Source: HEAD acoustics GmbH

Title: Loudness tests for ATIAS

Document for: Discussion

# Introduction

In the scope of the ATIAS work item, new measures for loudness of immersive audio have to be identified. Traditional send and receive loudness ratings (SLR, RLR) according to ITU-T P.79 [1] utilize a simple linearized loudness model for monaural telephony and may thus not be the optimum choice for these applications.

An alternative for receive direction might be the loudness calculation method of Recommendation ITU-T P.700 [2], "Calculation of loudness for speech communication", which is based on the time-varying loudness according to ISO 532-1 [3] and provides an integral loudness scale for handset/headset (monaural or binaural) as well as any hands-free mode, and is not limited to a certain audio bandwidth (from NB to FB).

# Loudness in RCV

## Introduction

In receive direction, it can be assumed that immersive communication terminals provide acoustical playback[[1]](#footnote-1)via loudspeakers (integrated/external) or headphones, which can both be captured with a head-and-torso-simulator (HATS), as illustrated in Figure 1.

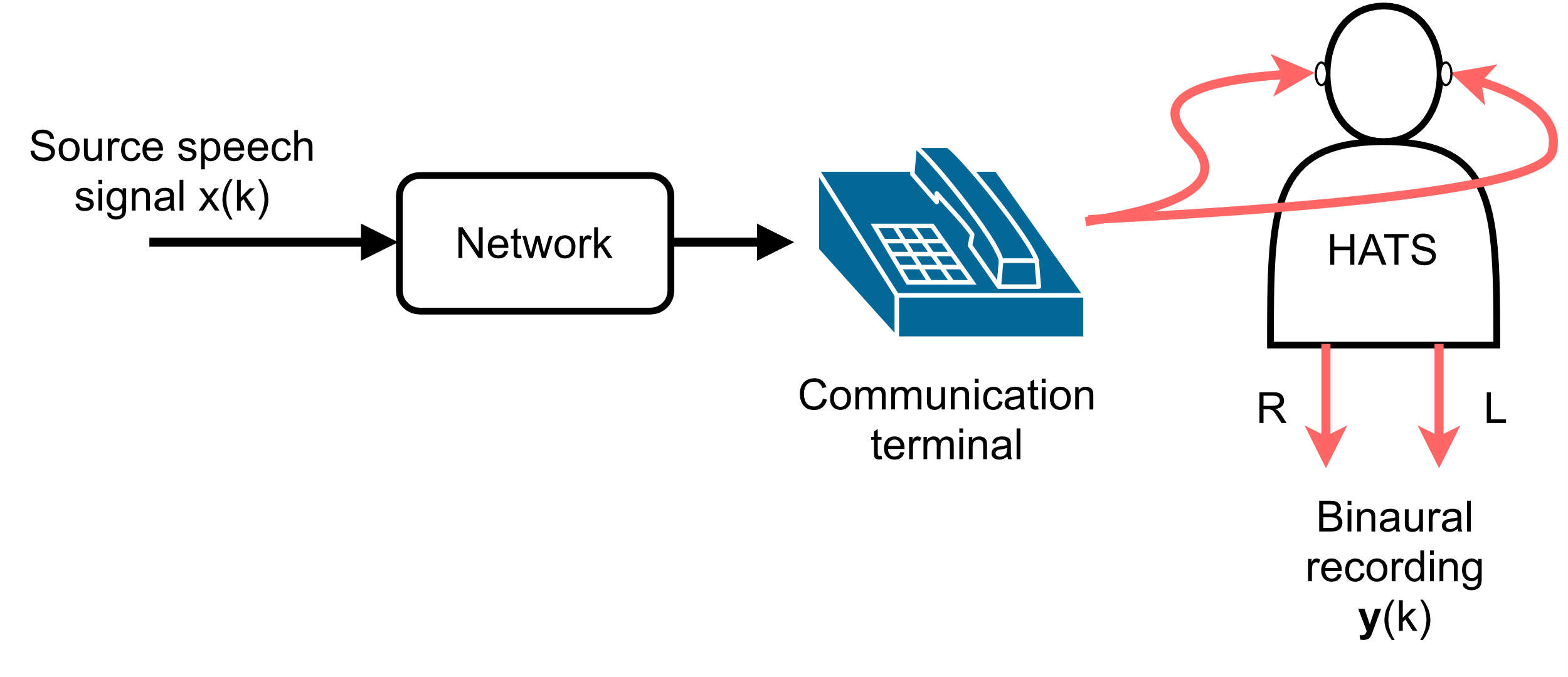


Figure : Simplified test setup for acoustic loudness measurement in RCV (from [2])

