

Technical study on transmission of sound and image through cloud-based systems for remote interpreting in simultaneous mode (Remote Simultaneous Interpreting-RSI)

Technical study commissioned by AIIC Technical and Health Committee.

Objective:

The objective of this technical study was the comparative analysis of the characteristics of input signals fed into the system at the source (source signals) and output signals at the sink (output signals transmitted to interpreters), related to essential parameters applicable to audio and combined audio/video signals for simultaneous interpreting, based on the requirements set out in ISO standards 20108 and 20109.

The study was carried out with data transmission of sound and image through the following cloud-based RSI systems:

1. **Interactio (Lithuania)**
2. **Catalava (Greece)**
3. **Olyusei (Spain)**
4. **Voiceboxer (Denmark)**
5. **KUDO (USA)**
6. **Interprefy (Switzerland)**

All platform providers were asked to facilitate log-in data for accessing the platform with the required profiles (speaker, interpreter) and to facilitate technical support for setting up the testing environment properly.

Platform providers were not informed in advance about the specific signal chain and the equipment used for testing by the audio engineers.

According to the objectives of the study, the following parameters were measured:

- **Frequency response** (excitation with Pink Noise and Sweep)
- **Spectrum** (excitation with White Noise),
- **Delay** (measured with XL2 Audioanalyzer),
- **STI value** (calculated on the basis of the impulse response and measured with XL2 Audioanalyzer),
- **Total Harmonic Distortion** (THD+N values).
- **Hearing protection** (Limitation when feeding fast and short impulses into the system)
- **Video quality** (visible artefacts, like blurring or freezing)

Technical setup:

For the testing, the following signal chain was used:

- Signal generator in ARTA Software
- ⇒ M-Audio USB Interface
- ⇒ Yamaha QL1 audio mixer
- ⇒ Lynx PDM 1383 SDI Embedder
- ⇒ Blackmagic Web Presenter
- ⇒ Individual Codec of the system
- ⇒ Internet
- ⇒ Individual Codec of the system
- ⇒ Steinberg UR22 USB Interface
- ⇒ M-Audio USB Interface
- ⇒ Evaluation in ARTA Software.

This signal chain includes several AD / DA transformations, which may have an effect on the precision of the measurements of high frequencies. Nevertheless, this setup was necessary in order to be able to evaluate the quality of transmission of the video signals under real life conditions¹.

Aprox. some 170ms of the measured values for the delay (latency) can be attributed to the use of Blackmagic Web Presenter². For the evaluation of the delay measured for the platform itself these 170ms always need to be deducted. On the other hand, when using the application with video transmission from the interpreter the Web Presenter is indispensable for feeding the video signal coming from an external high quality camera.

¹ For further tests, a special focus shall be put on minimizing the number of AD/DA transformations.

² Blackmagic Web Presenter makes any SDI or HDMI video source appear as a USB webcam for higher quality web streaming using streaming platforms

Results

Overview – set values ISO 20108 / 20109

Parameter	Set values as defined in ISO 20108/20109
Frequency response	125 Hz – 15 kHz
Delay (Latency)	Max. 500ms
Speech intelligibility	STI > 0,64
Total harmonic distortion	< 1 %
Hearing protection	< 94 dBA _{spl} over > 100ms
Video quality	No visible artefacts

Overview – measurements with platforms

Platform	FR - Pink Noise	FR - Sweep	Spectrum - White Noise	Tot. delay (ms)	Delay without Web Pres.	STI - Arta	STI - XL2	THD	Hearing protection
Interactio									
Catalava		XXX							
Olyusei (Zoom+)								XXX	
Voiceboxer									
KUDO (1)									
KUDO (2)*									
Interprefy									
	not compliant								
	compliant								
XXX	no results obtained due to specificities of platform								

* see note on page 20 of the report

General observations:

All platforms make use of a codec for the transmission of sound. All the codecs showed losses to achieve a better bitrate and make possible the transmission via Internet. In some cases, this leads to a strong alteration of the signals fed into the system. In some cases these alterations are constant over time (e.g. high-cut), in others variable in time (elimination of certain frequencies depending on the applied algorithm).

Depending on the applied measurement procedure, very different results could be observed between the different codecs for the same parameters. Very important variations in frequency response can be observed e.g. when measuring with *Pink Noise* or with a *Sweep* (sinusoidal signal with constantly increasing frequency). As the codecs are designed for masking effects, they more or less eliminate the masked frequencies. This means that you will obtain considerably better frequency response with a *Sweep* (where there is always only one single frequency applied with a constant amplitude) than it would be with the *Pink Noise*, where all the frequencies are fully applied at every moment). This means that measurement with *Sweep* shows the best possible frequency response when applying the respective codec, whereas the measurement with the *Pink Noise* would show the worst possible frequency response (*Pink Noise* was selected to underline the masking effect). As

the codec adapts the frequency response to the respective signal dynamically, in real time operation with speech signals you can expect a frequency response situated between these two extremes.

Furthermore, considerable differences between frequency responses of the different codecs and their spectrums were observed. In this context it needs to be explained that the measurement of the spectrum in this case is determined by the averaged sum of the 100 measurements. A continuous white noise signal is applied for excitation. With this measurement the alteration of the signal over time is considered and averaged, as well as quantization noise and dithering. Therefore, the spectrum measurement includes all interfering noise and side effects the codec can produce.

© AIIC 2019 - Reproduction for own use only

Detailed results for individual platforms (see also diagrams):

1. Interactio

The frequency range in the upper range shows important nonlinearities when measuring with Pink Noise. One can deduce from this that the codec carries out important signal alterations, not being compliant with the relevant ISO Standards.

The measured STI value was 0,69, which is compliant with the Standard.

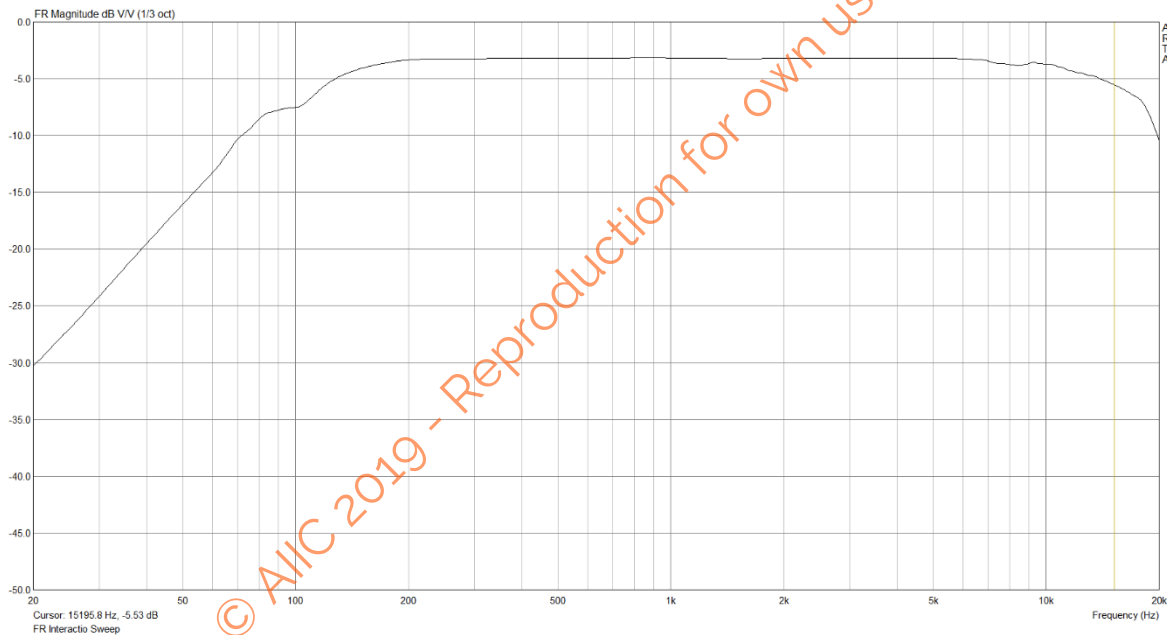
The value calculated from the impulse response is 0,48, situated below the nominal value of 0,63. Nevertheless, this is only an orientation value. Values measured with the XL2 Audioanalyzer are much more precise and reliable.

A delay (latency) of 415ms is compliant with the Standard.

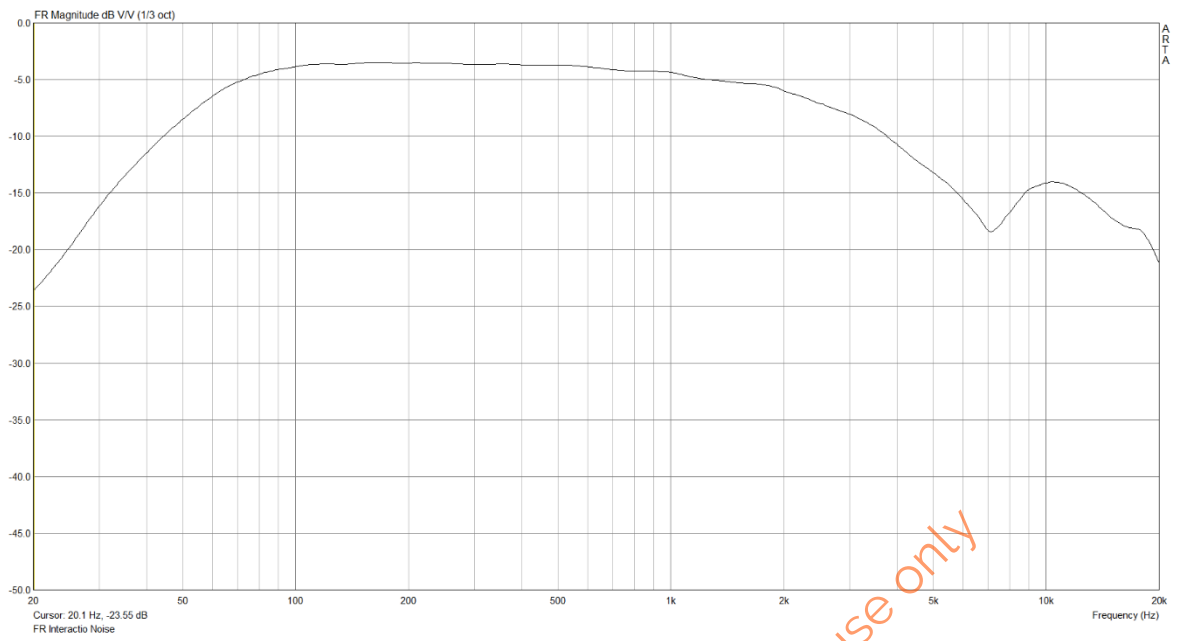
No hearing protection feature was detected. The impulses were transmitted through the platform without compression/limitation.

No visible artefacts were detected during the test.

