QUality Evaluation of micropones used for lecture capture in Universities

R Sanchez-Pizani London South Bank University/Kings College London, UK

Stephen Dance London South Bank University, UK

L Gomez-Agustina London South Bank University, UK

P Mapp London South Bank University, UK

# INTRODUCTION

This paper describes a recent series of experiments carried out with microphones commonly used for the live recording of lecturers in university teaching rooms.

Two experiments were designed to compare audio recordings from the microphones.  9 King’s College London Audio Visual engineers, that currently assess the perceived quality of lecture capture in the university, rated the recordings on a five-point mean opinion scale (MOS). Two different approaches to recording were followed, using simultaneous recording and substitution, with the latter showing better and more consistent results.

Results are then analysed by correlation to see if there is a significant relationship between the accuracy of a microphone, compared to results from a class 1 sound level meter, and the quality rating by the engineers.

# Previous REsearch

## On-Site Microphone Comparison

To this date it has not been possible to find any systematic methodology that compares microphones in lecture capture situations. The scarce literature seems to indicate procedures to compare specific parameters; however it is required that the experimentation is carried out in a low reverberation and noise environment 1. In recent years some researches have done some investigation into the most effective way to compare microphones. Establishing four main criteria for a fair comparison, i.e.

1. identical source;

2. identical sound field (at microphone);

3. no microphone limitation; and

4. practical and flexible.

Their conclusion was that there is no method that will allow all four criteria to be fulfilled and as such the best approach is to use an array configuration 2.

The standard procedure recommended by BS EN 60268-4:2014 is to use the substitution methodology as this produces more accurate results3; however all these methods assume a controlled acoustic environment.

As the interest of this research is to find a practical alternative to evaluate the quality of microphones in situ used for lecture capture of lessons, none of the previously described methods can be adopted fully.

A previous attempt to do on site comparison of recordings with quantitative metrics, in this case the speech transmission index for public address systems (STIPA), found no significant relationship between STIPA and the subjective quality evaluation 4.

Regarding the quality evaluation, the present research works with mean opinion scale (MOS), commonly used in the telecommunications industry, as recommended by International Telecommunication Union (ITU). This consist of a five-point scale rating of the recordings. Where the total quality is calculated as the arithmetic mean of the scores 5.

The main hypothesis of the experiment is based around the fact that, in order to estimate the quality of a recording correctly some form of artificial intelligence would be needed. A linear relationship between a single acoustic parameter and quality is not possible 6. This would be particularly true when said estimate is the product of measurement produced by non-laboratory calibrated equipment and microphones. As such, the main objective of the current research seeks to demonstrate, that what is understood as a good microphone doesn’t necessarily mean that it will produce higher quality results; therefore, justifying the use of machine learning algorithms to analyse the quality of an audio recording.

# Methodology

It was decided to separate the test into two experiments. In part 1 the set up was dictated by current practices and the manufacturers recommendations. It is described by the Figure 1. Part 2 uses laboratory conditions as dictated by the substitution method as per current standard.

The analysis is done by comparison between the scenarios using results from MOS tests correlation and linear regression techniques and see if there is any correlation between the results.

## Experiment 1

In a test room, at KCL several microphones were set up in their typical positions and connected to a sound card. Using a common setup as described in Figure 1 below.

Key:

Red test and symbol: beamforming/tracking microphones

Red Symbol and blue text: microphone arrays

Blue Text and Symbol: omnidirectional mic

Green Omni mic on stand

Red speaker: Louder recording and speaker for electrical distortion.



Figure 1 Floor plan of location of microphones in the test room

All microphones were installed and set up by professional AV integrators, following manufacturer’s instructions to ensure the best possible performance. A 1kHz tone was played at 1m of each speaker and all levels adjusted to similar values as indicated by the rms meter of the sound card.

Following indications by ITU recommendations, a single audio track of recordings from female and male speakers was playback using a Fostex monitor 601 from a PC running Adobe Audition.

The speaker was moved between 3 different positions for a total of 7 different directions and 49 combinations. Microphones remained fixed.

The audio track was recorded simultaneously, in an adapted version of an array configuration as shown in Figure 1, to ensure identical noise source and room conditions.

The audio track was composed using Adobe audition and the audio package from British English downloaded from the ITU website 7.

## Experiment 2:

The aim of this experiment was to investigate if there is a significant relationship in the subjective quality of the resulting recording when exposed to different levels of reverberation.

To achieve this, 6 different microphone types were compared to the track used in experiment 1. The audio track was loaded into an NTI minarator and the level was kept constant.

