QUALITY EVALUATION OF MICROPHONES USED FOR LECTURE CAPTURE IN UNIVERSITIES

R Sanchez-Pizani London South Bank University/King's College London, UK
S Dance London South Bank University, UK
L Gomez-Agustina London South Bank University, UK
P Mapp London South Bank University, UK

1 INTRODUCTION

Recently, lecture capture has become progressively more important to Higher Education institutions. They allow automated recordings of lectures to be taken and stored for consumption by remote students or students who wish to revise the class. The technology involved has become increasingly complex and hence the need to independently assess the quality of recordings made by the microphones.
This paper describes a recent series of experiments carried out with microphones commonly used for the live recording of lectures in university teaching rooms. Two laboratory-based experiments were designed to compare the audio recordings from the microphones; using nine Audio Visual engineers, that currently assess the perceived quality of lecture capture at universities. The ratings of the recordings were captured on a five-point mean opinion scale (MOS). Two different approaches to recording were followed, i.e. using simultaneous recording and substitution. Results were analysed by correlation to investigate if there is a significant relationship between the measurements of each of the six microphones, when compared to results from a class 1 sound level meter, and the quality rating by the engineers.

2 PREVIOUS RESEARCH

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 specify 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 research has been undertaken 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.

The 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 a substitution methodology as this produces more accurate results[3]; 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 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 using mean opinion scale (MOS), as commonly used in the telecommunications industry, and is 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.

3 METHODOLOGY

It was decided to separate the test into two experiments. In part 1 the set up follows existing practices and the manufacturers recommendations. Part 2 uses laboratory conditions as dictated in the substitution method as per the current British Standards.
The analysis is done by comparison between the scenarios using results from MOS tests, correlation and linear regression techniques, to investigate if there is any connexion between the results.

3.1 Experiment 1

In a test room, at King's College London (KCL) several microphones were set up in their typical positions and connected to a sound card. Using a common setup as seen in images below

Key for Figure 3:
Red name and symbol:
beamforming/tracking microphones
Red symbol and blue text:
microphone arrays
Blue text and symbol:
omnidirectional mic
Green:
mic on stand
Red speaker:
Louder recording level for electrical distortion.

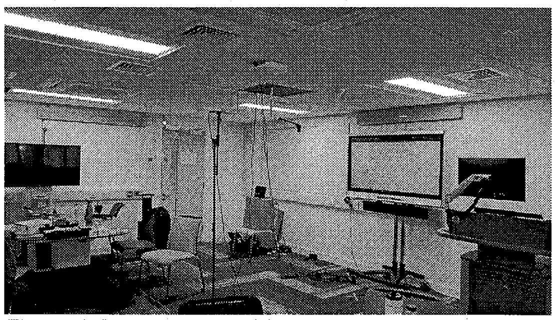

Figure 1 Classroom used for testing

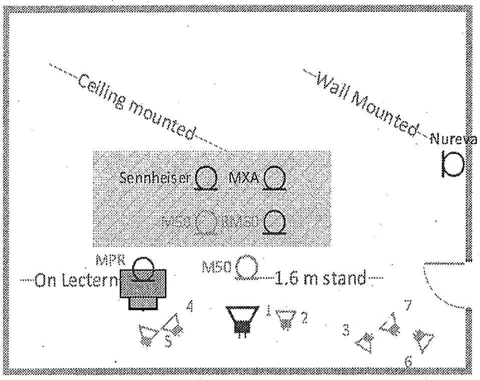

Figure 3 Test room floor plan (not to scale)

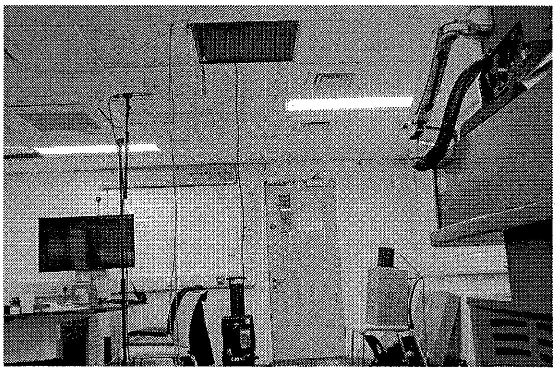

Figure 2 Classroom used for testing angle 2

Figure 1and Figure 2 show the room during the test, while Figure 3 shows a floor plan.
All microphones were installed and set up by professional AV integrators, following the manufacturer's instructions to ensure the best possible performance. A reference 1 kHz tone was played at 1m from each microphone and all levels adjusted to similar values as indicated by the rms meter of the sound card.
Following ITU recommendations, a single audio track of recordings from female and male speakers was played back using a Fostex monitor loudspeaker 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. To ensure identical noise source and room conditions all microphones remained fixed.
The audio track was composed using Adobe Audition and the audio package downloaded from the ITU website for British English [7].