If the terminal provides only an electrical interface (analogue or digital), either the IVAS-default or a custom binaural renderer could be used to generate corresponding binaural signals for analysis, as illustrated in Figure 2. Since it is typically desired to test a certain IVAS input format, this rendering step should always be applied after the electrical interface, i.e., as part of the signal analysis – even if the terminal could directly be configured for stereo/binaural output.

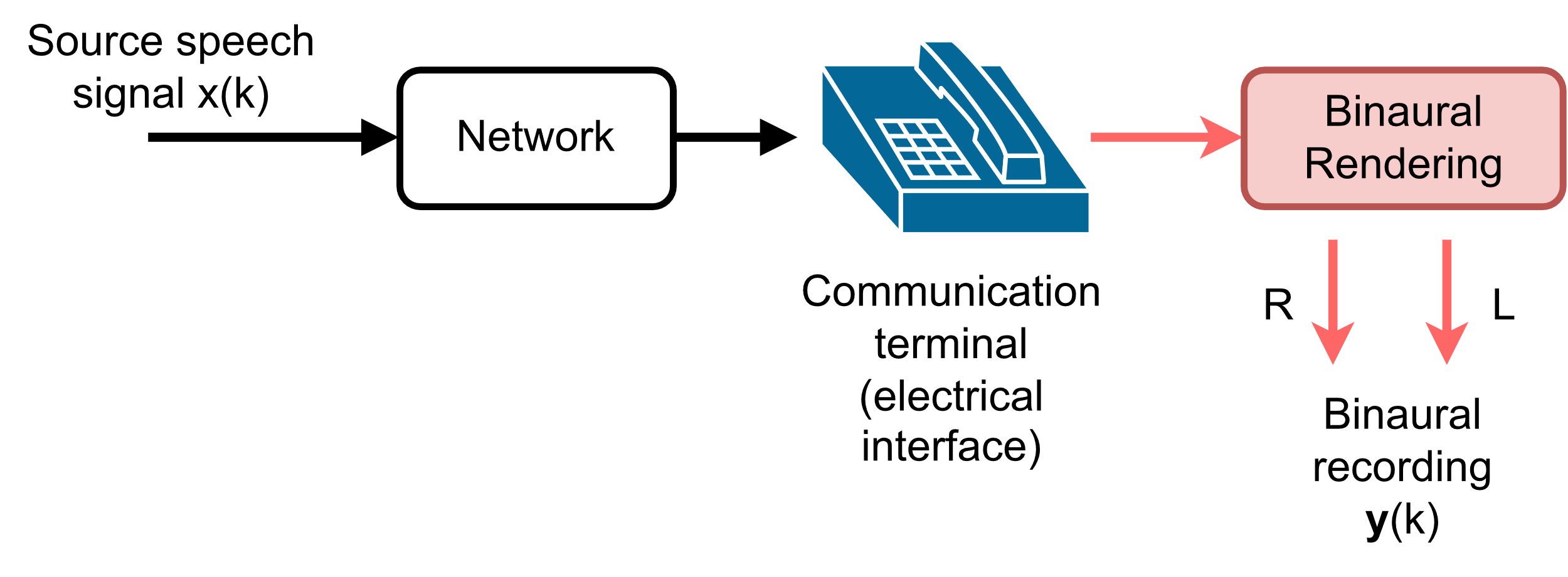


Figure : Simplified test setup for electrical loudness measurement in RCV (adapted from [2])

## Loudness according to ITU-T P.700

The calculation of loudness according to ITU-T P.700 [2] is based on the time-varying loudness ISO 532-1 [3], which provides a loudness-vs-time curve in the unit *sone* with a temporal resolution of 500 Hz. According to ISO 532-1, it is always assumed that a single-channel input to the loudness model would be diotically presented to subjects in a listening test.

