+0.88
+135	+30	−0.09
+180	+30	−0.26
0	+90	−0.62

Weightings for sound source directions not included in the 22.2 reproduction layout were estimated by smoothing the response data using local regression. ITU-R BS.2051 labels M±110 ($\theta = \pm 110°$, $\phi = 0°$) for 5.1 and 9.1 systems and U±110 ($\theta = \pm 110°$, $\phi = 30°$) for 9.1 systems yielded gains of 0.66 dB and 0.47 dB, respectively.

A modified version of ITU-R BS.1770 loudness with this set of weights is compared with the algorithm of reference by taking the differences between their measurements of the presented stimuli in Section 2.3, and plotting them against means and confidence intervals of participants. This is done in Fig. 5. Blue squares refer to differences in Loudness Units (LU) between measurements taken with the modified and the original algorithms, and jumps in the dashed line refer to ITU-R BS.1770 +1.5 dB gains in lateral incidences.

Differences between algorithms are more pronounced with sound sources on the upper plane, where measurements with the directional weights listed in Table 6 fall into subjects' confidence intervals in 8 out of 9 directions, against 3 out of 9 directions with the weights in Table 1. On the other hand, the modified algorithm performed worse than the original algorithm with sound sources in median sagittal plane, where predictors related to localization cues were

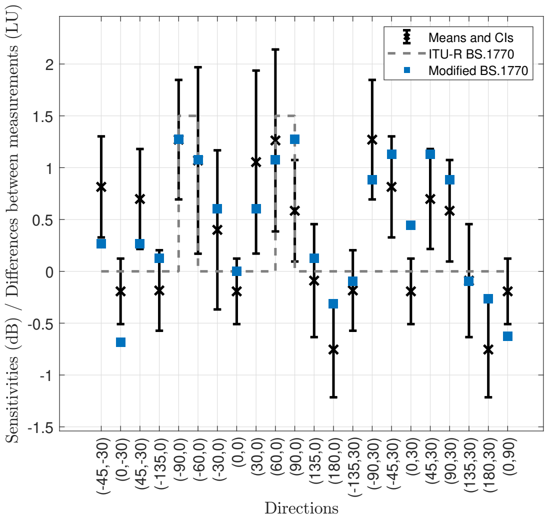

Fig. 5 – Differences between loudness measurements (LU) plotted against DLS means and confidence intervals (dB).

less effective at estimating gains of sources equidistant to the listener's ears.

4. TESTS WITH MULTICHANNEL AUDIO CONTENT

It was important to check if a modified loudness algorithm with a set of directional weights computed by Equation *(8)* can be generalized to measure program material items rendered to different spatial audio reproduction systems. This task was performed with the audio content used for the listening tests conducted in [15], kindly provided by the authors. In these tests, subjects were required to match the loudness of program items reproduced in mono, stereo, 9.1, 22.2 and cuboid[1] sound systems with the loudness of a reference 5.1 reproduction. Details on the program material and its production can be consulted in [16].

All rendered program items were measured by the ITU-R BS.1770 loudness algorithm and its modified version. These measurements were then fit to participant scores and the following performance statistics were computed: Pearson's correlation coefficients, RMSE, and the Epsilon-insensitive RMSE, or RMSE*, specified in Recommendation ITU-T P.1401 for evaluation in the context of subjective uncertainty [17]. RMSE* is the Euclidian distance between measurements and subjective data, considering only distances that fall into the 95% confidence intervals of listening test scores. In this assessment, the modified algorithm ($r = 0.9263$, RMSE $= 1.01$, RMSE$^* = 0.56$) performed better than standard ITU-R BS.1770-4 ($r = 0.9162$, RMSE $= 1.14$, RMSE$^* = 0.71$).

A comparison of model performances grouped by reproduction system is shown in Fig. 6. There is almost no difference between model scores in systems with fewer channels than the reference. As the number of channels increase, differences between scores also increase. This can be related to the number of loudspeaker positions with elevations different than zero, to the point that the largest difference is seen in cuboid system, where $\phi \neq 0°$ in every channel.

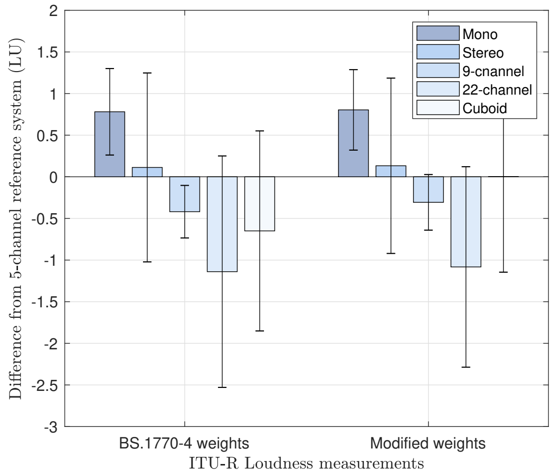

Fig. 6 – Differences between loudness measurements (LU) in relation to 5.1 reference system, broken down into reproduction systems.

5. CONCLUSION

Discussions on further development of ITU-R BS.1770 multichannel loudness model to address object and scene-based audio are taking place in Radiocommunication Sector Study Groups. Originally designed for stereo and 5.1 content, the algorithm was extended to an unrestricted number of channels in its latest update. However, it has no directional weighting for broader elevation angles and the method used to estimate its weighting coefficients was based on binaural summation gains derived from subjective data on narrowband sounds.