3.2 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, six different microphone types were compared to the track used in Experiment 1.
The audio track was loaded into an NTi minirator and the level was kept constant.
To ensure correct and fair comparison all microphones were calibrated by comparison following the guidance in BS EN 60268-4:2014; 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 NTi, Minirator Pro was set at -20 dBF with a pre-recorded file. The beginning of the audio track had pink noise which was used to compare the SPL by the speaker.
The first measurement was done at 1200 mm from the acoustical centre of the loudspeaker 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 mm	0 degrees from microphone centre
2200 mm	0 degrees from microphone centre
2200 mm	90 degrees right from microphone centre
2200 mm	90 degrees Left from microphone centre
Reverberant Environment	
3200 mm	0 degrees from microphone centre
Semi-Reverberant Environment	
3200 mm	0 degrees from microphone centre

Table 1 Summary of laboratory conditions used for microphone testing

3.2.1 Calibration

A challenge was to calibrate the microphones in question, coupling a calibrator to them was not possible, so this had to be done by using the substitution method.
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[8] as shown in Figure 4.

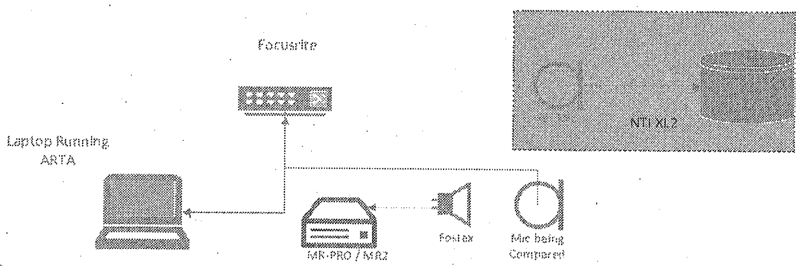

Figure 4 Connectivity of microphones to the audio system

To test that the methodology was sound, the sensitivity of two Earthworks M50 microphones and an NTi Class 1 microphone M2230, all factory calibrated, was determined using a Class 1 Norsonic 1 kHz calibrator at 114 dB.
Earthworks 1 = 37.2 mv/Pa *Earthworks 2 = 34.2 mv/Pa* *NTi microphone = 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. See Figure 5.

Sound Card and output levels were also calibrated following guidance from the software developer. Using Arta, a 400 Hz signal was played back, and the input channel adjusted until a reading of approximately -3 in the RMS meter in front of the Focusrite Scarlet 18i/20 sound card was obtained. 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.

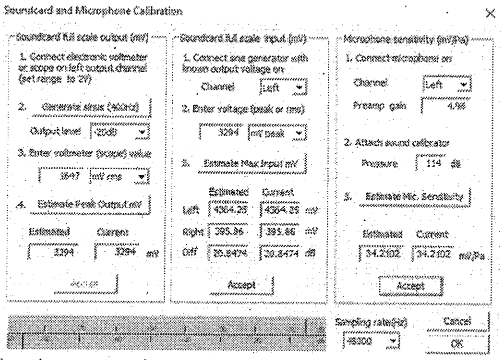

Figure 5 Calibration example results, using Arta. Example shows earthworks

Once known values for sensitivity were obtained from the calibration process, the array methodology could be tested.
Using a Minirator MR-PRO, a 1 kHz signal was sent to the omnidirectional source at 240mV (omniamp of the dodecahedron speaker set at 0 attenuation).
The NTi microphone, M2230, was connected to a Sound Level Meter NTi XL2, at 1 meter from the omni source and measured for 15 seconds L_{eq}. The value obtained was 78.1 dB.
Using the previously calibrated gain and sensitivity values for the Earthworks microphone, this was placed very close to the NTi and using Arta's SPL tool, a L_{eq} of 15 seconds was taken, resulting in a 79.6 dB reading, as observed in Figure 6.