Two extensions are described in ITU-T P.700, which are different or not specified in ISO 532-1:

1) As a temporal aggregation value, the arithmetic mean (average) over loudness-vs-time is chosen instead of the 95% percentile (N5).

2) Similar as in ITU-T P.56 [4], only signal ranges of active speech are considered for the temporal aggregation. Pauses longer than 400 ms are excluded.

The average loudness value versus time is first determined in sone. The loudness level (in unit *phon*) is finally reported by transforming the loudness value via equations (1)-(3) in clause 5.3 of ISO 532-1. An overview of the algorithm is shown in Figure 3.

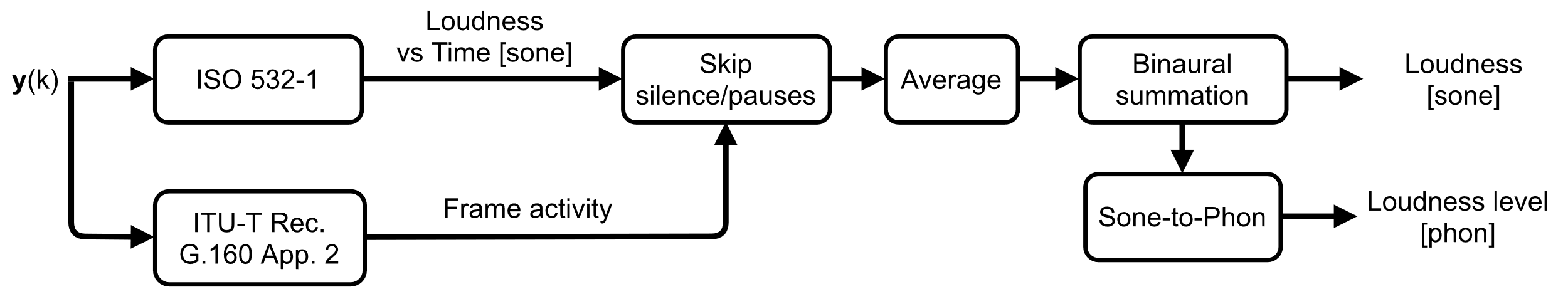


Figure : Analysis of P.700 loudness (from [2])

## Loudness level vs Loudness rating

RLRs (and also SLRs) according to Recommendation ITU-T P.79 [1] are widely used in measurement specifications to quantify the loudness of terminals by means of a linearized loudness model. These were designed and evaluated several decades ago and are based on "traditional" NB telephony and was later adapted/extrapolated for WB. Instead of developing another adaption/extrapolation for SWB and FB applications, ITU-T SG12 decided to work on a new universal loudness model for speech communication terminals, which should overcome the following drawbacks of loudness ratings by using a state-of-the-art loudness prediction model:

- LRs were developed mainly as a fundamental parameter for transmission planning and is actually an attenuation to an implicit reference listening system.

- The original development of loudness ratings considered mainly the frequency responses of handsets, determined over a long-term test signal. Typical modern signal processing components like e.g., AGC, compressor, noise reduction, etc., were not available at that time.

- On the other hand, the time-varying loudness according to ISO 532-1 as a base analysis of ITU-T P.700 is in general able to detect short-time signal impairments (see also results from P.700 validation experiments in Appendix I-III of [2]).

- Binaural extensions of RLRs (for e.g., headsets) were also only adapted/extrapolated from the monaural listening situation. Even though the binaural loudness summation that is used in P.700 might not be a highly sophisticated approach, but it is at least better documented in literature than just assuming a 6 dB offset.

- Since loudness perception is an absolute measure independent of the bandwidth, there should be no distinction between NB, WB, and SWB/FB transmission systems for the loudness measure. The aforementioned non-linear and/or short-term degradations have most likely also more influence on loudness than the bandwidth.