To ensure correct and fair comparison all microphones were calibrated by comparison following the guidance in (add British Standard); however, given the size, shape and technical limitations of some of the microphones, some assumptions were necessary.

Two assumptions were made:

* The most sensitive area of all microphone was on axis and at their centre.
* The acoustic environment is approximately constant for all cases.

The was minarator set at -20 dBF with a pre-recorded file. The beginning of the audio track had pink noise used to compare the SPL by the speaker.

The first measurement was done at 1200 mm from the acoustical centre of the source to the front of the microphone to replicate a typical distance in teaching situations.

The same process was repeated with all microphones at distances of:

|  |  |
| --- | --- |
| Anechoic Environment | Direction |
| 1200 | 0 degrees from microphone centre |
| 2200 | 0 degrees from microphone centre |
| 2200 | 90 degrees right from microphone centre |
| 2200 | 90 degrees Left from microphone centre |
| Reverberant Environment |  |
| 3200 | 0 degrees from microphone centre |
| Semi-Reverberant Environment |  |
| 3200 | 0 degrees from microphone centre |

Table 1 Summary of Laboratory Conditions used for Microphone Testing

### Calibration

A challenge to calibrate the microphones in question is that, coupling a calibrator to them is impossible, so this had to be done by substitution.

To ensure the least amount of unwanted interference, the process was tested in an anechoic room at London South Bank University (LSBU). The calibration process was done using ARTA as shown in Figure 2 below.



Figure 2 Connectivity of Microphones to the Audio System

To test that the methodology was sound, the sensitivity of two Earthworks and an NTI microphone, all factory calibrated, was determined using a Norsonic calibrator at 114 dB at 1kHz.

Earthworks 1 = 37.2 mv/Pa

Earthworks 2 = 34.2 mv/Pa

NTI mic = 42.9 mv/Pa

These values were compared with the stated calibration by the manufacturer. As the result was less than 1mv/Pa from the lab value, the process was considered within acceptable error to proceed.

Sound Card and output levels were also calibrated following guidance from the software developer.

Using Arta, a 400 Hz signal was played back and adjusted the input channel until obtaining a reading of approximately -3 in the RMS meter in front of the Focusrite Scarlet 18i/20. The output signal was then measured with a multi-meter by connecting pins 2 and 3 to positive and negative terminals of the Multi-meter respectively.

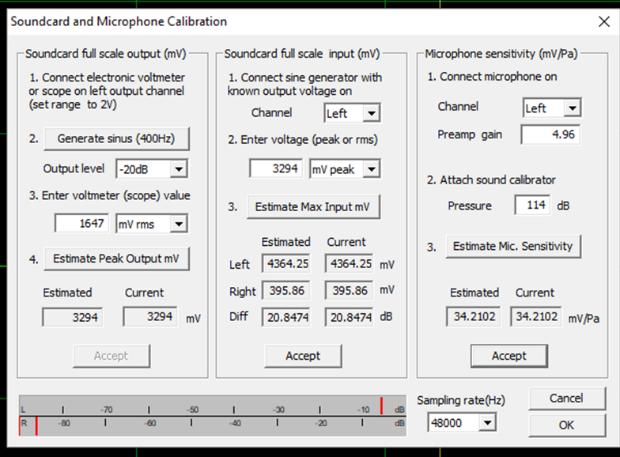


Figure 3 Calibration Example Results, using ARTA. Example shows Earthworks

Once known values were obtained from independent processes of calibration, the array methodology was tested.

Using a minirator MR-PRO a 1K was sent to the omnidirectional source at 240mV (omniamp of the dodecahedron spiker set at 0 attenuation).

NTI microphone was connected to a Sound Level Meter NTI XL2, at 1 meter from the omni source and measured for 15 seconds Leq. The value obtained was 78.1 dB.

Using the previously calibrated gain and sensitivity values for the Earthworks mic, this was placed very close next to the NTI and using ARTA’s SPL tool a Leq of 15 seconds was taken.