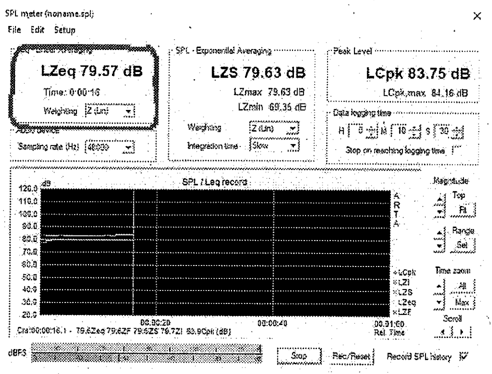

Figure 6 Measurement using Earthworks

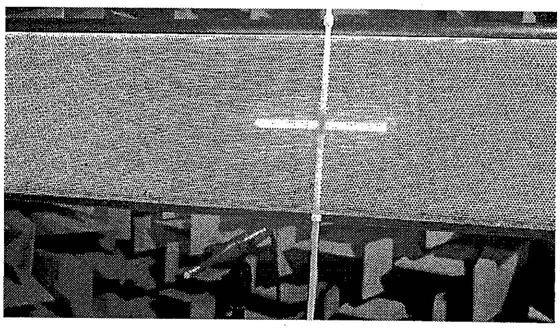

Figure 7 Nureva behind Earthworks during test

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 large Nureva microphone which was behind the setup as per Figure 7.

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 (see Figure 8). Reducing the distance until all three microphones were as close as possible (see Figure 9).

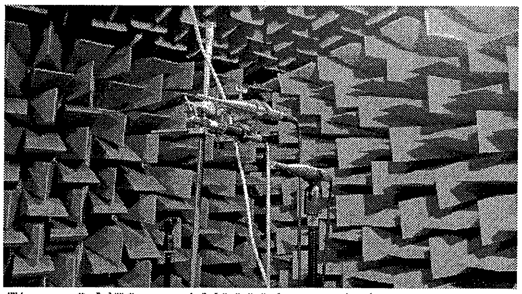

Figure 8 M50s and M2230 in proximity

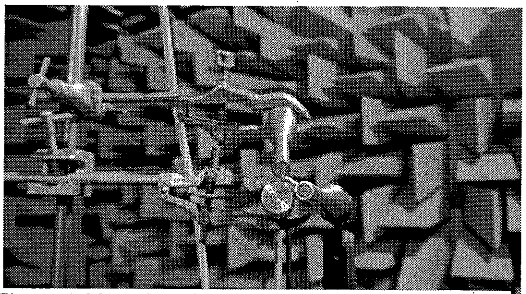

Figure 9 M2230 and M50 as close as possible

Removing the speaker improved the results, and the difference became less noticeable when the microphones were at minimal distance; however, there was a larger improvement by replacing the reference microphone by the one under test, effectively agreeing with the current standard recommendation of using substitution. As result, it was decided to keep this technique for the remaining of the experiment as it seems as if the process holds firm.
All microphones were then calibrated using a dodecahedron and a 400 Hz tone at 1m (Figure 11). The standard deviation observed between all test was of 0.15 dB as seen in Table 2; therefore, making the results comparable.

1m Source, 400 Hz Tone	NTi Value $L_{eq,15s}$	Mic result $L_{eq,15s}$	Sensitivity 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

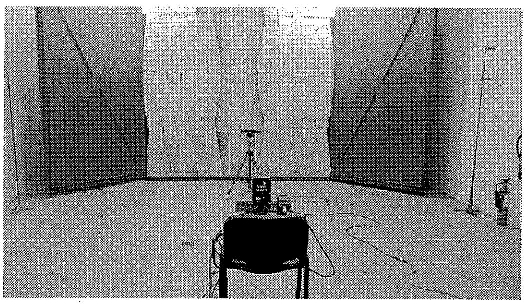

Figure 10 Recording of the audio track in semi reverberant conditions using MPR 210

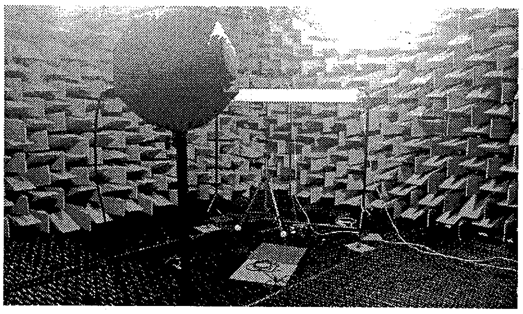

Figure 11 Testing calibration in anechoic conditions using the Nureva microphone and Omni source

4 RESULTS

4.1 Experiment 1

Nine 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 fixed afterwards.
They compared each microphone against the others for every position recorded in the room. 6 microphones were tested in 7 different positions by each expert listener, see Figure 12. The expert listeners were specifically instructed to ignore intelligibility.