In the terminal measurement specifications ETSI TS 103 737/738/739/740 (handset/hands-free terminals, for NB/WB, [5] [6] [7] [8]), the P.700 loudness level was already adopted (complementary to LRs). In a previous study to this update [9], a comparison between LRs and P.700 loudness levels (in phon) was made for different types of devices:

- Handset (HA)

- Binaural headset (HE-Bin)

- Monaural headset (HE-Mon)

- Table-mounted speakerphones (SP)

The results are shown as a scatter plot in Figure 4. Note that the RLR results of the different RLR calculation methods for hands-free, binaural headset, and monaural handset/headset were corrected for a common target RLR value of +2 dB. The resulting linear regression between loudness level in receive (LLR) and RLR was then determined according to equation 1:

(1)

This relation between LLR and RLR was used to derive a nominal loudness level from the "usual" nominal RLR requirement of +2 dB, which corresponds to ~74 phon. The equation can also be used to determine limits for the nominal range, e.g., RLR = +2 dB ± 3 dB would correspond to a range of LLR = ~71-77 phon.

Note that in [5] [6] [7] [8] the nominal loudness level range was slightly extended (± 4 phon) to consider the variance between LLR and RLR in this study, i.e., devices with a RLR in the nominal range should also obtain a nominal LLR.