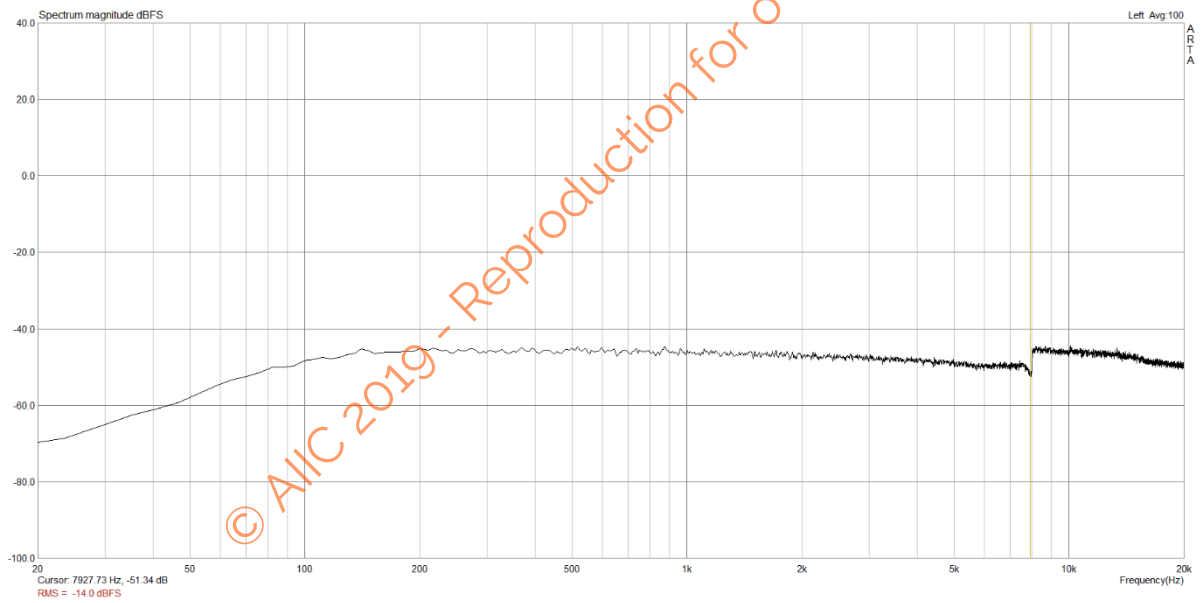
Frequency response with Sweep:



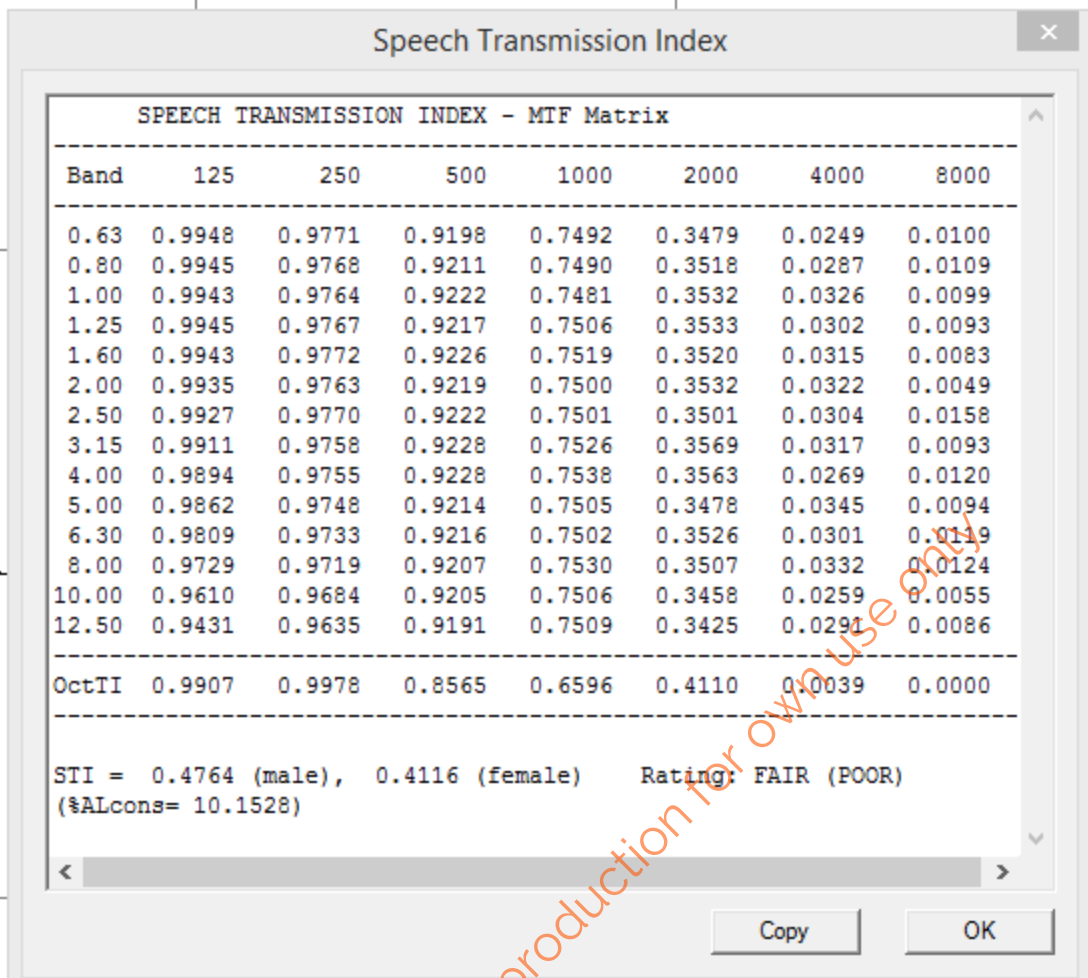
Frequency response with Pink Noise:



Spectrum with White Noise:

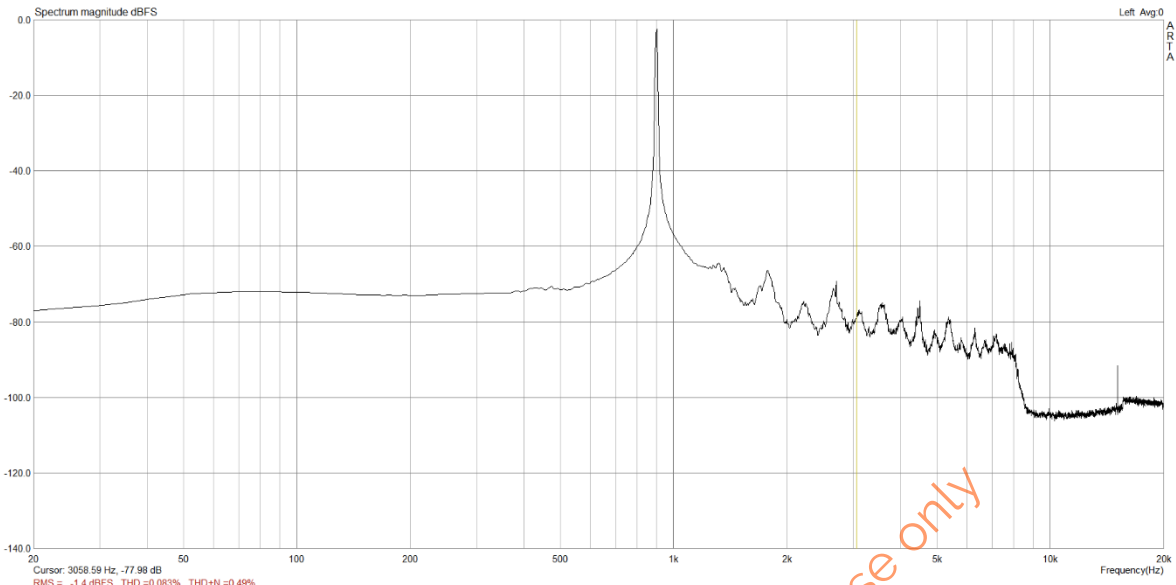


Calculated STI values:



© AIIC 2019 - Reproduction for own use only

Total harmonic distortion (THD + N):