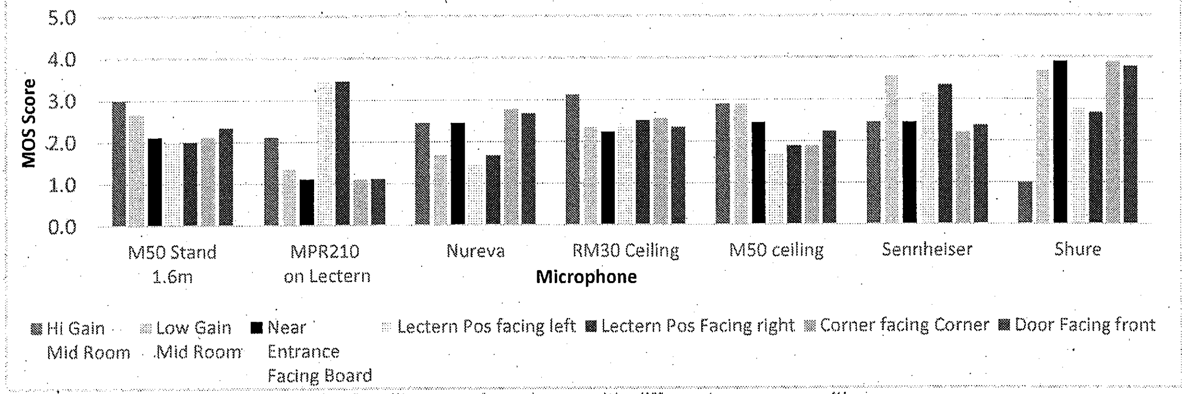

Figure 12 Quality per microphone with different source positions

It is no surprise that the microphones performed 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's position, see Table 3.

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

4.2 Experiment 2

Several microphones were set up and tested in three separate acoustical conditions (Figure 10 and Figure 11 show examples of the rooms and set up for recording and calibration respectively). An anechoic room, reverberant and semi-reverberant conditions. The NTi Minirator was loaded with an audio track with the structure:

- Pink noise generated by audition,
- Speech from for different talkers 4 male and 4 females, were downloaded from the ITU online resource[7]
- Single tones, generated with Audition 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.
While the measurement in Figure 13 took place, an audio file was also recorded to allow a comparison between the opinion scores and the L_{eq} measurements. Table 4 shows the results summarised as the average of the quality scores for all microphones and conditions. 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.
Figures 14-16 below show: the number of times a microphone in a specific condition (explained by the name: e.g. *F[final]_Microphone [Ear2, Earthworks 2]_Condition [A: anechoic, SR: semi-reverberant, FR: fully reverberant]*) obtained a rating (vertical axis: number of votes, horizontal axis: the microphone-condition pair, and series: the quality score in question given to that pair).