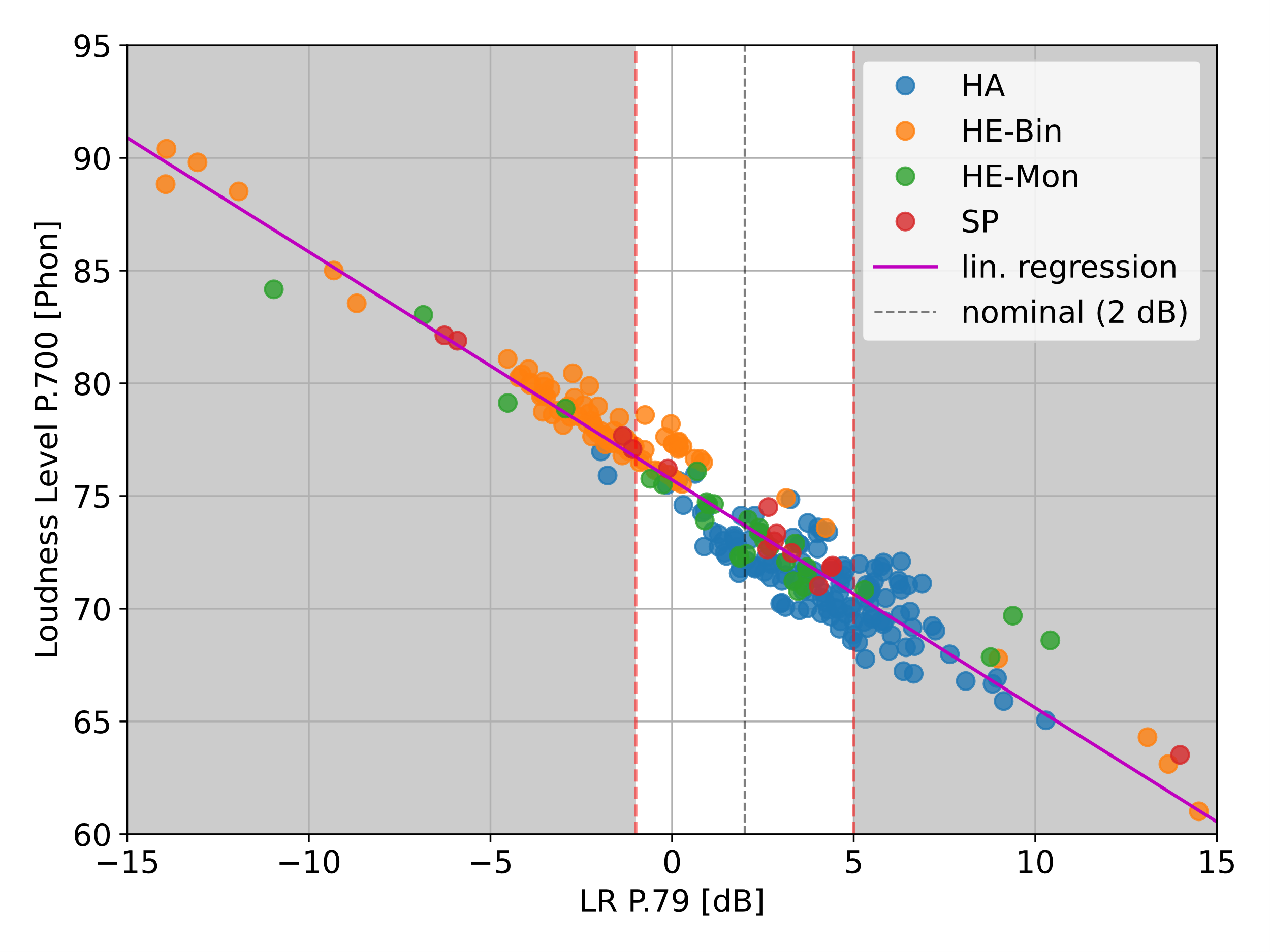


Figure : P.700 loudness level vs RLR (from [9])

## Proposal for test method and requirement

Editor's Note: Test method and requirements are in general applicable to all IVAS formats.

### Requirement

The nominal value of Loudness Level in Receive (LLR) shall be [75] ± [3-4] phon for all UE types. In case a user controlled receive volume control is provided, for at least one setting of the control the LLR shall meet the nominal value.

When the control is set to maximum, the LLR shall not be louder than [89] phon. With the volume control set to the minimum position the LLR shall not be quieter than [52] phon and should not be quieter than [58] phon.

Editor's Note: Values calculated based on the RLR vs LLR equation above.

Performance requirements and objectives apply for source position of azimuth 0° and elevation 0° in the test signal and should also be evaluated for the source positions listed in Table 1.

Table 1: Additional source positions for loudness

|  |  |
| --- | --- |
| Azimuth [°] | Elevation [°] |
| 0 | +90 |
| 0 | -90 |
| 0 | 180 |
| +90 | 0 |
| -90 | 0 |

### Test method

1) The test signal to be used for the measurements shall be the British-English single talk sequence described in clause 7.3.2 of Recommendation ITU-T P.501 [xx], calibrated to an active speech level according to Recommendation ITU-T P.56 [xx] of [-26 dBov / TBD].

Editor's Note: The level calibration is a preparation for the following rendering step, so ASL in dBov is correct?

2) The source signal for the measurement is generated by virtually positioning the mono test signal at azimuth 0°, and elevation 0° and a distance of 1.0 m and then render it to the corresponding IVAS input format of the UE.

Editor's Note: Add better description / reference to IVAS command line tools?! (use ISM with metadata?)

3) The source signal is calibrated to a [level/loudness] of [xx dB].

Editor's Note: How to define default levels for the source signals? See also loudness in send…

4) The UE is setup according to clause(s) [xx], the source signal is encoded by the reference client, and inserted at the POI to the UE.

5) The capture of the UE output is carried out via …

a) acoustical interface (headphones or loudspeakers): recording via diffuse-field equalized HATS.

b) electrical interface: recording via corresponding reference interface. [If the captured audio format is not stereo,] the default IVAS binaural renderer [xx] is used to generate a binaural signal.

Editor's Note 1: Assume that stereo == binaural and directly analyse it – or can/should it also be rendered to binaural?

Editor's Note 2: Add better description / reference to IVAS default renderer. We might need an additional factor to scale the renderer output to the acoustic domain (like e.g., 73 dB SPL == -21 dB Pa == -26 dBov 🡪 apply +5 dB)?

6) The LLR in phon is calculated according to clause 8.3.3 of Recommendation ITU-T P.700 with the captured or rendered binaural signal.

7) The same binaural signal should be used to calculate Receive Loudness Rating (RLR) according to Recommendation ITU-T P.79 [xx] for comparison to 3GPP TS 26.131/132.

a) The inverse diffuse-field correction according to Recommendation ITU-T P.58 is applied on left and right channel of the recording to obtain the signal at DRP. Then DRP-to-ERP correction is applied.

b) The reference signal used for the RLR calculation is the original test signal specified in step 1), calibrated to ‑16 dBm0.

c) The RLR is calculated according to clause 8.2.3.2 of 3GPP TS 26.132.