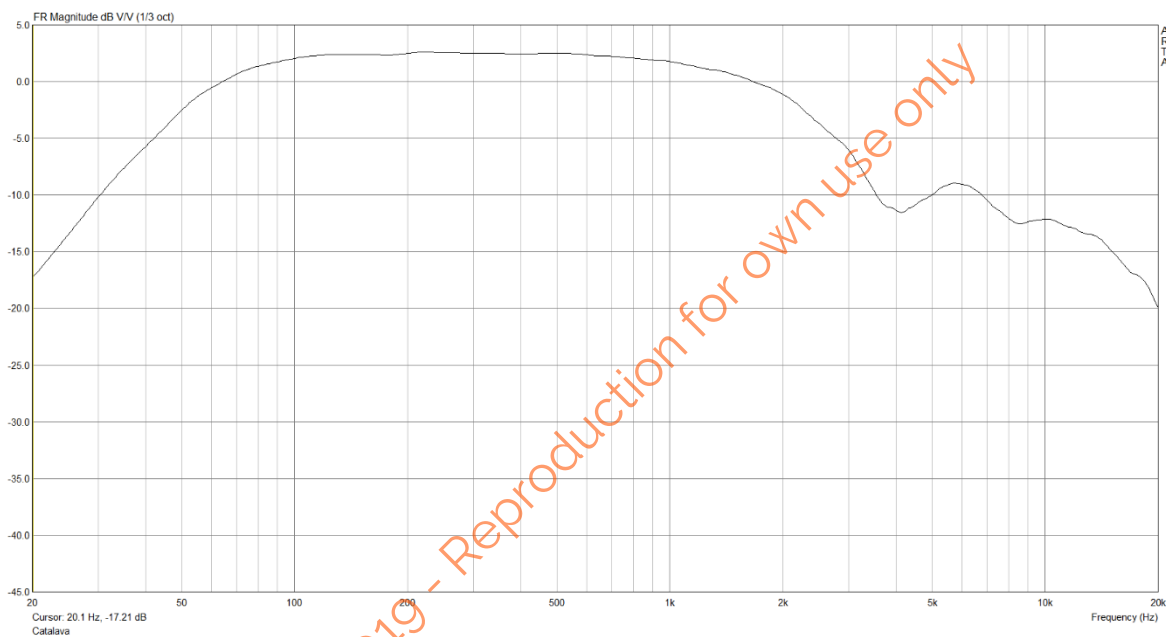


2. Catalava

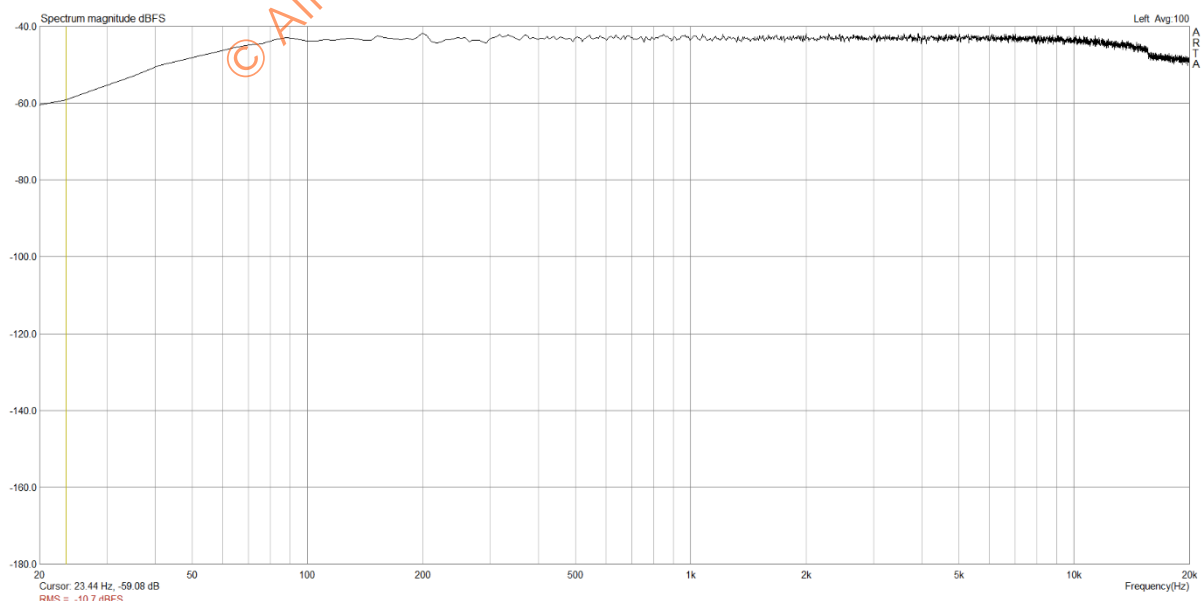
For Catalava, important nonlinearities could be observed. Specifically, important nonlinearities were observed for the frequency response at higher frequencies, which means that this platform is not compliant with the ISO Standards. On the other hand, this platform shows high values for THD+N measurements. The MTF Matrix shows that the modulation depth to the high frequencies gets very low, which means that the calculated STI value is only 0,47. This value, again, is not compliant with the ISO Standards. Measured STI value for this platform is 0,71 , which is compliant with the Standard.

Delay (latency) was measured with 399ms, meeting the requirements of ISO Standards. No hearing protection feature was detected.

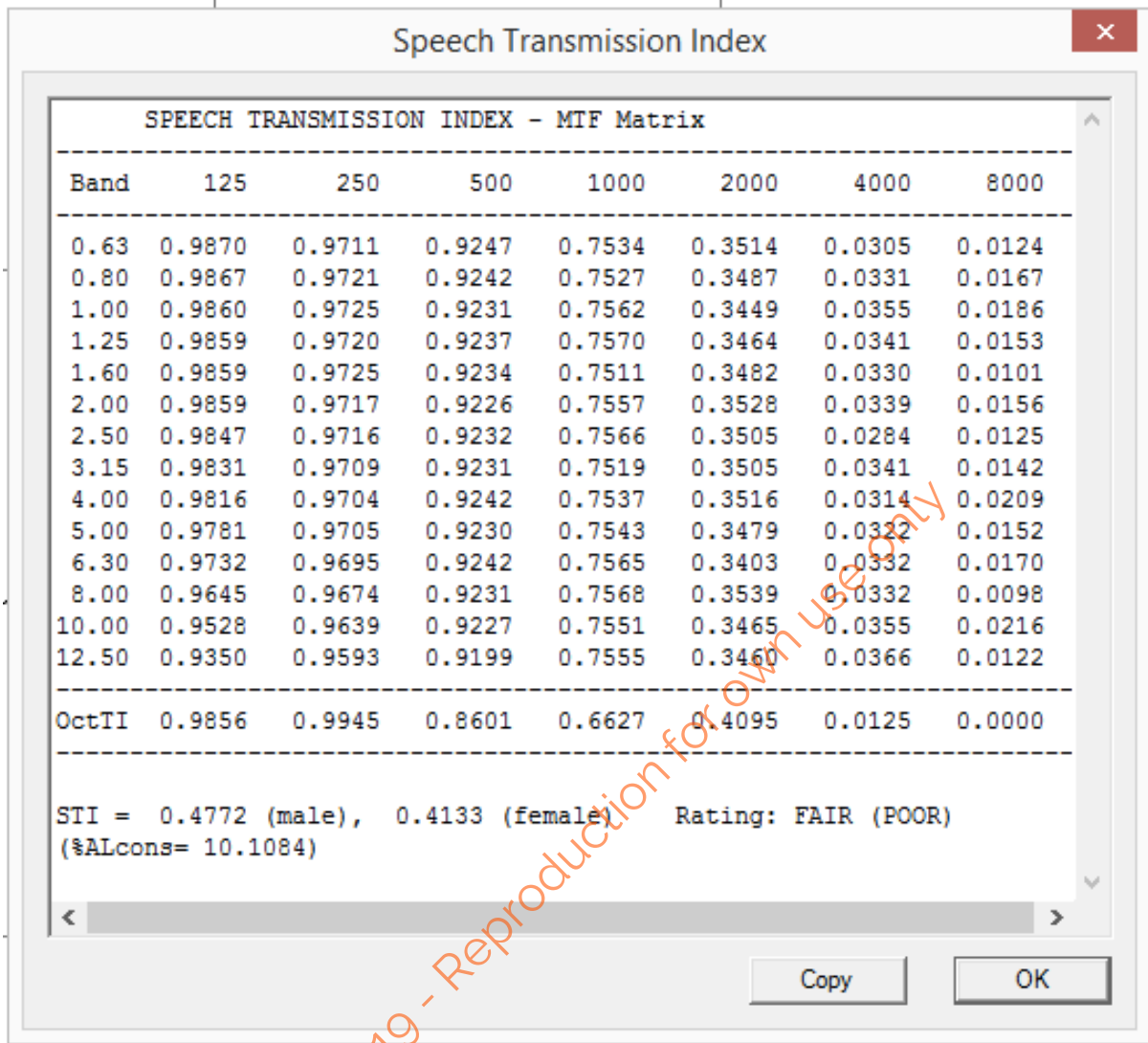
Frequency response with Pink Noise:



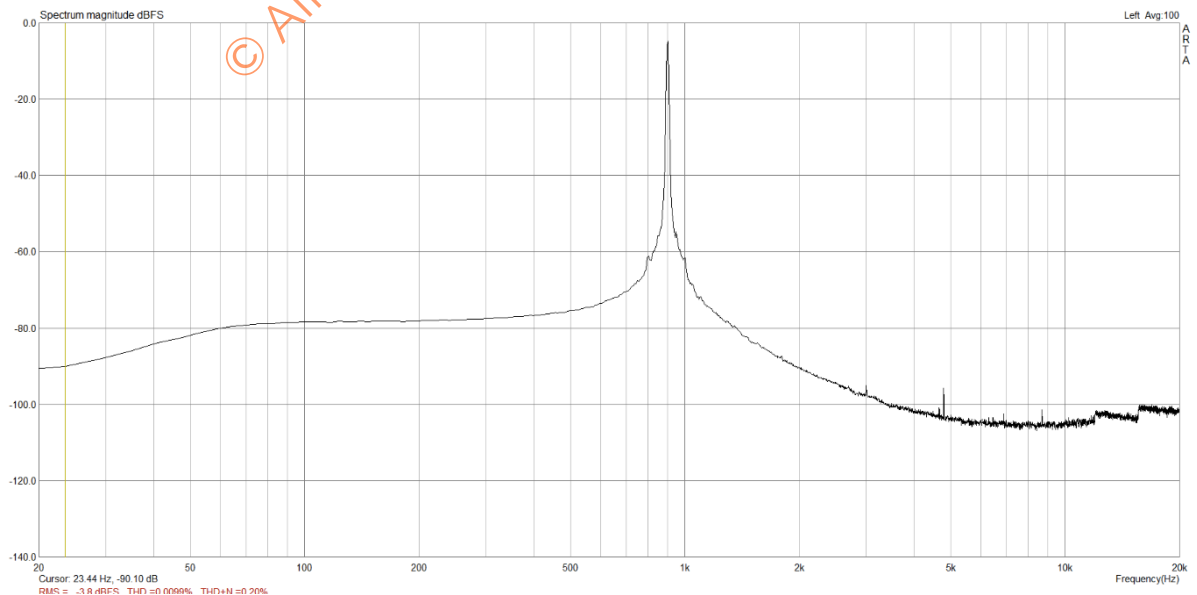
Spectrum with White Noise:



Calculated STI values:



Total harmonic distortion (THD + N):



3. Olyusei

Olyusei is different from all the other platforms tested. This platform manipulates signals significantly. Amongst other effects, we could identify a multiband compressor, feedback suppression and noise cancellation, but no hearing protection against short impulses. The platform constantly applies level adjustment, making impossible a THD+N measurement. That is why there are no values for this parameter. This can also be seen as an advantage, as the input level is kept more or less constantly over time. The platform shows high values for delay (latency) with 886ms. Even when deducting 170ms (for the Web Presenter), the value is still exceeding the nominal 500ms. This means that this platform is not compliant with the ISO Standards.

Up to about 8 kHz frequency is relatively straight. At 8 kHz high cut begins to intervene in order to not exceed the maximum transmitted frequency of 12,5kHz. Also here the frequency response drops out earlier than the nominal value of 15 kHz.

Nevertheless, the difference between the measurement with Sweep and the measurement with Pink Noise is relatively low in this case. This leads to the conclusion that this codec cuts out relatively few frequencies in the upper range.

Both measured and calculated STI values appear to be very high, with 0,96 and 0,99. This leads to the conclusion that despite the limitation of the transmitted frequency range to 12,5 kHz the speech intelligibility of this platform is very good and complies with the requirements set out in ISO Standards 20108/20109. Hence, all STI measurements done with XL2 were qualified as not precise³. This is certainly due to the fact that the platform executes signal processing continuously, falsifying the measurement results. Report of measurements with XL2 are attached.

No visible artefacts were detected during the test.