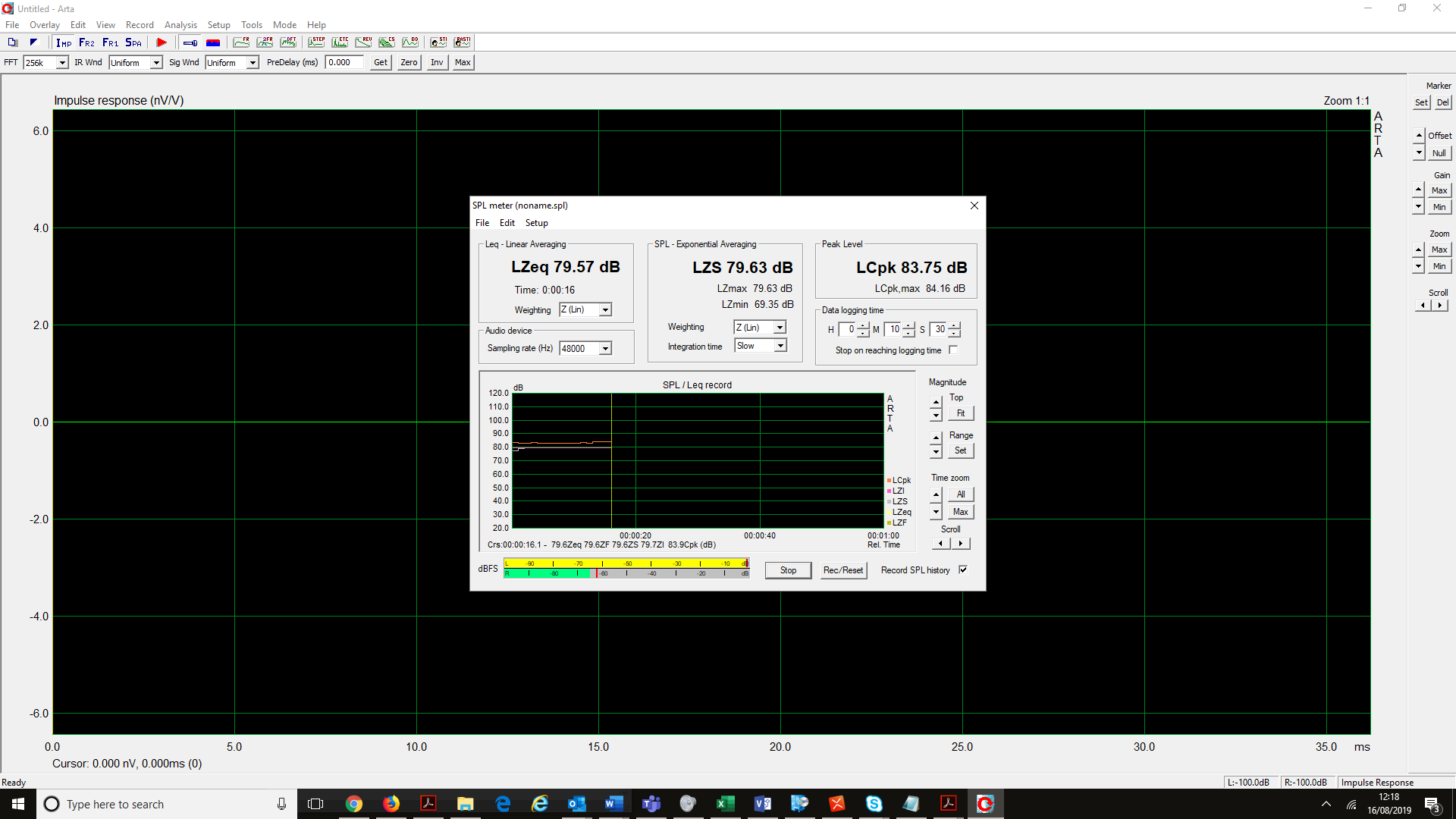




Figure 4 First measurement using Earthworks as input and NTI as reference

This discrepancy was considered too large; therefore, requiring further investigation. It was determined that the error was the product of the acoustic environment and reflection from the Nureva microphone behind them.

To confirm this was the case, the process was repeated using 3 separate NTI SLM and measuring the levels of the tone in three different positions simultaneously.

Removing the speaker improved the results, as difference became less noticeable; however, there was a larger improvement by replacing the reference microphone by the one under test, effectively agreeing with the current standard recommendation. As a result, it was decided to keep this technique and use this to calibrate all the other microphones and do the recordings as it seems as if the process holds.

Following the recommended process all microphones were then calibrated using a dodecahedron and a 400 Hz tone at 1m. The standard deviation observed between all test was of 0.15 dB as seen in Table 1; therefore, making the results comparable.

|  |  |  |  |
| --- | --- | --- | --- |
| 1m Source, 400 Hz Tone | NTI Value | Mic result | Sensitivity |
| Leq,15s | Leq,15s | mV/Pa |
| Earthworks | 85.1 | 85.55 | 38.4 |
| ES97 | 86.1 | 86.09 | 7.66 |
| RM30 | 86.0 | 86.01 | 26.09 |
| MPR210 | 82.0 | 82.09 | 310.67 |
| Nureva | 81.5 | 81.32 | 125.62 |
| MXA910 | 86.0 | 86.04 | 197.69 |

Table 2 Results from Calibration, using Omni source

# Results

## Experiment 1

9 professional sound engineers, all working on Audio Visual Services at King’s College London (KCL), volunteered to take part in a survey to evaluate the quality of the recordings (under ethical approval from LSBU: ETH1819-0086 and KCL: MRA-18/19-11895).