Editor's Note: RLR may be calculated optionally from the same recording?

8) Steps 2-7 should be repeated for additional source positions as described in Table 1.

# Loudness in SND

## Introduction

A loudness measure for send direction was not proposed or discussed for ATIAS-1 [10]. Several options are described in the following sections, which all assume that an encoded IVAS bitstream is captured at the point-of-interconnect (POI) and decoded by a reference client.

## ITU-R BS.1770 (+ rendering)

The calculation according to ITU-R BS.1770 provides loudness/level results in LKFS (Loudness, K-weighted, relative to nominal full scale) and can be applied for stereo/binaural and channel-based. However, for MASA and scene-/object-based audio, an additional step would be required to render these formats into a certain channel-based format. A similar approach was already used in the processing of listening test samples that were generated in the selection phase (see IVAS-7a/b) The block diagram of this method is shown in Figure 5.

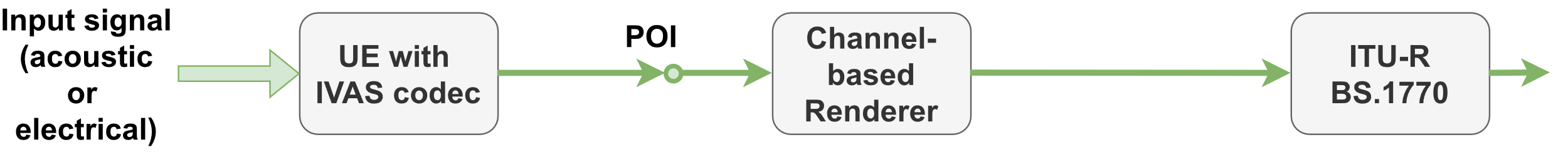


Figure : BS.1770 LKFS calculation with channel-based rendering

## ITU-T P.700 in send mode

The Recommendation ITU-T P.700 also provides a loudness calculation for the send direction. This approach assumes that a mono signal is analyzed, which is then calibrated with a default receive sensitivity in order to obtain a (pseudo‑)acoustical signal. This would require an additional mono downmixing step (see section 3.5), as shown in the block diagram of Figure 6.

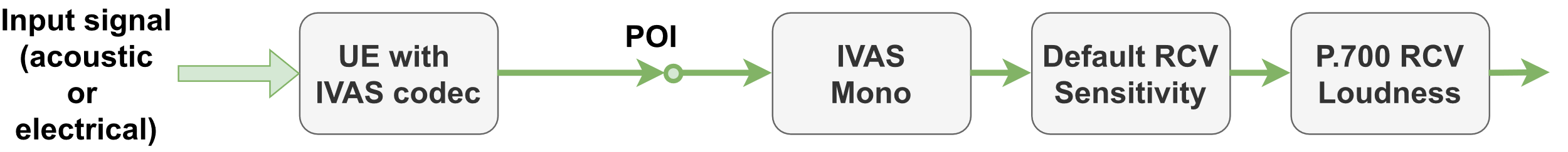


Figure : P.700 loudness calculation for sending

## ITU-T P.700 with binaural rendering

As an alternative to the default send loudness calculation, a binaural renderer (default?) could be applied on the immersive audio format to obtain a binaural signal prior to the actual analysis. The P.700 loudness calculation would then be applied as for the receive direction, as shown in the block diagram in Figure 7.

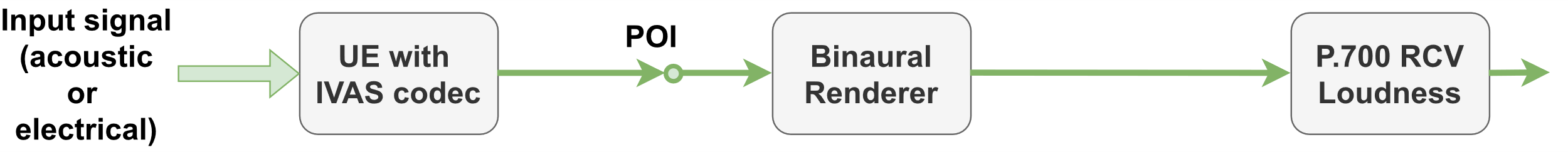


Figure : P.700 loudness calculation with (default) binaural rendering

## IVAS Mono-Downmix

The following approach was originally proposed in [11] (but for the receive direction): the IVAS mono mode (EVS interoperable mode) is either applied on the immersive audio to generate a mono signal – or the reference client directly decodes/renders the received bitstream to mono. The resulting mono signal can then be used to calculate "traditional" metrics like send loudness rating (SLR) or active speech level according to ITU-T P.56 [4], as shown in Figure 8.