³ Regarding possible lack of precision of measurement of STI values one can find the following statements in the user manual of XL2 Audioanalyzer:

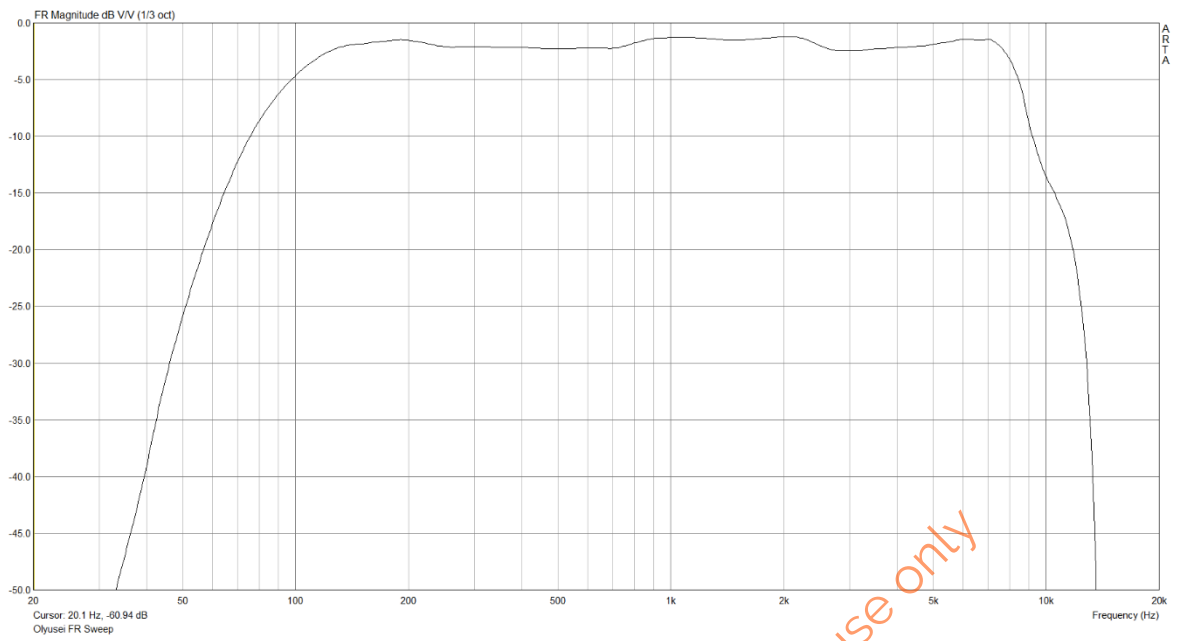
The measuring device automatically checks plausibility of individual results obtained. Thereby, unvalid measurements, primarily caused by impulsive ambient noise, might be detected.

Specifically, XL2 verifies:

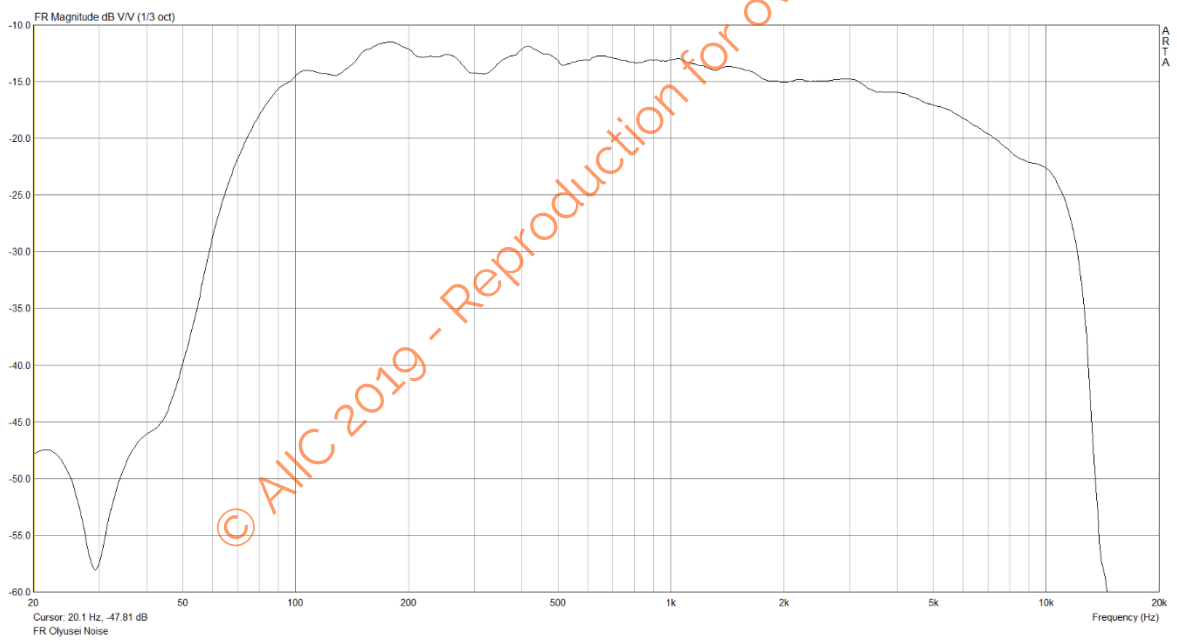
- invalid modulation relations in individual octave bands ($mr1$ or $mr2 > 1,3$)
- unsteady level ratios or impulsive conditions while measuring (therefore, a comparison between the first and the second half of the measuring period is made).

Due to the fact that all tested codecs either cut out certain frequencies at a given moment, or dynamically alterate those or permanently readjust the total level, this leads to measurements for certain codecs being marked as erroneous. Measurements were not carried out using a microphone, but directly and electrically, meaning that the STI value cannot be falsified. Furthermore it needs to be considered that measuring of STI was not developed originally for evaluation of codecs, but for measuring of speech intelligibility in public buildings, train stations etc. This means that the STI measuring does not consider the specificities of the codec operation and that the measured values might be altered by this operational circumstances.

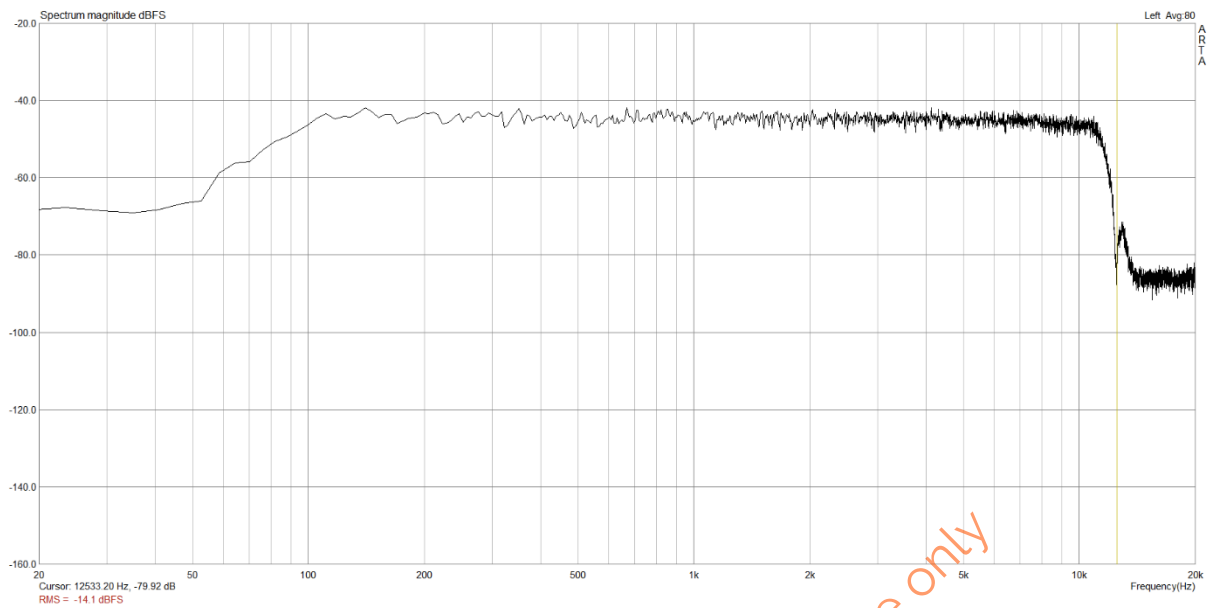
Frequency response with Sweep:



Frequency response with Pink Noise:



Spectrum with White Noise:



STI values calculated with ARTA:

Speech Transmission Index

SPEECH TRANSMISSION INDEX - MTF Matrix							
Band	125	250	500	1000	2000	4000	8000
0.63	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
0.80	0.9996	0.9999	0.9999	0.9999	0.9999	0.9999	0.9998
1.00	0.9996	0.9999	0.9999	0.9999	0.9999	0.9999	0.9998
1.25	0.9991	0.9998	0.9999	0.9999	0.9999	0.9999	0.9997
1.60	0.9986	0.9997	0.9999	0.9999	0.9999	0.9999	0.9995
2.00	0.9979	0.9985	0.9999	0.9999	0.9999	0.9999	0.9992
2.50	0.9971	0.9993	0.9999	0.9999	0.9999	0.9999	0.9989
3.15	0.9951	0.9988	0.9998	0.9999	0.9999	0.9999	0.9982
4.00	0.9926	0.9982	0.9997	0.9999	0.9999	0.9999	0.9972
5.00	0.9880	0.9971	0.9995	0.9999	0.9999	0.9999	0.9954
6.30	0.9802	0.9952	0.9991	0.9998	0.9999	0.9998	0.9922
8.00	0.9706	0.9928	0.9986	0.9997	0.9999	0.9997	0.9884
10.00	0.9530	0.9883	0.9977	0.9994	0.9998	0.9995	0.9812
12.50	0.9285	0.9819	0.9964	0.9991	0.9997	0.9994	0.9707
OctII	0.9862	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

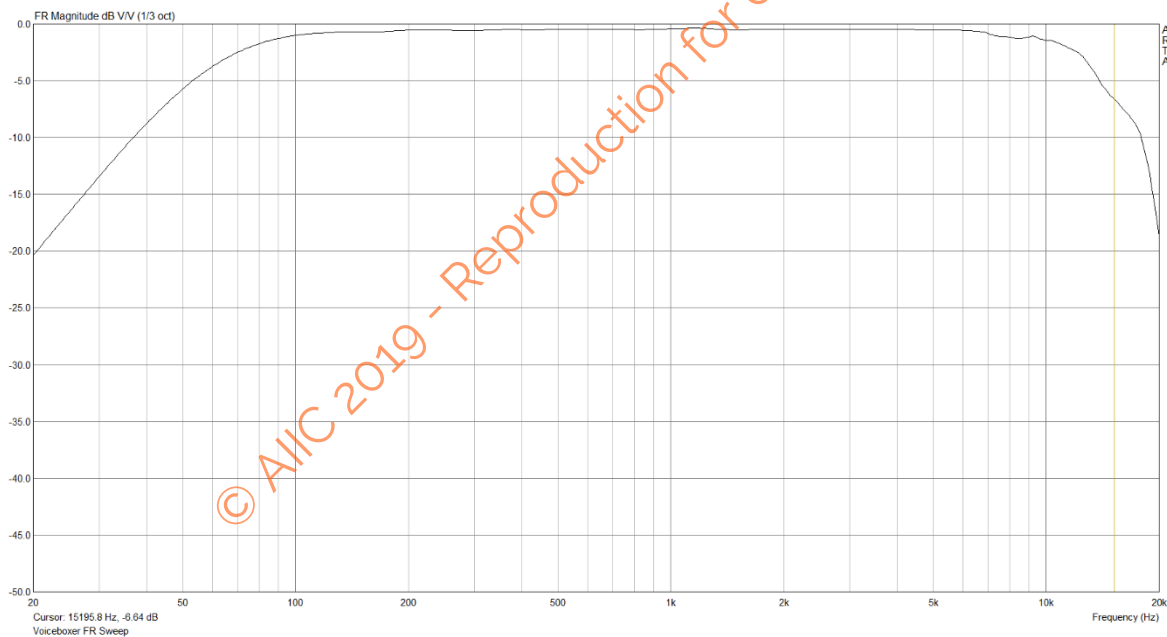
STI = 0.9994 (male), 1.0000 (female) Rating: EXCELLENT (EXCELLEN
(\$ALcons= 0.7631)