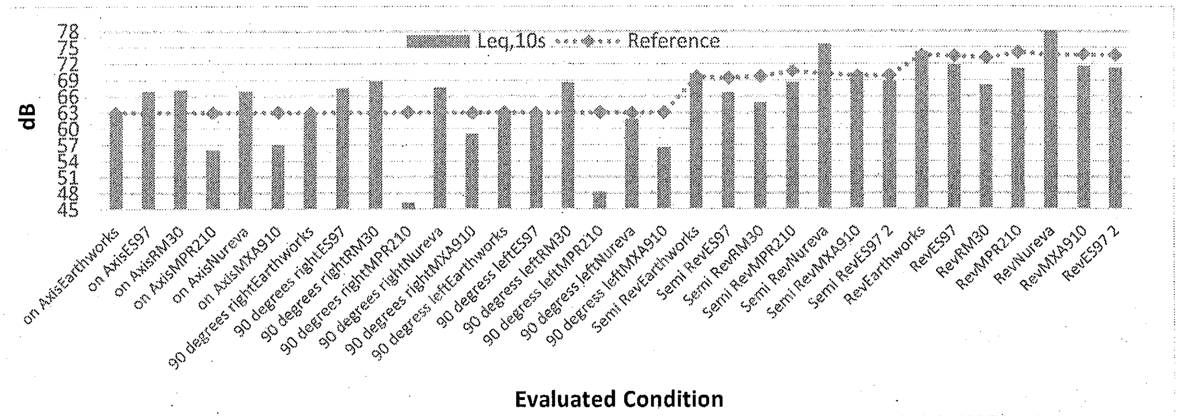

Figure 13 Difference between microphones and reference value

	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 under Lab conditions

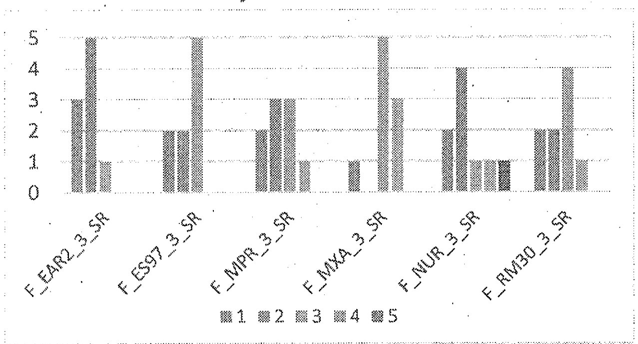

Figure 14 Quality rating semi reverberant conditions

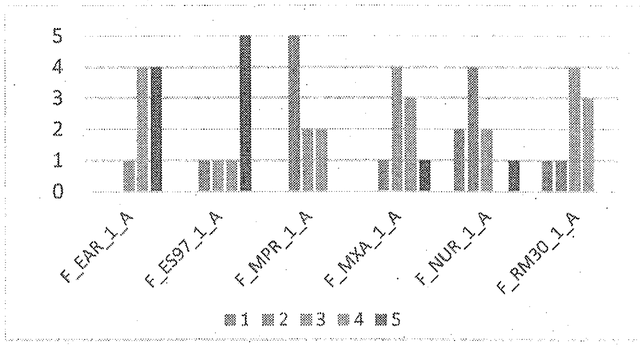

Figure 15 Anechoic microphone quality rating

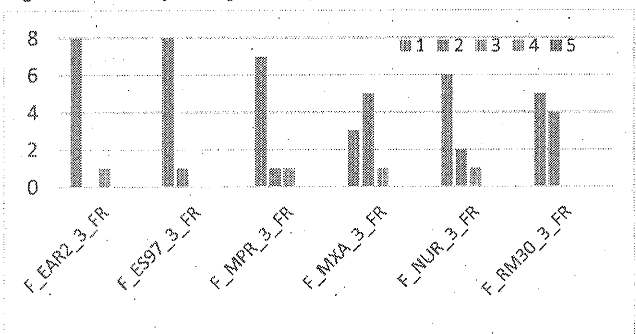

Figure 16 Semi-reverberant microphone quality rating

rating given by expert listeners.

5 DISCUSSION

An analysis of the resulting quality (Figures 14-16) 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 5 shows, there was no significant relationship between, how accurate the microphone was, and the subjective quality

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 5 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 speaker is facing away, and high reverberation lab conditions, as seen in Table 6.The same cannot be said for the lab tests under anechoic conditions.
This is consistent with the calibration results; all microphones were very accurate 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 13; 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		0.281	-.846*	-.848*	-0.396	-0.331	-0.567	-0.515	-0.454	-0.533	-0.520
Anechoic	0.281		-0.320	-0.423	0.651	0.644	0.187	-0.197	-0.302	-0.084	0.097
Rev	-.846*	-0.320		.965**	0.478	0.492	.782*	0.325	0.280	.844*	.772*
Semi_rev	-.848*	-0.423	.965**		0.407	0.375	0.691	0.370	0.359	.738*	0.653
Total	-0.396	0.651	0.478	0.407		.974**	.801*	0.016	-0.095	0.569	0.683
Low gain Middle	-0.331	0.644	0.492	0.375	.974**		.853*	-0.098	-0.218	0.670	.775*
Entrance Facing Away	-0.567	0.187	.782*	0.691	.801*	.853*		-0.181	-0.267	.936**	.978**
Lectern Facing Left	-0.515	-0.197	0.325	0.370	0.016	-0.098	-0.181		.986**	-0.186	-0.265
Lectern Facing Right	-0.454	-0.302	0.280	0.359	-0.095	-0.218	-0.267	.986**		-0.248	-0.349
Corner Facing Corner	-0.533	-0.084	.844*	.738*	0.569	0.670	.936**	-0.186	-0.248		.974**
Door Facing Front	-0.520	0.097	.772*	0.653	0.683	.775*	.978**	-0.265	-0.349	.974**	

Table 6 Correlation matrix of all quality ratings

6 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 L_{eq} by the microphone under test and a calibrated laboratory system, was compared to the quality rating from expert listeners and the resulting analysis showed that there was no significant relationship.
There was no microphone that was always the best for the subjective ratings, with omnidirectional mics (Earthworks M50) being the top 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 under anechoic conditions.

7 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 rate the subjective quality of recordings and microphones used in teaching rooms.

8 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) Equipments. (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. ITU, ITU-T, T-REC-P.800 TELEPHONE TRANSMISSION QUALITY Methods for objective and subjective assessment of quality. (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.
8. ARTA Software. http://artalabs.hr/.