This paper presented an alternative set of directional weights from subject data on broadband sounds reproduced at different azimuths and elevations from the listener. Directional loudness sensitivities from listeners, sound pressure level measurements at the ears of a dummy head placed in the listener position, and loudness measurements in binaural recordings of reproduced stimuli were inputs to two weight estimation approaches.

The optimization approach was an attempt to reproduce the method that derived a binaural gain used to estimate directional weighing in ITU-R BS.1770-4. Despite the fact that the method yielded reasonable results, it did not provide any insights on elevation effects. On the other hand, a regression model using localization cues as predictors resulted in a better modeling of directions with $|\phi| \geq 30°$.

[1]Loudspeakers at ±45° and ±135° azimuth, ±30° elevation. $g(\mathrm{B} \pm 135) = g(\pm 135°, -30°) \approx 0.00$ dB.

A modified version of the loudness algorithm including these new weights was tested fairly on the collected data and against its benchmark on immersive audio content rendered to different spatial reproduction layouts. This modified algorithm performed better than ITU-R BS.1770-4 based on correlations with subjective data and Epsilon insensitive RMSE (RMSE*) measurements. Also, a relationship between better performance scores and reproduction systems with more loudspeakers positioned out of the horizontal plane was observed.

ACKNOWLEDGMENT

This research was partially supported by Agência Nacional de Telecomunicações (Brazil). The authors would like to thank the subjects for their time in participating in the loudness matching test.

REFERENCES

[1] Radiocommunication Sector of ITU (2019). Recommendation ITU-R BS.2127-0. Audio Definition Model renderer for advanced sound systems. https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.2127-0-201906-I!!PDF-E.pdf

[2] Radiocommunication Sector of ITU (2015). Recommendation ITU-R BS.1770-4. Algorithms to measure audio programme loudness and true-peak audio level. https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1770-4-201510-I!!PDF-E.pdf

[3] Scott G. Norcross, Sachin Nanda, and Zack Cohen (2016). ITU-R BS.1770 based Loudness for Immersive Audio. Audio Engineering Society Convention 140. http://www.aes.org/e-lib/browse.cfm?elib=18199

[4] Tomoyasu Komori, Satoshi Oode, Kazuho Ono, Kensuke Irie, Yo Sasaki, Tomomi Hasegawa and Ikuko Sawaya (2015). Subjective loudness of 22.2 multichannel programs. Audio Engineering Society Convention 138. http://www.aes.org/e-lib/browse.cfm?elib=17643

[5] D. W. Robinson and L. S. Whittle (1960). The loudness of directional sound fields. Acta Acustica united with Acustica, v. 10, n. 2, p. 74-80. https://www.ingentaconnect.com/content/dav/aaua/1960/00000010/00000002/art00004

[6] Ville Pekka Sivonen and Wolfgang Ellermeier (2006). Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation. The Journal of the Acoustical Society of America 119, 2965. http://dx.doi.org/10.1121/1.2184268

[7] Ville Pekka Sivonen and Wolfgang Ellermeier (2006). Effect of Direction on Loudness for Wideband and Reverberant Sounds. Audio Engineering Society Convention 120. http://www.aes.org/e-lib/browse.cfm?elib=13625

[8] Ville Pekka Sivonen and Wolfgang Ellermeier (2008). Binaural Loudness for Artificial-Head Measurements in Directional Sound Fields. JAES Volume 56 Issue 6 pp. 452-461. http://www.aes.org/e-lib/browse.cfm?elib=14394

[9] Radiocommunication Sector of ITU (2018). Recommendation ITU-R BS.2051-2. Advanced sound system for programme production. https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1770-4-201510-I!!PDF-E.pdf

[10] Radiocommunication Sector of ITU (2015). Recommendation ITU-R BS.1116-3. Advanced sound system for programme production. https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1116-3-201502-I!!PDF-E.pdf

[11] Russell Mason (2016). Installation of a Flexible 3D Audio Reproduction System into a Standardized Listening Room. Audio Engineering Society Convention 140. http://www.aes.org/e-lib/browse.cfm?elib=18160

[12] Brian C. J. Moore (2012). An introduction to the psychology of hearing (sixth edition). Bingley: Emerald Group Publishing.

[13] Phillip Jackson, Martin Dewhirst, Robert Conetta, Slawomir Zielinski, Francis Rumsey, David Meares, Søren Bech and Sunish George (2008). QESTRAL (Part 3): System and metrics for spatial quality prediction. Audio Engineering Society Convention 125. http://www.aes.org/e-lib/browse.cfm?elib=14748

[14] Brian C. J. Moore, Brian R. Glasberg, Ajanth Varathanathan and Josef Schlittenlacher (2016). A Loudness Model for Time-Varying Sounds Incorporating Binaural Inhibition. Trends in Hearing, v. 20, p. 1–16. https://doi.org/10.1177/2331216516682698

[15] Jon Francombe, Tim Brookes, Russell Mason and Frank Melchior (2015). Loudness Matching Multichannel Audio Program Material with Listeners and Predictive Models. Audio Engineering Society Convention 139. http://www.aes.org/e-lib/browse.cfm?elib=18020

[16] Jon Francombe, Tim Brookes, Russell Mason, Rupert Flindt, Philip Coleman, Qingju Liu, and Phillip Jackson (2015). Production and Reproduction of Program Material for a Variety of Spatial Audio Formats. Audio Engineering Society Convention 138. http://www.aes.org/e-lib/browse.cfm?elib=17634

[17] Standardization Sector of ITU (2013). Recommendation ITU-T P.1401. Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models. https://www.itu.int/rec/T-REC-P.1401-201207-I/en