Copy OK

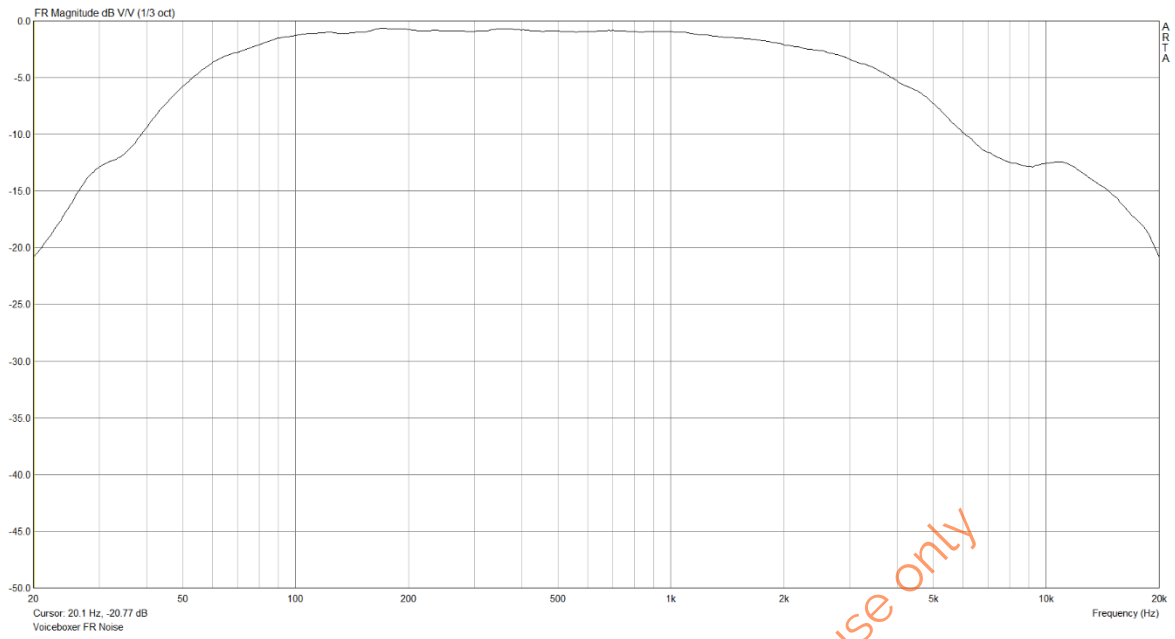
4. Voiceboxer

Voiceboxer shows relatively few differences between the frequency response measured with Pink Noise and the frequency response measured with Sweep. On the other hand, we can observe a definition of the corner frequency of a low or high cut of 3dB. For Voiceboxer level drop when making measurement of frequency response with Sweep is situated at 6dB at 15 kHz, which means that Voiceboxer misses the nominal values very sharply. A better signal chain could possibly lead to measured values compliant with the Standard, although this refers only to the measurement with Sweep, representing the ideal case. When measuring with Pink Noise, Voiceboxer would not be compliant with Standard even if there was an ideal signal chain. Both STI values are noticeable. The calculated STI value is of 0,99 and the measured one is situated at 0,69. This is the only platform where the measured value is clearly worse than the calculated value. Most probably this is a particular feature of this codec. Nevertheless, both values are compliant with the ISO Standards. In addition, Voiceboxer shows good values for the THD measurement, situated at 0,23 %. The delay (latency) is 412ms, being ISO compliant. No hearing protection feature could be detected for Voiceboxer. No visible artefacts were detected during the test.

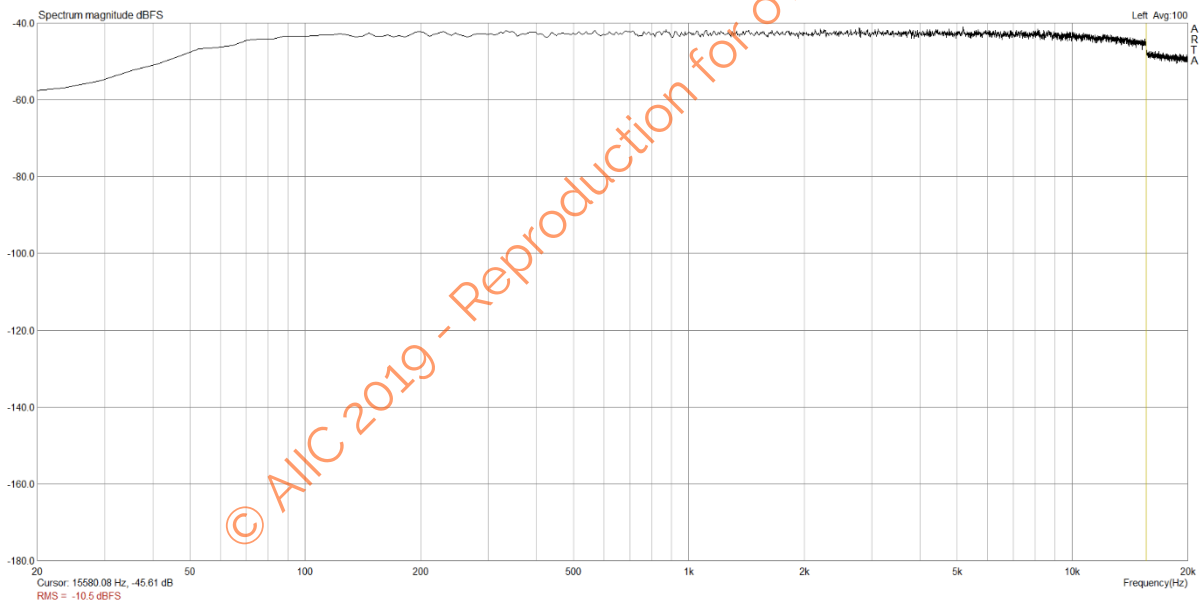
Frequency response with Sweep:



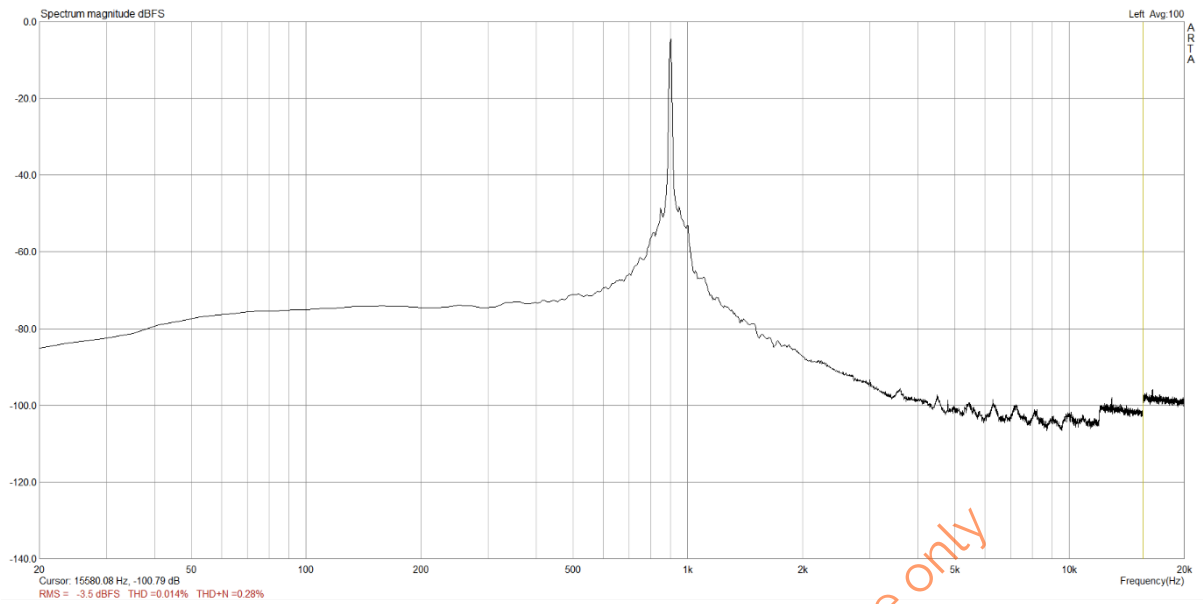
Frequency response with Pink Noise:



Spectrum with White Noise:



Total harmonic distortion (THD + N):



© AIIC 2019 - Reproduction for own use only

5. KUDO

KUDO (1)

KUDO is not compliant with some of the requirements of the ISO Standard 20108/20109, which can be seen by the results of the measurement of the frequency response with Pink Noise. THD+N measurement needs to be pointed out. Even with a level of -16,8 dBFs KUDO achieved 1,07 % as result. At higher levels this value appeared to be even higher. Delay (latency) was measured with 624 ms, not being compliant with the required 500ms.

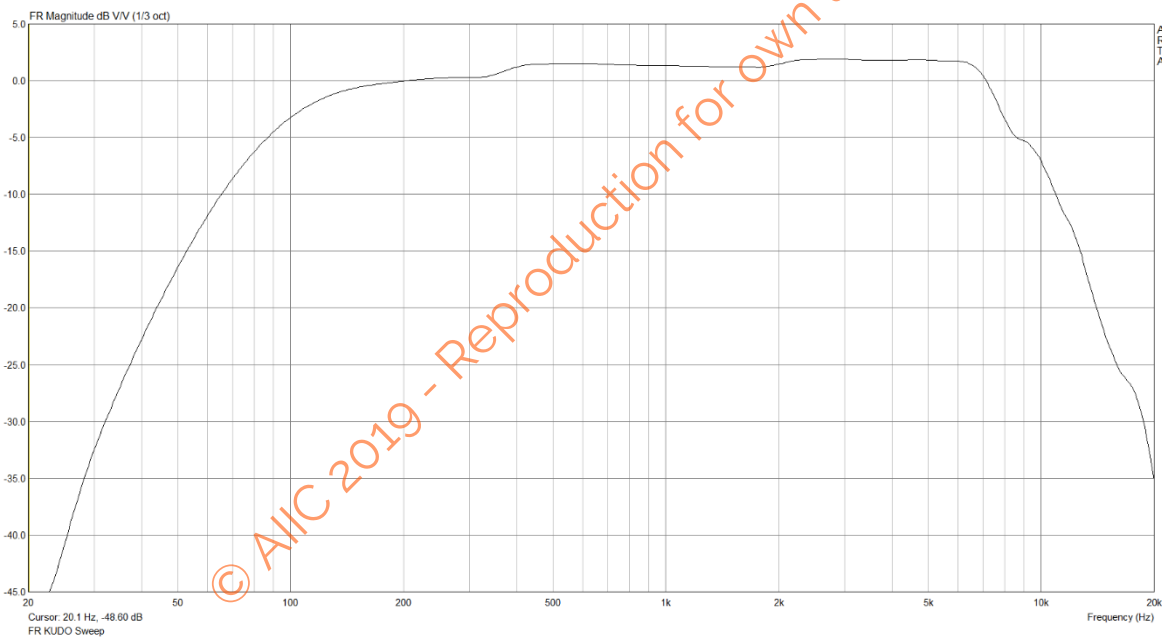
Deducting the 170ms originated by the use of Web Presenter, the obtained value for delay (latency) would be of 454ms, which would be compliant with the Standard.