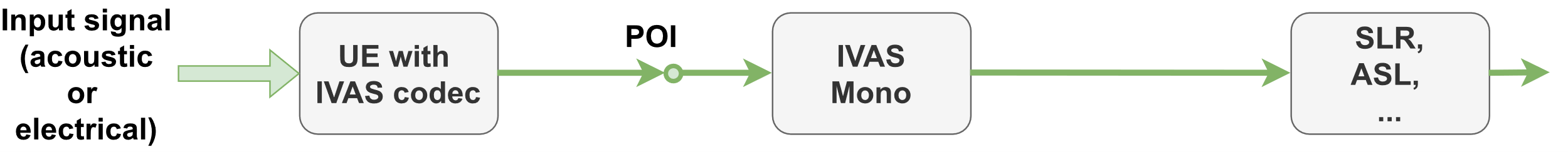


Figure : SLR/Level calculations based on IVAS mono downmix

# Conclusion

Proposals and an overview of possible loudness tests for receive and send direction were made. The source suggests considering the input of the present document for the next update of the permanent document ATIAS-1 [10].

In receive, the loudness according to P.700 has several benefits over the traditional RLR values and is applicable for test setups that include acoustic playback of immersive audio. For electrical interfaces, an additional binaural renderer has to be used.

In send, there are several options to calculate the loudness of transmitted immersive audio signals. These methods should also be considered in general for calibrating test signals in all receive tests currently proposed in ATIAS-1.

# References

|  |  |
| --- | --- |
| [1] | Recommendation ITU-T P.79, „Calculation of loudness ratings for telephone sets,“ 11/2007. |
| [2] | ITU-T Recommendation P.700, „Calculation of loudness for speech communication,“ 06/2019. |
| [3] | ISO 532-1, „Methods for calculating loudness - Part 1: Zwicker method,“ 2016. |
| [4] | ITU-T Recommendation P.56, „Objective measurement of active speech level,“ 12/2011. |
| [5] | ETSI TS 103 737 v1.4.1, „Speech and multimedia Transmission Quality (STQ); Transmission requirements for narrowband wireless terminals (handset and headset) from a QoS perspective as perceived by the user,“ 2021-10. |
| [6] | ETSI TS 103 738 v1.4.1, „Speech and multimedia Transmission Quality (STQ); Transmission requirements for narrowband wireless terminals (handsfree) from a QoS perspective as perceived by the user,“ 2021-10. |
| [7] | ETSI TS 103 739 v1.4.1, „Speech and multimedia Transmission Quality (STQ); Transmission requirements for wideband wireless terminals (handset and headset) from a QoS perspective as perceived by the user,“ 2021-10. |
| [8] | ETSI TS 103 740 v1.4.1, „Speech and multimedia Transmission Quality (STQ); Transmission requirements for wideband mobile wireless terminals (hands-free) from a QoS perspective as perceived by the user,“ 10/2021. |
| [9] | ETSI STQ(21)067014, „ITU-T P.700 Loudness for speech communication terminals,“ HEAD acoustics GmbH, 06/2021. |
| [10] | 3GPP S4-231418, „Permanent Document ATIAS-1,“ 3GPP SA4 Audio SWG. |
| [11] | 3GPP S4aA230108, „ATIAS receive loudness and frequency sensitivity characteristics,“ Dolby Sweden AB. |
| [12] | 3GPP S4-231701, „UE classification and test structure for ATIAS,“ HEAD acoustics GmbH. |

1. This assumption is based on the proposals made in S4-231701 [13]. [↑](#footnote-ref-1)