A listening station was set up. This consisted of a Dell Latitude laptop running Windows 10, a multitrack session from Adobe audition CC 2018 and a pair of Sennheiser headphones HD 4.20s. Each engineer was given verbal instructions on how to mark the quality of the recordings on a fixed scale and what to evaluate. Everyone adjusted the monitoring audio to a comfortable level at the beginning of the test, keeping this fix afterward.

They compared each microphone against the others for every position recorded in the room.

They were instructed to ignored intelligibility.

6 microphones were tested in 7 different positions.

Figure 5 Quality per Microphone with Different Source Positions

There seems to be no surprise that microphones looked as if they perform better when the source was closer, e.g. the MPR210 located on the lectern.

On average, microphone arrays in the middle of the room appeared able to provide a more consistent experience, regardless of the speaker position.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| M50 Stand  1.6m | MPR210  on Lectern | Nureva | RM30 Ceiling | M50 ceiling | Sennheiser | Shure |  |
| 2.3 | 2.0 | 2.2 | 2.5 | 2.3 | 2.8 | 3.1 | **AVG** |
| 146 | 123 | 136 | 154 | 143 | 175 | 195 | **Total** |

Table 3 Total Quality Average per Microphone in Classroom

Figure 6 Quality Average Classroom

## Experiment 2

Several microphones were set up and tested in three separate acoustical conditions. An anechoic room, reverberant and semi-reverberant conditions. The minarator was loaded with an audio track with the structure:

* noise,
* speech from for different talkers 4 male and 4 female
* single tones for 250,500, 1000, 2000, 4000 and 8000 Hz.

This was done to ensure a known value was used for the recording to allow for comparison.

The recordings are shown in the figure below for all three spaces.

Figure 7 Difference Between Microphones and Reference Value

While the measurement above took place, an audio file was also recorded to allow a comparison between the opinion scores and the specific results from the measurement. The table below shows the results summarised as the average of the scores for all microphones and situations. In this case, the AV engineers were told to compare the quality of each recorded audio with the original file only, i.e. ignoring the other microphones.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Anechoic | Rev | Semi-Rev | Total |
| M50 Stand 1.6m | 4.3 | 1.2 | 1.8 | 2.4 |
| MPR210 on Lectern | 2.7 | 1.3 | 2.3 | 2.1 |
| Nureva | 2.3 | 1.4 | 2.4 | 2.1 |
| RM30 Ceiling | 3.0 | 1.4 | 2.4 | 2.3 |
| M50 ceiling | 4.3 | 1.2 | 2.1 | 2.6 |
| Shure | 3.4 | 1.8 | 3.1 | 2.8 |
| ES97 | 4.3 | 1.1 | 2.3 | 2.6 |

Table 4 Total Quality Score in Lab conditions

Figure 8 Anechoic Microphone Quality Rating

Figure 9 Semi-Reverberant Microphone Quality Rating

Figure 10 Reverberant Microphone Quality Rating

# Discussion

An analysis of the resulting quality and the accuracy of the measurement microphone, with the hypothesis than the more accurate the microphone will produce better subjective quality results proved inconsistent. As Table 6 shows, there is no significant relationship between, how accurate the microphone is, and the subjective quality rating given by expert listeners.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **R** | **R Square** | **Adjusted R Square** | **Std. Error of the Estimate** | **Change Statistics** | | | | |
| **R Square Change** | **F Change** | **df1** | **df2** | **Sig. F Change** |
| 1 | .074a | .005 | -.057 | 1.00421 | .005 | .087 | 1 | 16 | .772 |

Table 6 Linear Regression Comparison between Accuracy of the microphone and quality rating

Similarly, if results are analysed by comparing the quality rating between the different scenarios, it is possible to identify some correlation between, the quality rating of the situations where the microphones are facing away, and high reverberation lab conditions, as seen in Table 7. The same cannot be said for the lab tests in anechoic conditions.

This is consistent with the calibration results, that even though all microphones had very accurate outcomes, when calibrated on axis while using a 400 Hz sinusoidal tone in an anechoic environment, adding complex noise, changing the direction of the source or increasing reverberation; all appear to influence the accuracy of the results.