STI values are very good for this platform, although all measurements with XL2 were marked as not precise, which can be attributed to the specificities of the codec (see above).

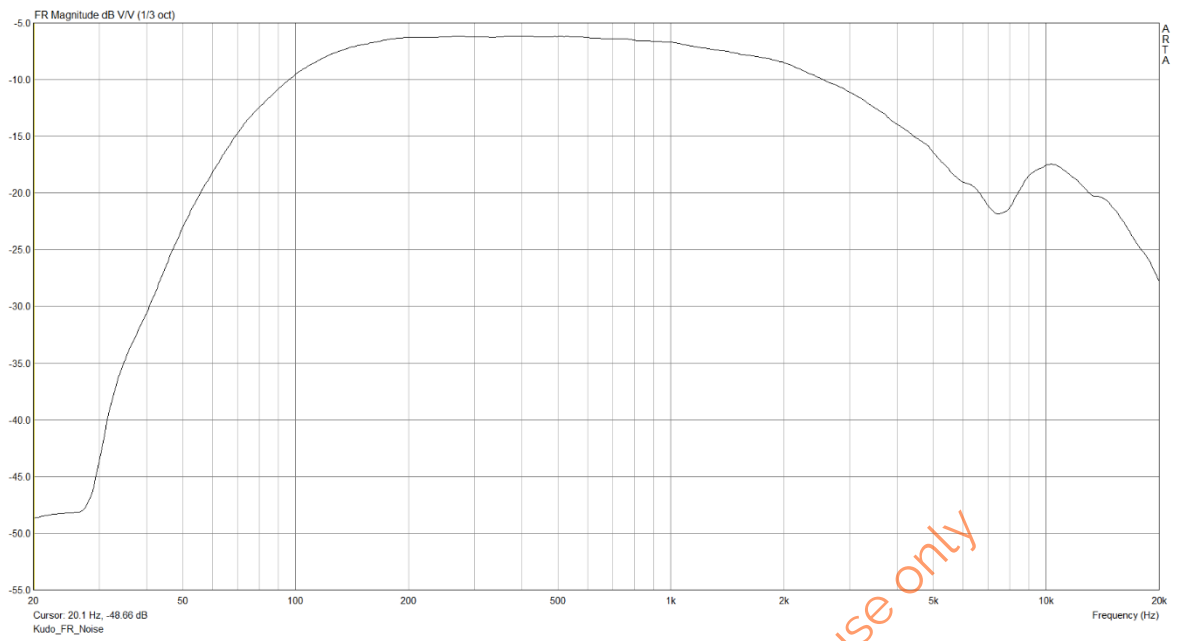
Also for KUDO, no hearing protection feature could be detected. The impulses got through without any limitation.

No visible artefacts could be detected during testing.

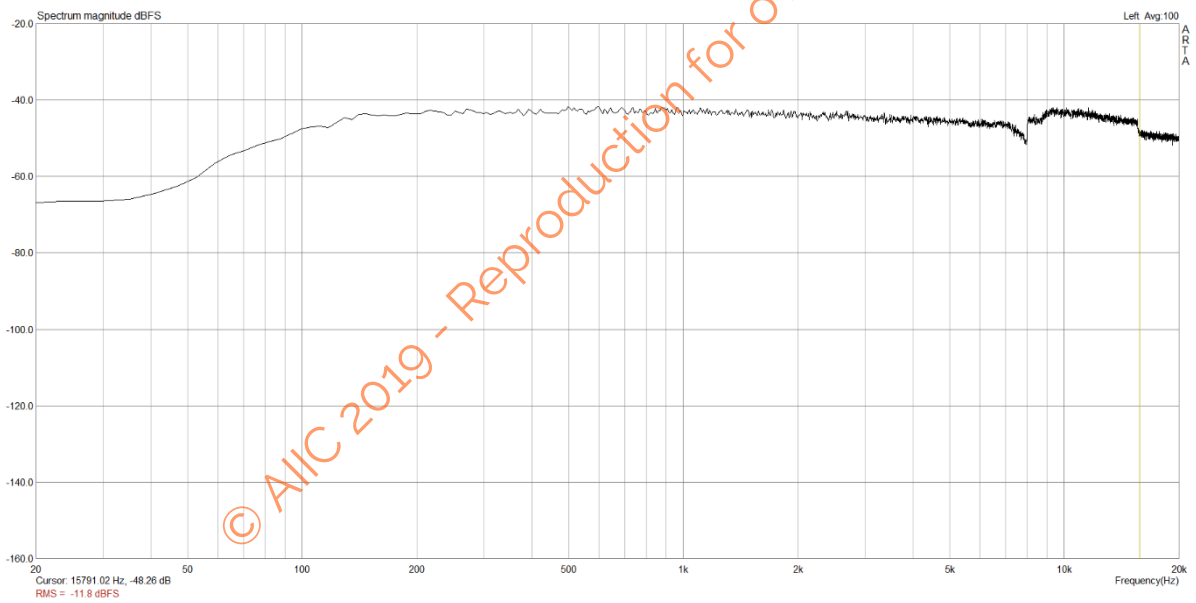
Frequency response with Sweep:



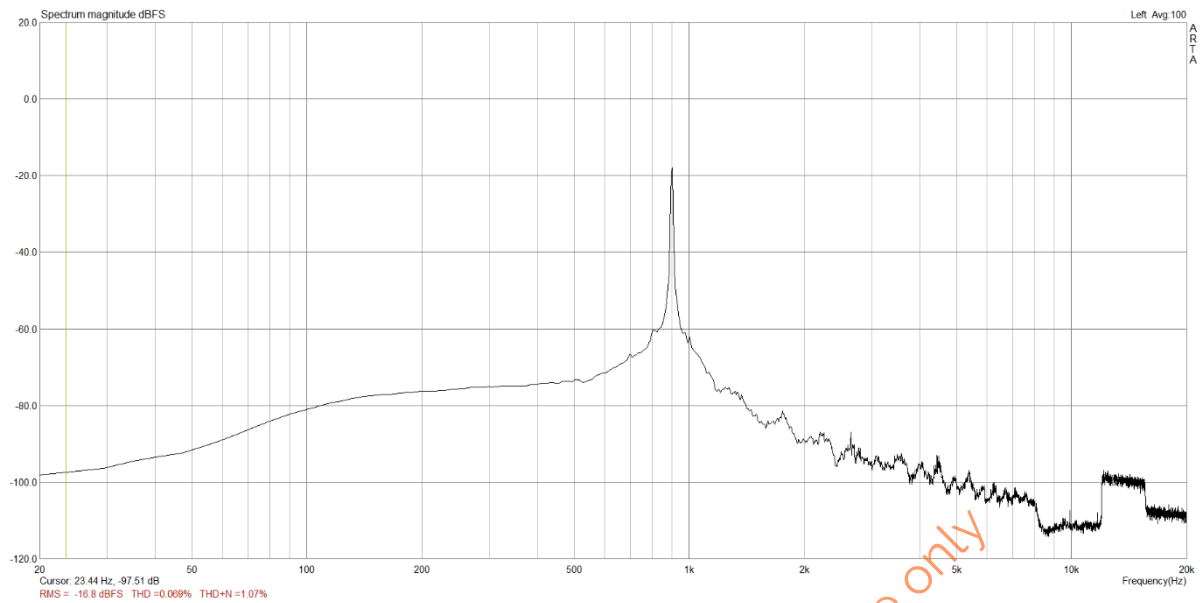
Frequency response with Pink Noise:



Spectrum with White Noise:



Total harmonic distortion (THD+N):



STI values calculated by ARTA:

Speech Transmission Index

SPEECH TRANSMISSION INDEX - MTF Matrix

Band	125	250	500	1000	2000	4000	8000
0.63	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9989
0.80	0.9997	0.9999	0.9999	0.9999	0.9999	0.9999	0.9982
1.00	0.9996	0.9999	0.9999	0.9999	0.9999	0.9999	0.9974
1.25	0.9995	0.9999	0.9999	0.9999	0.9999	0.9999	0.9964
1.60	0.9991	0.9998	0.9999	0.9999	0.9999	0.9999	0.9940
2.00	0.9986	0.9997	0.9999	0.9999	0.9999	0.9999	0.9913
2.50	0.9977	0.9994	0.9999	0.9999	0.9999	0.9999	0.9867
3.15	0.9962	0.9990	0.9998	0.9999	0.9999	0.9998	0.9806
4.00	0.9943	0.9986	0.9997	0.9999	0.9999	0.9997	0.9751
5.00	0.9907	0.9976	0.9995	0.9999	0.9999	0.9997	0.9694
6.30	0.9855	0.9963	0.9991	0.9998	0.9999	0.9996	0.9664
8.00	0.9772	0.9941	0.9986	0.9997	0.9999	0.9996	0.9657
10.00	0.9648	0.9909	0.9979	0.9995	0.9998	0.9996	0.9662
12.50	0.9459	0.9857	0.9966	0.9992	0.9997	0.9995	0.9675
OctTI	0.9924	1.0000	1.0000	1.0000	1.0000	1.0000	0.9962

STI = 0.9992 (male), 0.9994 (female) Rating: EXCELLENT (EXCELLEN
 (%ALcons= 0.7620)

Copy OK

KUDO (2)

Immediately after the testing session carried out on January 8, KUDO approached Neumann & Müller / Klaus Ziegler explaining that KUDO technicians had chosen erroneous settings for the operation of the platform during the first session. Following KUDO, these settings included additional new features that KUDO had not yet enough tested and officially released when the testing was done. KUDO asked for a repetition of the testing with their platform. A second testing session was carried out by Neumann & Müller audio engineers on February 13, measuring the same audio parameters again. The following results were obtained:

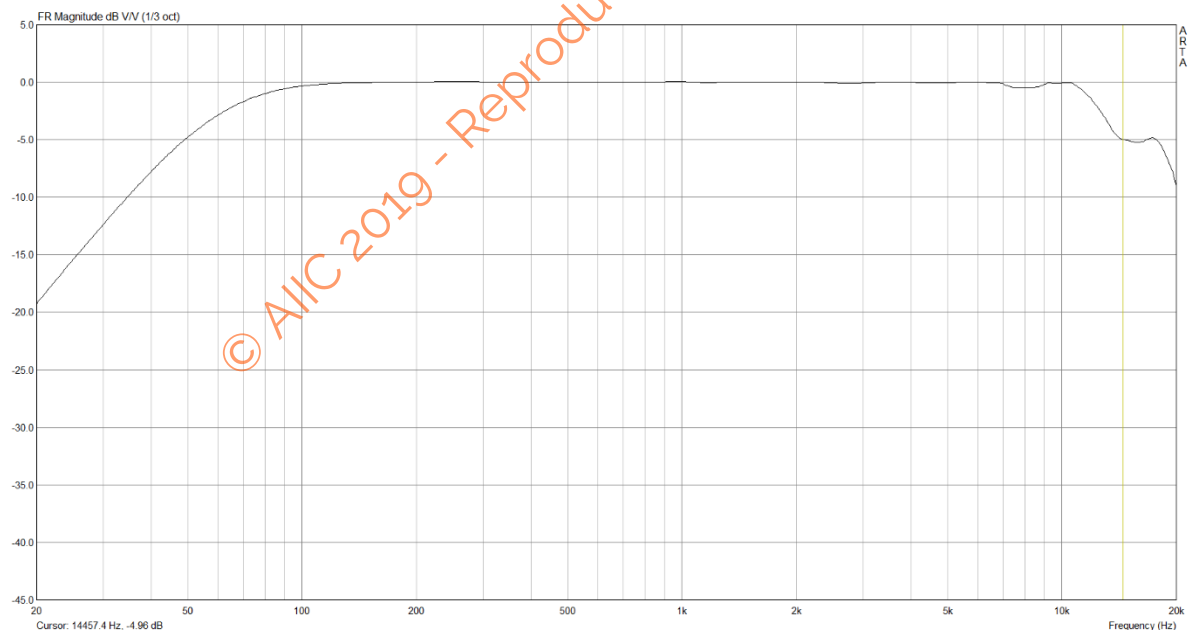
As we can observe, the frequency response shows a completely linear behavior up to aprox. 12 kHz, with a relatively steep decline beyond this level. This can be observed especially when measuring with Pink Noise and ten times averaging, as well as based on the spectrum. This frequency response is not compliant with ISO 20108⁴.

The STI values are higher than 0,90 measured both with XL2 Analyzer, but also calculated on the basis of the impulse response through ARTA. The result is very good, being compliant with the ISO Standard.

The THD+N value is at 0,41%, which is ok. For the latency/delay we measured 170ms without Web presenter. This value is compliant with the requirement of ISO 20108. Even if we added the delay caused by the use of a Web presenter, the total delay would still be kept within the limits as set out in ISO 20108.

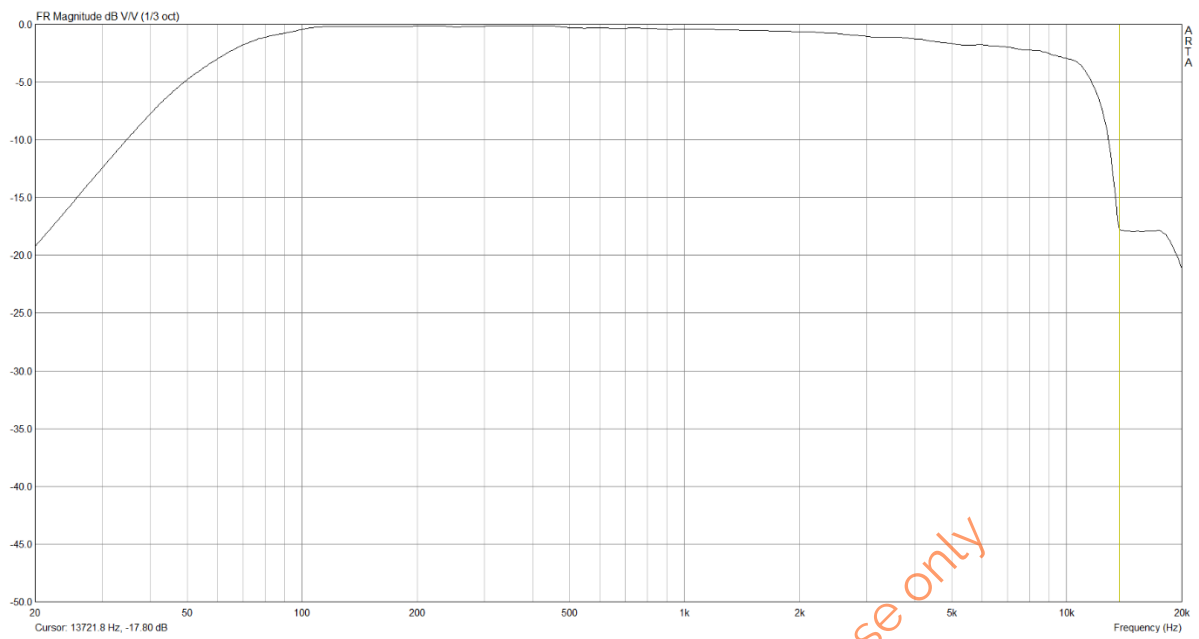
We could not detect any features regarding hearing protection.

Frequency response with Sweep:



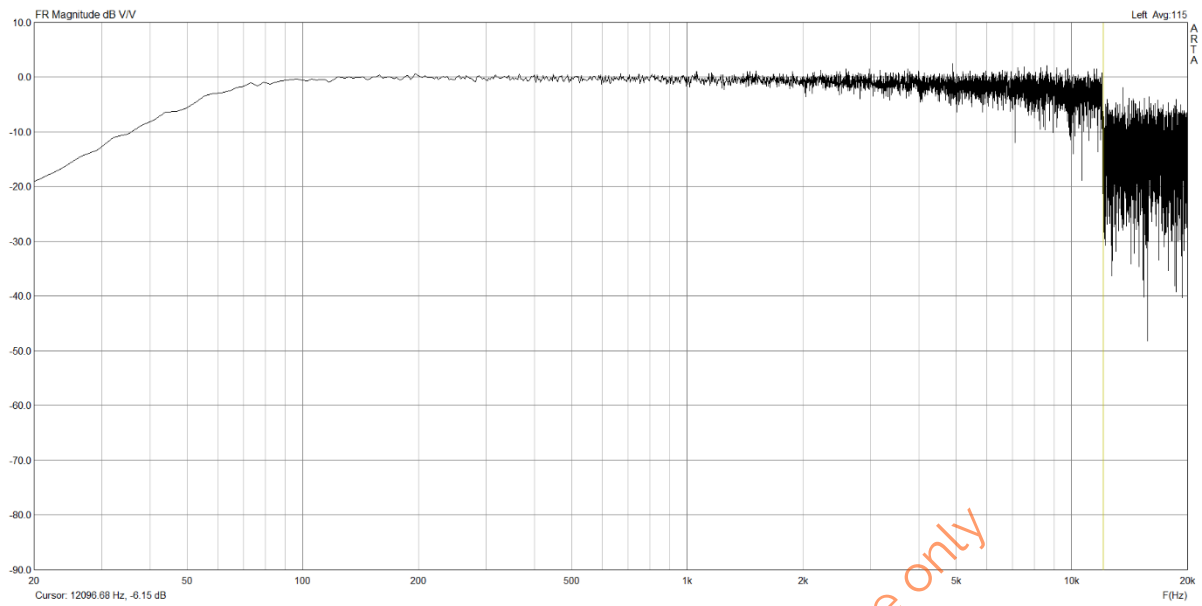
⁴ We assume that a sampling frequency of 24-30 kHz is applied to minimize the bandwidth for the audio signal. Raising sampling rate to the usual 44.100 Hz, or, even better, to 48.000 Hz would probably allow to adjust the anti-aliasing High Cut to a higher frequency as well. This would lead to a linear behavior of the frequency response up to aprox. 20 kHz, thus being compliant with the requirements in ISO 20108.

Frequency response with Pink Noise:

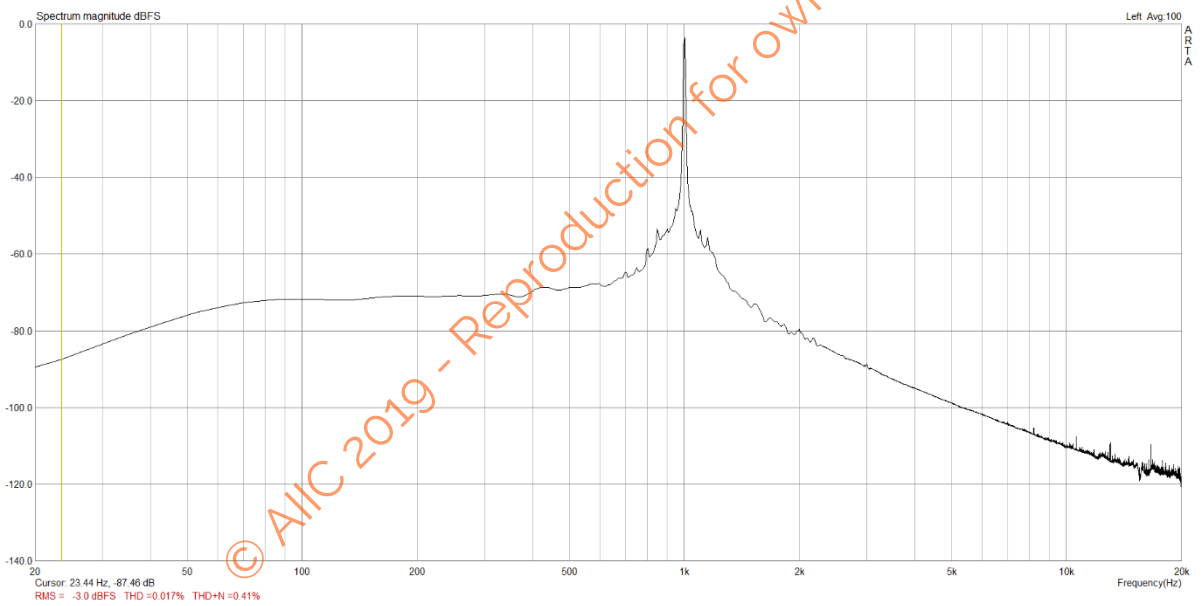


© AIIC 2019 - Reproduction for own use only

Spectrum with White Noise:



Total harmonic distortion (THD + N):



STI values calculated by ARTA:

Speech Transmission Index ×

SPEECH TRANSMISSION INDEX - MTF Matrix							
Band	125	250	500	1000	2000	4000	8000
0.63	0.9996	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
0.80	0.9993	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
1.00	0.9989	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
1.25	0.9985	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
1.60	0.9976	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999
2.00	0.9963	0.9997	0.9999	0.9999	0.9999	0.9999	0.9999
2.50	0.9940	0.9995	0.9999	0.9999	0.9999	0.9998	0.9999
3.15	0.9901	0.9991	0.9998	0.9999	0.9999	0.9997	0.9998
4.00	0.9851	0.9987	0.9997	0.9999	0.9998	0.9996	0.9997
5.00	0.9759	0.9979	0.9995	0.9998	0.9998	0.9994	0.9996
6.30	0.9626	0.9967	0.9992	0.9998	0.9997	0.9992	0.9994
8.00	0.9420	0.9947	0.9987	0.9996	0.9996	0.9990	0.9991
10.00	0.9120	0.9917	0.9980	0.9994	0.9995	0.9988	0.9988
12.50	0.8681	0.9871	0.9968	0.9991	0.9995	0.9987	0.9984
OctII	0.9632	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
STI = 0.9985 (male), 1.0000 (female) Rating: EXCELLENT (EXCELLEN							
(%ALcons= 0.7756)							

© AIIC 2019 - Reproduction for own use only

6. Interprefy

The frequency response measured with Pink Noise does not respond to the requirements set out in ISO 20108/20109. Also, when measuring with Sweep one can observe a limitation of the frequency range at aprox. 8 kHz.