For array and beamforming microphones this phenomenon seems less evident for diffuse spaces, as can be seen in Figure 7; however, it is not clear if this is a result of microphones being less susceptible to reverberation or the fact that all measurements in the reverberant room were on axis; therefore, more research is needed in this area.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | High Gain Middle | Anechoic | Rev | Semi\_rev | Total | Low gain Middle | Entrance Facing Away | Lectern Facing Left | Lectern Facing Right | Corner Facing Corner | Door Facing Front |
| High Gain Middle | **1** | 0.281 | -.846\* | -.848\* | -0.396 | -0.331 | -0.567 | -0.515 | -0.454 | -0.533 | -0.520 |
| Anechoic | 0.281 | **1** | -0.320 | -0.423 | 0.651 | 0.644 | 0.187 | -0.197 | -0.302 | -0.084 | 0.097 |
| Rev | -.846\* | -0.320 | **1** | .965\*\* | 0.478 | 0.492 | .782\* | 0.325 | 0.280 | .844\* | .772\* |
| Semi\_rev | -.848\* | -0.423 | .965\*\* | **1** | 0.407 | 0.375 | 0.691 | 0.370 | 0.359 | .738\* | 0.653 |
| Total | -0.396 | 0.651 | 0.478 | 0.407 | **1** | .974\*\* | .801\* | 0.016 | -0.095 | 0.569 | 0.683 |
| Low ga Middle | -0.331 | 0.644 | 0.492 | 0.375 | .974\*\* | **1** | .853\* | -0.098 | -0.218 | 0.670 | .775\* |
| Entrance Facing Away | -0.567 | 0.187 | .782\* | 0.691 | .801\* | .853\* | **1** | -0.181 | -0.267 | .936\*\* | .978\*\* |
| Lectern Facing Left | -0.515 | -0.197 | 0.325 | 0.370 | 0.016 | -0.098 | -0.181 | **1** | .986\*\* | -0.186 | -0.265 |
| Lectern Facing Right | -0.454 | -0.302 | 0.280 | 0.359 | -0.095 | -0.218 | -0.267 | .986\*\* | **1** | -0.248 | -0.349 |
| Corner Facing Corner | -0.533 | -0.084 | .844\* | .738\* | 0.569 | 0.670 | .936\*\* | -0.186 | -0.248 | **1** | .974\*\* |
| Door Facing Front | -0.520 | 0.097 | .772\* | 0.653 | 0.683 | .775\* | .978\*\* | -0.265 | -0.349 | .974\*\* | **1** |

Table 7 Correlation Matrix of all Quality Ratings

# Conclusion

Two experiments were conducted to test microphones in the laboratory and in real-life conditions. The accuracy of the microphones, expressed as the distance between the measured Leq by the microphone under test and a calibrated laboratory system, was compared to the quality rating and the resulting regression showed that there was no significant relationship.

There was no microphone that was always the best, with omnidirectional mics (Earthworks M50) working the best in anechoic situations, while array microphones (Sure MXA 910) provided better quality in semi reverberant conditions. All microphones scored lower with high levels of reverberation than in anechoic conditions.

# Recommendations and Future work

Results show that there is no significant relationship between the accuracy of a microphone and the perceived quality; therefore, it is justified to investigate the implementation of some form of supervised machine learning algorithm, to find an objective way to rate subjective quality of recordings and microphones used in teaching rooms.

# References

1. Schmidle, G., Beach, M. & MacMillan, B. Challenges and best practices for microphone end-of-line testing. *New York* 10 (2018).

2. Pearce, A., Brookes, T. & Dewhirst, M. Validation of experimental methods to record stimuli for microphone comparisons. *New York* 10 (2015).

3. BSI. Sound system equipment - Part 4: Microphones (IEC 60268-4:2014) Equipements. (2014).

4. Bradley, J. S. & Gover, B. N. Subjective and Objective Rating of Intelligibility of Speech Recordings. in *Convention Paper 7178* 12 (AES, 2007).

5. T-REC-P.800 TELEPHONE TRANSMISSION QUALITY Methods for objective and subjective assessment of qualityi. (08/96).

6. Zieliński, S., Rumsey, F. & Bech, S. Towards Unification of Methods for Speech, Audio, Picture and Multimedia Quality Assessment. 16 (2015).

7. ITU, T. ITU-T Test Signals for Telecommunication Systems.Test Vectors Associated to Rec. ITU-T P.50 Appendix I. *ITU* http://www.itu.int/net/itu-t/sigdb/genaudio/AudioForm-g.aspx?val=1000050.