Very good STI values, both measured and calculated, could be observed for Interprefy (0,94 measured with XL2; 0,99 calculated with ARTA). Both values are compliant with the requirement set out in ISO Standards.

On the other hand, the values of XL2 are marked as not precise (see above).

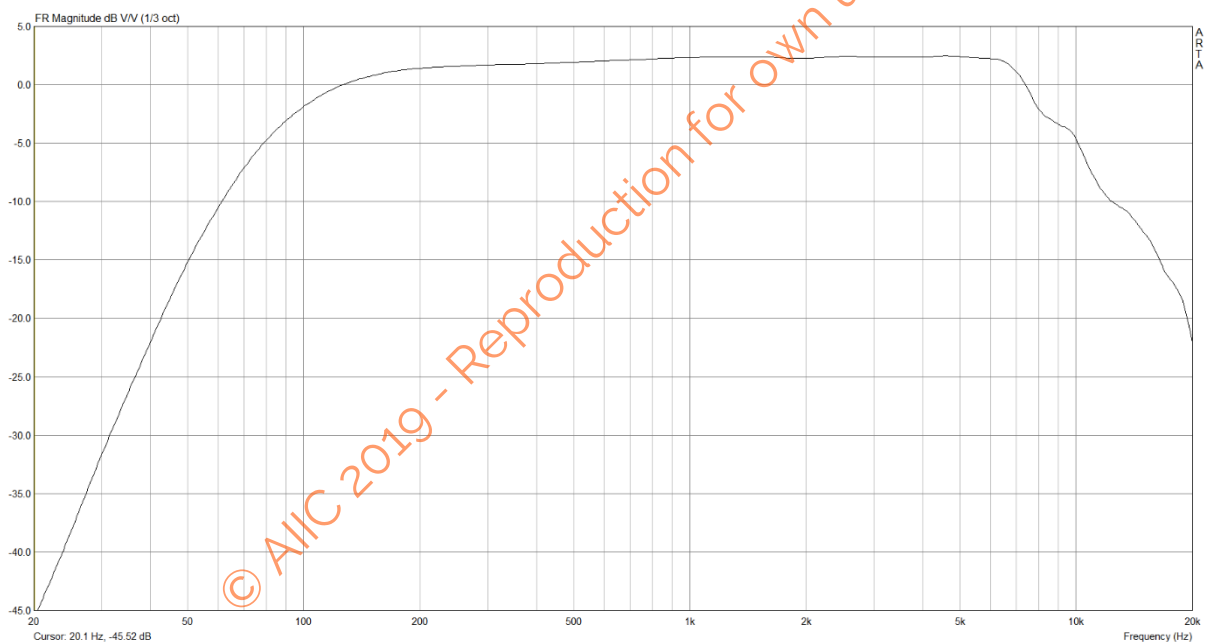
The values of the THD+N measurement are relatively high and achieve 0,44 %, thus being compliant with applicable ISO Standards.

Delay measurement gave a result of 436ms. This value is situated within the admissible ISO range.

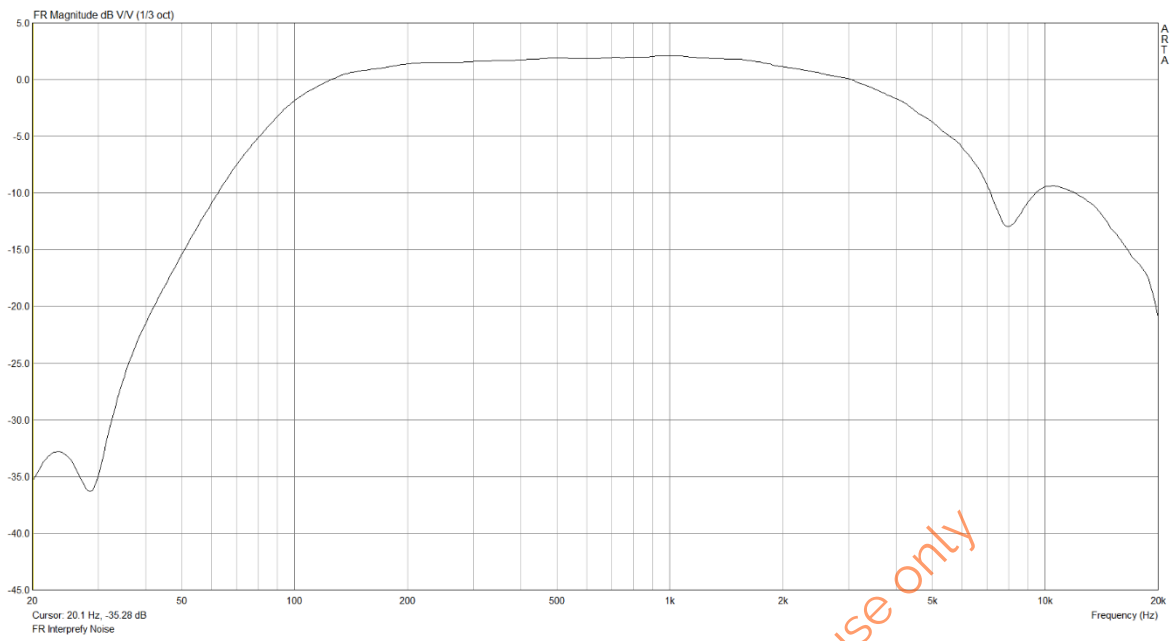
No hearing protection feature was detected. The impulses were transmitted through the platform without compression/limitation.

No visible artefacts were detected during the test (ISO compliant).

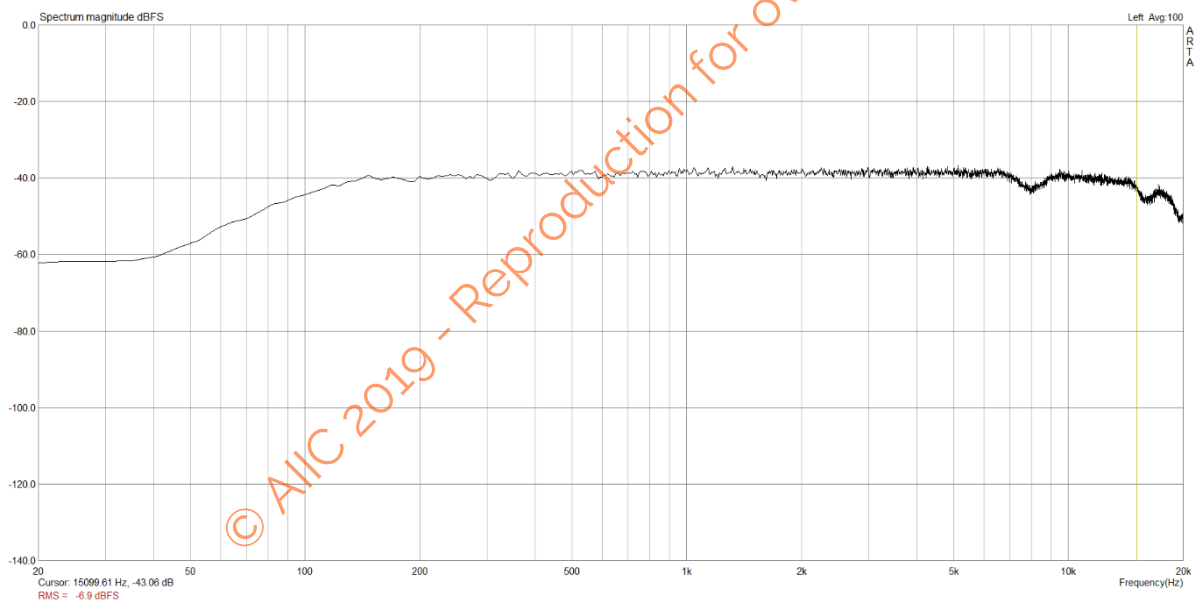
Frequency response with Sweep:



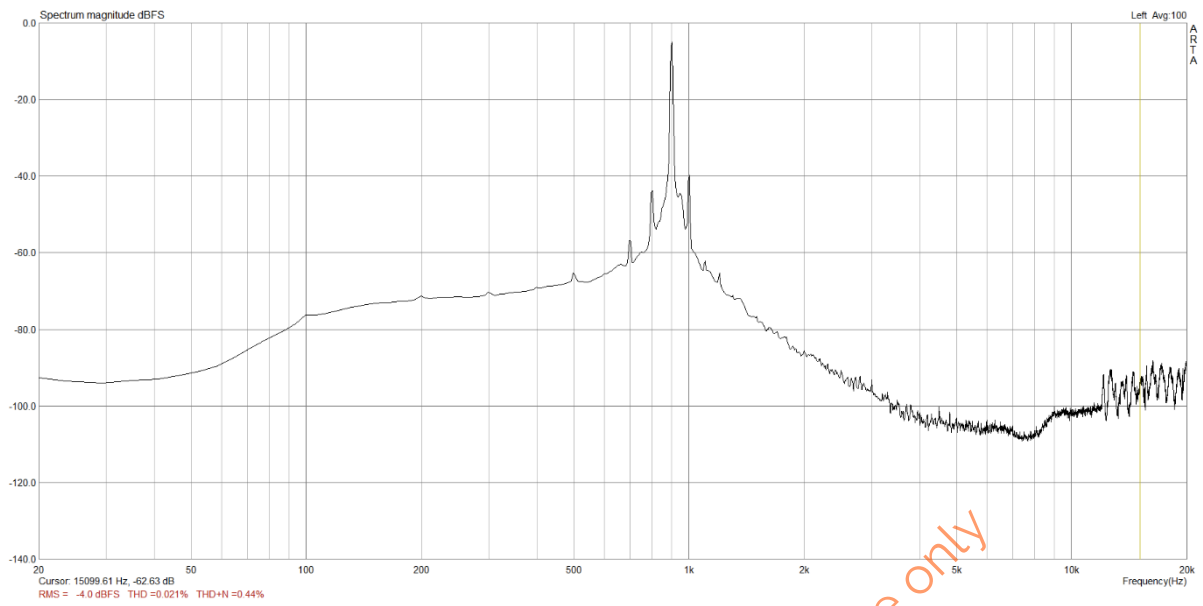
Frequency response with Pink Noise:



Spectrum with White Noise:



Total harmonic distortion (THD + N):



Measuring was carried out by audio engineers from Neumann & Müller (Esslingen, Germany).

Results revised and approved by AIIC Technical and Health Committee (Klaus Ziegler, Coordinator).

Esslingen/Hamburg, February 17 of 2019

© AIIC 2019 - Reproduction for own use only